GI-Edition

Lecture Notes
in Informatics

Gesellschaft für Informatik (Hrsg.)

SKILL 2021

Studierendenkonferenz Informatik

28. September und 01. Oktober 2021

Gesellschaft für Informatik (Hrsg.): SKILL 2021

17

Seminars

GESELLSCHAFT
FÜR INFORMATIK

Gesellschaft für Informatik (Hrsg.)

SKILL 2021
Studierendenkonferenz Informatik

28. September/01. Oktober 2021

Gesellschaft für Informatik e.V. (GI)

Lecture Notes in Informatics (LNI) - Seminars
Series of the Gesellschaft für Informatik (GI)

Volume S 17

ISSN 1614-3213
ISBN 978-3-88579-751-7

Volume Editors
Gesellschaft für Informatik e.V.
Ahrstraße 45
53175 Bonn
E-Mail: bonn@gi.de
Redaktion: Michael Becker
E-Mail: becker@infai.org

printed by Köllen Druck+Verlag GmbH, Bonn

Vorwort

Die Studierendenkonferenz Informatik (SKILL) ist eine jährliche Konferenz für Studentinnen und Studenten der Informatik sowie angrenzender Disziplinen aus ganz D-A-CH. Die Intention des SKILL-Konferenzformats ist es, sehr guten studentischen Arbeiten eine öffentliche Plattform zur Diskussion zu bieten. Die Studierenden können Erfahrungen zum wissenschaftlichen Publizieren sammeln und ihre Ergebnisse vor einem breit gefächerten Publikum vorstellen.

Nach einer ersten coronabedingten virtuellen SKILL im Jahr 2020, findet auch die SKILL 2021 wieder ausschließlich online statt. Es hat sich gezeigt, dass auch im Online-Format kritische und konstruktive Diskussionen nicht zu kurz kommen. Darüber hinaus ermöglicht der einfache Zugang auch eine Öffnung für Teilnehmerinnen und Teilnehmer, die bisher aus organisatorischen Gründen nicht an der SKILL teilnehmen konnten. Trotz allem fehlt natürlich der unmittelbare und persönliche Austausch, der sich nur sehr schwer mit digitalen Mitteln unterstützen lässt.

Die Erfahrungen der letzten zehn Jahre zeigen, dass ein Format wie die SKILL wichtig ist, um Studierende für eine wissenschaftliche Diskussion und in letzter Konsequenz für eine wissenschaftliche Arbeit zu begeistern. Die SKILL wird daher auch in Zukunft offen für alle Themen der Informatik bleiben und die gesamte Vielfalt unseres Faches widerspiegeln.

In diesem Jahr wurden insgesamt 26 Beiträge als Full oder Short Paper eingereicht und wissenschaftlich begutachtet. In diesem Band erscheinen 15 Beiträge, die auf zwei Konferenztagen durch die Studierenden präsentiert werden. Neben dem großen Thema Machine Learning präsentiert die SKILL in diesem Jahr auch eine Reihe von Beiträgen, die sich mit der Anwendung der Informatik in verschiedenen Domänen sowie der gesellschaftlichen Relevanz der Informatik beschäftigen.

Die Mitglieder des Organisationskomitees der SKILL 2021 bedanken sich bei den Autorinnen und Autoren, ohne deren hochwertige Beiträge die Konferenz nicht möglich wäre. Wir freuen uns darüber hinaus, dass wir auch in diesem Jahr namhafte Gutachterinnen und Gutachter gewinnen konnten, die den Studierenden mit hilfreichen und ausführlichen Kommentaren zu ihren Arbeiten zur Seite standen.

Leipzig & Berlin, 17. September 2021

Organisationskomitee der SKILL 2021

- Michael Becker, Institut für Angewandte Informatik e.V.
- Judith Michael, Rheinisch-Westfälische Technische Hochschule Aachen
- Thomas Riechert, Hochschule für Technik, Wirtschaft und Kultur Leipzig
- Ludger Porada, Gesellschaft für Informatik e.V.

Gutachterinnen und Gutachter der SKILL 2021

Michael Bader	Technische Universität München
Matthias Book	University of Iceland
Julia Friedrich	Institut für Angewandte Informatik e.V.
Kurt Geihs	Universität Kassel
Martin Grimmer	Universität Leipzig
Otthein Herzog	Tongji University, Shanghai, und Universität Bremen
Walter Hower	Hochschule Albstadt-Sigmaringen
Paul Jähne	Fraunhofer-Institut für Zelltherapie und Immunologie
Christine Jakobs	Technische Universität Chemnitz
Stephan Klingner	Universität Leipzig
Christian Kücherer	Hochschule Reutlingen
André Langer	Technische Universität Chemnitz
Jan-Patrick Lehr	Technische Universität Darmstadt
Christian Martin	Universität Leipzig, ScaDS.AI
Ludger Martin	Hochschule RheinMain Wiesbaden
Lars-Peter Meyer	Universität Leipzig
Helmut Neukirchen	University of Iceland
Michael Perscheid	Hasso-Plattner-Institut
Olaf Resch	Hochschule für Wirtschaft und Recht Berlin
Petra Sauer	Beuth Hochschule für Technik Berlin
Thomas Schmid	Universität Leipzig
Andreas Schmidt	Saarland Informatics Campus
Michael Schöttner	Universität Düsseldorf
Peter Thiemann	Albert-Ludwigs-Universität Freiburg
Ralf Wimmer	Albert-Ludwigs-Univertsität Freiburg

GI Mentoringprogramm

Das Mentoringprogramm der GI bietet motivierten und engagierten GI-Mitgliedern eine Hilfestellung zur Erreichung ihrer persönlichen Karriereziele – betreut von kompetenten und erfahrenen Mentorinnen und Mentoren.

VON MITGLIEDERN FÜR MITGLIEDER

Um was geht's?

- Unterstützung bei der Planung und Umsetzung konkreter Vorhaben und Karriereschritte
- Persönlichkeitsentwicklung und Coaching
- Wege in Beruf oder Wissenschaft aufzeigen

Für wen?

- Masterstudentinnen und -studenten der Informatik sowie angrenzender Disziplinen mit dem Schwerpunkt Informatik
- Lehramtsstudierende mit dem Fach Informatik
- Berufsanfängerinnen und Berufsanfänger aus der Informatik
- Nachwuchswissenschaftlerinnen und -wissenschaftler

Anmeldung und mehr Infos:

Das Mentoringprogramm ist für GI-Mitglieder natürlich kostenlos.

GESELLSCHAFT FÜR INFORMATIK

mentoring.gi.de

Inhaltsverzeichnis

Informatik in der Anwendung

Concolic-Fuzzing of JavaScript Programs using GraalVM and Truffle

Robert Delhougne[1]

Abstract: The scripting language JavaScript has established itself as a central component of the modern internet. However, the dynamic execution model of the language limits the support for source-code analysis, which leaves a developer without essential tools to maintain safety and security requirements. This paper describes a concolic-fuzzer based on the GraalVM to automatically test JavaScript programs. The fuzzer shows promising results in both code coverage and runtime evaluations and provides developers with additional features such as special analysis targets.

Keywords: Software-Verification; Fuzzing; Concolic-Fuzzing; JavaScript; GraalVM; Truffle

1 Introduction

Originally intended as a simple language for interactive media on a web application's client-side, the scripting language JavaScript is now also used for server- and desktop applications. However, a study by Ocariza et al. [OPZ11] shows that programming with JavaScript can be difficult because even subtle programming errors can lead to severe repercussions later in the program's execution. Moreover, due to the aggressive type-coercing, these errors can propagate through the program for a long time before being detected and thus hiding the initial cause [PS15]. These features of the language and their consequences complicate the development process and point to the need for more precise tools for checking code quality.

In this paper, the author presents a tool that can automatically test JavaScript programs for errors with a technique called *concolic-fuzzing*, an advanced software-testing method that has gotten more attention in recent years [Ba18; Ka15]. This fuzzer uses the virtual machine GraalVM and its language implementation framework Truffle to execute the JavaScript programs and keep track of the inner workings of the program to guide the testing process. One advantage of this method is that it is fully automated and does not require any modification of the language or program to guide the testing process. To address the difficulties of JavaScript, the fuzzer has special analysis targets to detect the errors characterized by Ocariza et al. [OPZ11] or Pradel and Sen [PS15] as early as possible. In a preliminary evaluation, the fuzzer reached between 50% to 98% branch coverage without the need to specify any information about the input structure of the test programs.

[1] TU Dortmund, August-Schmidt-Straße 4, 44227 Dortmund, Germany, robert.delhougne@tu-dortmund.de

Related Work One of the first tools using concolic-execution is DART, short for *directed automated random testing* [GKS05]. DART can be used to test the programming interfaces of C-programs automatically. Build on the ideas of DART, the testing framework jDART proposed by Luckow et al. [Lu16] can test Java programs. The concolic-execution of the program is performed using the Java-Pathfinder [Vi03] framework. Another well-known program for symbolic execution is SAGE [GLM08]. Godefroid et al. are stating that it was extensively used in the development of the operating system Windows 7, claiming that about one-third of all errors found with testing based on files are found by this tool. SAGE works by directly utilizing x86 instructions and thus is capable of testing all programming languages that can be compiled to x86 machine-code. The support for dynamic languages like JavaScript in symbolic execution tools is much less common. Saxena et al. [Sa10] presented a symbolic execution framework for JavaScript called *Kudzu*. Kudzu uses a simplified version of the JavaScript language to reduce the complexity of the symbolic model called *JASIL*. The tool is used to test inputs for web applications and utilizes a modified version of the WebKit engine to observe the application at runtime. Sun et al. [Su18] also presented a software testing framework for Node.js applications that utilizes source-code instrumentation, but does not use true symbolic or concolic execution techniques like they are employed in this work.

2 Motivational Example

```
function factorial(n) {
  if (n >= 0) {
    fac = 1;
    while (n > 1) {
      fac = fac * n;
      n = n - 1;
    }
    return fac;
  }
}
```

List. 1: Algorithm to calculate the factorial of *n*.

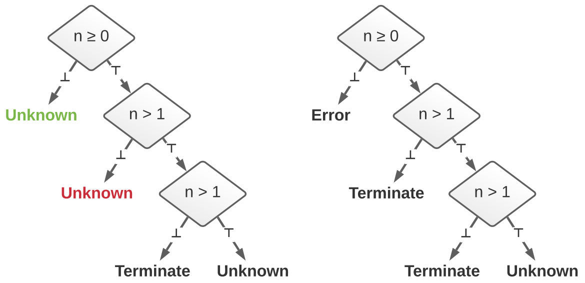


Fig. 1: Execution trees of `factorial(n)`.

To visualize the methodology of fuzzing in general and specifically the principles of concolic-fuzzing, we look at the small JavaScript function shown in Listing 1. The function call of `factorial(n)` with $n < 0$ introduces an error state into the surrounding program (e. g. `factorial(-1)` returns the value `undefined`). To test this function, besides approaches like unit testing, the automated technique *fuzzing* can be used. This method aims to generate a vast number of (pseudo-) random input data to test the software. If the software reaches a failure state, the fuzzer detects it and saves the corresponding input data. However, a naive, completely random fuzzer often has a low probability of triggering errors in a program

and needs redundant test iterations. Modern fuzzers use additional information about the inputs or the program to narrow down the generated input data and speed up the testing circumventing this issue. One of these advanced techniques used in this paper is called *concolic-fuzzing*.

Let's demonstrate concolic-fuzzing using the example program in Listing 1. In the first step, the fuzzer executes the program with random inputs, in this example the input of function `factorial(n)` is set to $n = 2$. Simultaneously with running the program with this concrete value, the fuzzer treats the input as a *symbolic* value and records all modifications to this variable. This implies that the fuzzer can fully observe all procedures of the tested program at runtime, down to single operators and variable accesses. With the help of the symbolic variables, the fuzzer constructs an execution tree to keep track of the already executed paths of the program. Figure 1 (left) shows the execution tree for function `factorial(n)` after the first iteration with $n = 2$. For every new test iteration, the fuzzer chooses a different, unknown execution path from this model, e. g., the red "Unknown" node. All the branch conditions starting from the root of the tree down to the unknown location are then collected and combined to form a *path-condition*:

$$n \geq 0 \quad \wedge \quad \neg(n > 1). \tag{1}$$

To find a value for n that fulfills these conditions, the fuzzer uses an *SMT-Solver* like the solver Z3[2] developed by Microsoft. In this example, the only possible solution is $n = 1$. If the program is executed with this value as an input, it takes exactly[3] this predicted execution path, and the fuzzer can extend the execution tree. To trigger the failure state in the program, the fuzzer only has to collect the conditions for the green node (only $\neg(n \geq 0)$) and then run the program with an n satisfying this condition. Figure 1 (right) also depicts the execution tree after these two iterations.

With this method, a concolic-fuzzer quickly achieves very high code coverage because all new inputs are tailored for a new execution path. Code paths that check the input data for distinct patterns and usually pose a hard to overcome barrier for naive fuzzers can easily get around.

3 Implementation

As demonstrated in the example, a concolic-fuzzer needs to observe the tested program at runtime and must be able to modify the execution e. g. to inject new input values directly into the running program instead of relying on reading the data from the file system. In this implementation, I use the GraalVM and the Truffle framework for this purpose. The

[2] `https://github.com/Z3Prover/z3`

[3] There are exceptions, e. g., if the program flow depends on a non-deterministic variable or the fuzzer does not exhaustively model the program flow.

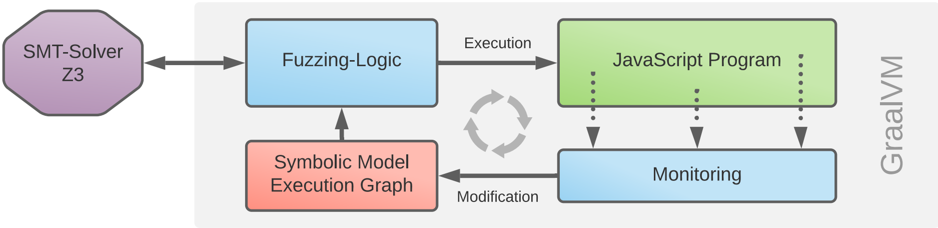


Fig. 2: Overview of the fuzzer's architecture.

GraalVM is a virtual machine for Java, developed by Oracle, based on the HotSpot VM. Its first production-ready version was released in 2019. Besides Java, it supports a wide variety of additional languages through Truffle, which is a framework to implement an interpreter for *abstract-syntax-trees* (ASTs) [Wü13]. The behavior of the interpreted language, the so-called *guest-language*, is implemented through Java classes. One benefit of the Truffle architecture is that it also allows developers to deeply inspect and alter the execution of a guest-language program at runtime, based on the AST representation of the program. In the concept of GraalVM, these extensions are called *tools*.

Figure 2 shows the simplified architecture of the fuzzer. All of the components, except the SMT-Solver, are running inside of the GraalVM. The program logic of the fuzzer creates a *language-context*, that parses the source code of the executed program, builds the AST, and then executes it. Simultaneously, the fuzzer monitors all operations of the AST and creates the symbolic model of the program (cf. Section 3.1). The execution tree grows with every iteration and based on that, the fuzzer chooses a new execution path according to a specific strategy (cf. Section 3.2), solves the path constraints with the help of an SMT-Solver (in this case, Z3), and executes the JavaScript program again with the newfound input values. When the program is running, the fuzzer checks it for exceptions or special error analysis targets (cf. Section 3.3). The three main components of the implementation are presented in more detail in the following subsections.

3.1 Symbolic Flow

To keep track of the variable modifications inside the running JavaScript program, the fuzzer uses the Truffle API concept of *wrapper-nodes*. In this work these nodes are used to define a symbolic behavior for every AST-node of the program, in addition to the concrete behavior that is already given by the language implementation. The symbolic behavior depends on the type of the AST-node, for example, arithmetic nodes, variable read/write or constants. The wrapper-nodes are dynamically attached to nodes in the (JavaScript-) AST and can listen to specific *events*, e. g., before the corresponding AST-node gets executed, a new input is available, or execution of the node has finished. With the help of the wrapper nodes, a developer can extend the behavior of the AST-nodes as desired. The symbolic model of the

program, modified by the wrapper-nodes, consists of two central data structures that are essentially representing a load-store-architecture.

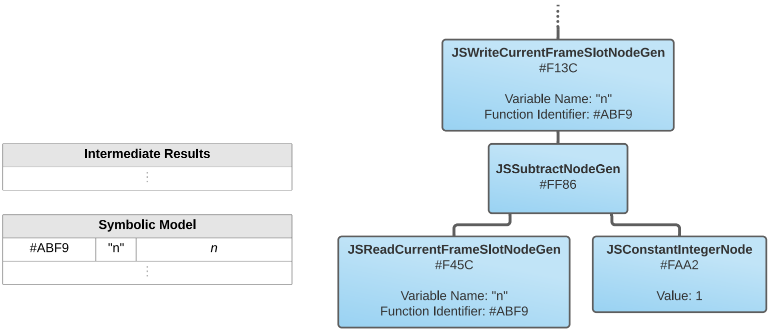
Intermediate Results The fuzzer saves all symbolic counterparts to intermediate results, computed by the AST-nodes, into a data structure consisting of a unary function $i \rightarrow S$, mapping integer identifiers i onto a data structure for symbolic expressions S. The integer i is a unique identifier of an AST-node of the instrumented JavaScript program. This can be thought of as symbolic "register".

Symbolic Model The symbolic model is a binary function $k \times s \rightarrow S$, which contains a flat hierarchy of the symbolic representation of all the data structures currently present in the observed JavaScript program. The integer identifier k, in this case, represents a JavaScript object, the string s represents the attribute name. This data structure can model attributes inside objects, arrays, and stack frames. This data structure is a symbolic "memory", where the intermediate results get written to when their concrete counterpart is assigned to a JavaScript variable.

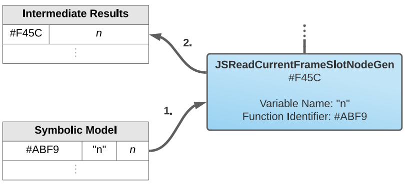
Figures 3a to 3e show an example of the technique for the JavaScript expression `n = n - 1` that is part of the small JavaScript function `factorial(n)`, shown in Listing 1. The expression reads a local variable n, subtracts 1 and saves the result back to the local variable. The initial situation is shown in Figure 3a. First, given the information about the current function context, the `JSReadCurrentFrameSlotNodeGen` loads the symbolic expression n from the variable "n" of the current function from the symbolic model into the intermediate results (Figure 3b). Then, the `JSConstantIntegerNode` is evaluated (Figure 3c). The wrapper of this node only stores a symbolic constant into the intermediate results. Both of these intermediate results are then used by the instrumentation of the `JSSubtractNodeGen` to construct the symbolic subtraction operation (Figure 3d). In the last step, this symbolic result is transferred back to the symbolic model, just like it is saved as a local variable in the simultaneously running concrete execution of the program (Figure 3e).

3.2 Path Exploration

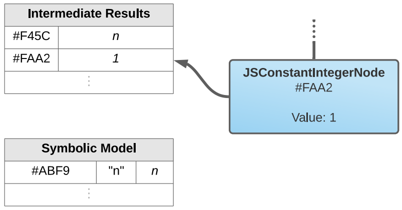
To construct the execution tree of the program, two types of AST-nodes have to be instrumented: The node type `IfNode` for branching instructions and the node type `WhileNode` to cover `while` and `for` loop statements. A separate node type handling `for`-loops does not exist; this loop statement is also handled by the `WhileNode`. The execution tree in this fuzzer is modeled as a state-machine, meaning every new iteration of the JavaScript program, the current position inside the execution tree is reset to the root of the tree. While the fuzzer runs the JavaScript program, the instrumentation of the `IfNodes` and `WhileNodes` keep track of the current position inside the tree. When the program enters an unknown part



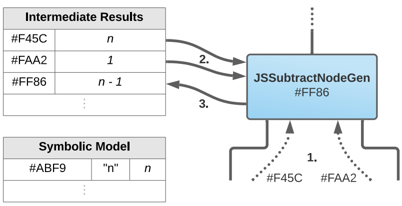
(a) Initial condition for the expression `n = n - 1` from the function `factorial(n)`.



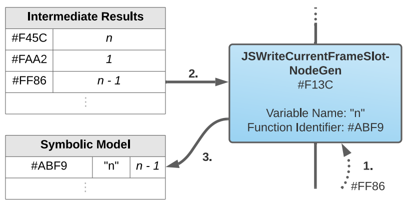
(b) Transfer of the symbolic value *n* from the model to the intermediate results.



(c) Creation of a new symbolic constant.



(d) Calculation of the symbolic subtraction operation.



(e) Transfer of the intermediate result back to the model.

Fig. 3: Example of the concolic execution based on AST-nodes.

of the tree, meaning this execution path is run for the first time, the tree gets extended by the instrumentation. The wrapper-node transfers the matching branch condition from the intermediate results to the newly created execution tree node.

One of the challenges of symbolic execution is the selection of the next possible execution path. In this work, I implemented three different strategies to select a new, unknown execution path. They are listed in Table 1. For the comparison of the strategies, refer to Section 4. The user can further configure these strategies and other parts of the fuzzer with

Strategy	Description
DEPTH_SEARCH	Of all the unknown candidate leaves of the execution tree, select the one that is the farthest away from the root of the tree.
IN_ORDER_SEARCH	Traverse the tree in-order and return the first unknown node that is found.
RANDOM_SEARCH	Traverse the tree from the root. At each node, switch into one of the child nodes with a possibility of $\frac{1}{2}$ until an unknown node is found. If a leaf is reached but it is not unknown, use backtracking.

Tab. 1: List of the implemented search strategies for unknown execution paths in the execution tree.

a YAML configuration file. One example of a parameter is the maximum search depth in the execution tree, which prevents the fuzzer from forming too long path predicates that are hard to solve in later stages of the test process.

3.3 Special Analysis Targets

To further adapt the fuzzer to the JavaScript language, I extended the fuzzer to analyze the tested program for common programming mistakes. A known problem of JavaScript, which is also described by Pradel and Sen [PS15], is that the aggressive type-coercing of JavaScript can suppress errors, for example, in the form of an undefined value, for a long time, before they are finally causing an unrecoverable error that is noticeable by the developer or user. This hides the actual cause of the error. To circumvent this problem, this work extends the fuzzer by enhanced error guards. These guards describe a valid JavaScript behavior under normal circumstances that the fuzzer treats as an error condition when the user enables the error guard during the test process. An example is the guard `division_op_no_zero` that prevents the evaluation of a zero-division to `+Infinity` and instead throws an exception that the fuzzer handles as a faulty state of the program. With the corresponding input value and line number, the developer can reproduce the error at the root cause. These extensions to the expected behavior of JavaScript applications also testify to the power and versatility of the code instrumentation with the Truffle API.

4 Results

The evaluation objectives for the project are the the effectiveness of the three implemented search strategies for path exploration and the total runtime of the testing process. The evaluation set consists of about 15 JavaScript programs, ranging from simple functions to complex programs with hundreds or thousands of lines. Table 2 shows a selection of these programs and their properties. The programs are from three different sources: First, programs written by the author, which are small with few inputs. Secondly, the work used programs from the “Benchmarksgame”[4] project. Third, the evaluation uses several

[4] `https://benchmarksgame-team.pages.debian.net/benchmarksgame`

Name	Symbolic Inputs	LOC	Max. cycl. Compl.	# Func.
`fasta.js`	1 String	106	6	4
`calculator.js`	2 Integer, 1 String	296	23	22
`infusion.js`	37 Integer, 22 Boolean	1165	70	23
`alarm.js`	65 Integer, 43 Boolean	2143	66	28
`minepump.js`	2 Integer, 16 Boolean	331	19	38
`addition01.js`	2 Integer	36	6	2
`triangle.js`	3 Integer	46	8	2

Tab. 2: A selection of test programs and their properties. Listed are the number and types of the symbolic input variables for each program, the number of lines (LOC), the maximum cyclomatic complexity per function, and the total number of functions.

programs from the SV-COMP[5] benchmark, which have been converted manually from Java to JavaScript. All of the measured coverage metrics and runtimes were calculated on a test system, consisting of an AMD Ryzen 7 4800HS (16 vCores) and 16 GiB of RAM. For detailed information about the used software versions please refer to the project page[6] on GitHub.

4.1 Code Coverage Metrics

To evaluate the effectiveness of the fuzzing technique in general and compare the different path exploration strategies, the fuzzer calculates three different coverage metrics: statement coverage (C_0), function coverage, and branch coverage (C_1). This evaluation will focus on branch coverage. Figure 4 shows the calculated metric for some of the example programs. I configured the fuzzer to stop the testing on a maximum of 3000 iterations or if it reaches 95% branch coverage. Moreover, I configured the search strategies to search in a maximum depth of 64 in the execution tree. As the `RANDOM_SEARCH` strategy is nondeterministic , I performed three different runs for this strategy to visualize the differences per run. In the first execution of the test programs, the fuzzer always executes the programs with default values for all inputs, namely `true` for boolean inputs, `1` for numeric values, and the empty string for string inputs.

While the fuzzer quickly improves the code coverage against the first run in some of the programs, namely `calculator.js`, `infusion.js`, or `triangle.js`, other programs like `alarm.js` or `minepump.js` are less suited for this testing technique. This is likely since these programs are simulating state machines. These programs prevent the execution of certain code paths due to their structure, so a coverage like in `alarm.js` cannot be improved much, even with many iterations. The fuzzer improves the test coverage of the program `minepump.js` from 71% to 79%, the coverage of `alarm.js` is improved by 48.8% from 33.6% to 50%. The coverage metric can be improved much more on the other programs,

[5] `https://sv-comp.sosy-lab.org`
[6] `https://github.com/rdelhougne/Amygdala`

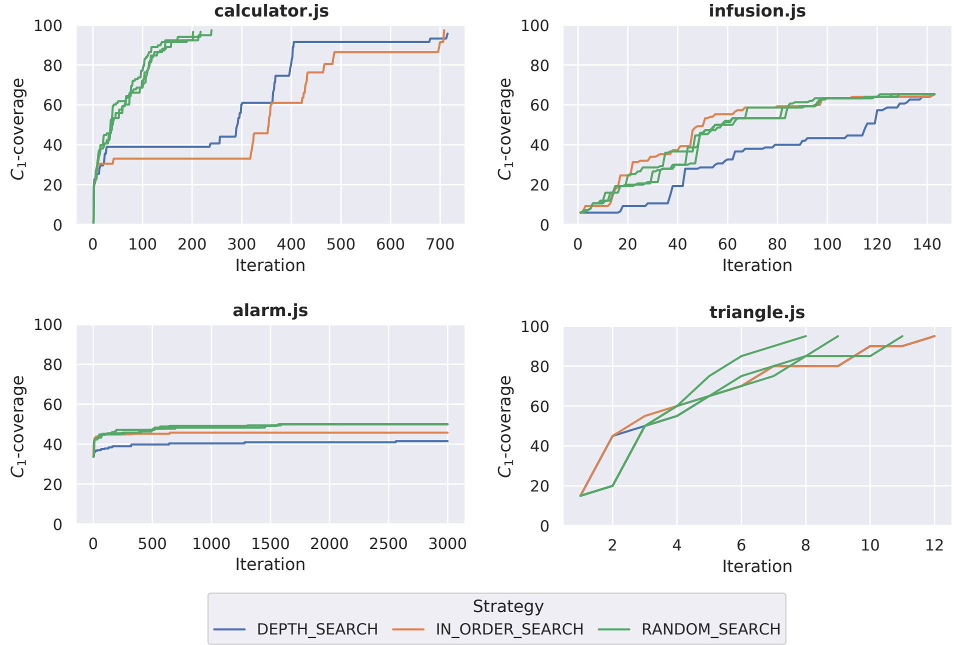


Fig. 4: Branch coverage of various test programs.

with `calculator.js` and `triangle.js` almost reaching 100% branch coverage. Notably, the tactic `RANDOM_SEARCH` always performs equal, if not better, than the other tactics. Especially with the program `calculator.js`, where it can consistently reach a high code coverage in about one-third of the iterations, or with program `alarm.js`, where the random tactic always reaches a higher code coverage, although by a small margin.

4.2 Runtime Evaluation

The total run time measurements further underline the performance of this strategy. Figure 5 shows the total runtime of the fuzzing processes. The random strategy needs needs much less time to test `alarm.js` and `minepump.js`, despite running the same number of iterations. The runtime differences in these cases were caused by path constraints that take much more time to be solved by the SMT-solver and are simply less likely to be hit by the `RANDOM_SEARCH` strategy because this strategy does not exhaustively explore the execution tree with respect to boundaries like maximum depth. This effect is particularly noticeable in the program `calculator.js`. In addition, the smaller number of iterations (cf. Figure 4) ensures that the random strategy with a runtime of 7.5 seconds only needs a fraction of the time compared to the other strategies.

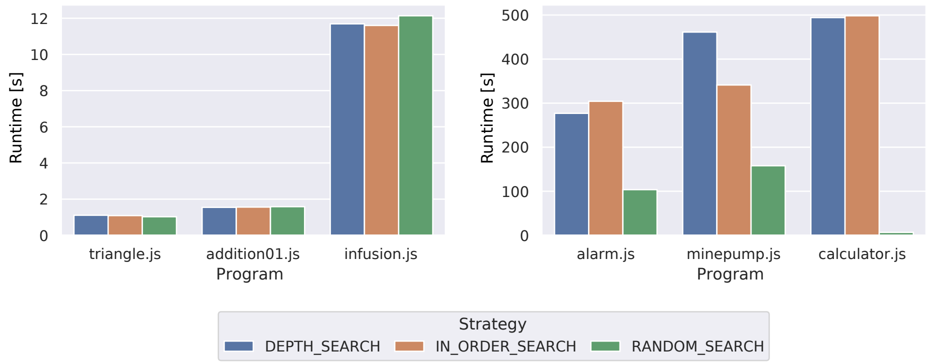


Fig. 5: Total runtime of the fuzzing process for various test programs.

5 Discussion

Section 4 shows a preliminary evaluation of the fuzzer. The applicability of the fuzzer in real-world applications is, however, limited at this time. This has two leading causes: First, the fuzzer only does support a subset of the ECMAScript standard to this date. This subset includes all arithmetic, logical, and comparison operators (except for bitwise operators) and several frequently used string and math operations. Despite the extensive support of operators, the presented fuzzer is relying purely on concolic-execution. Therefore, it needs to model the whole execution flow of the tested program exhaustively. If the fuzzer or the SMT-Solver does not support just one statement and a branch depends on this statement, the path conditions for this branch cannot be constructed or solved. As a result, every part of the program lying behind this branch cannot be reliably reached and tested by the fuzzer. A solution to this problem is the extension of the support for the ECMAScript standard, which is work-intensive as explained below, or to build a *hybrid* fuzzer. Such a fuzzer uses a more straightforward method like grammar-fuzzing for most of the testing procedure and only relies on concolic/symbolic execution at difficult to reach sections of the program [MS07].

The results from chapter 3 showed that the Truffle API is an elegant way to monitor programs at runtime. The tested programs had a high execution speed despite the instrumentation due to the simultaneous optimization of AST-interpreter and tool by the GraalVM. However, the usage of wrapper-nodes imposes a drawback to this architecture: The detailed information needed to successfully implement concolic-execution forces the programmer to directly work with the low-level language implementation, instead of one of the other high-level interfaces Truffle provides. As a result, the programmer has to define a symbolic behavior for many different nodes, each of which can be arbitrarily complex. A search for class definitions for nodes in the JavaScript language implementation resulted in 634 definitions, most of which potentially can appear in an executed AST. This quantity is a lot more than the maximum of 256 opcodes of the Java VM [Li20], or the strictly defined set

of operations in LLVM-IR [Ll21]. The problem is worsened by nodes with extremely complex behavior like `JSArrayJoinNodeGen`[7]. The behavior of this node has to be manually symbolically modeled because the execution of the node is only observable as a whole. The dependency on the low-level JavaScript implementation also prevents the fuzzer from being easily ported to other languages. To make this possible, Truffles's high-level polyglot API would have to provide more specific information, which contradicts its deliberately abstract implementation.

6 Conclusion and Outlook

Due to the dynamic nature of the JavaScript language, the support for sophisticated code analysis tools lacks compared to other, more strict languages like C/C++ or Java. This work shows the implementation of a concolic-fuzzer for JavaScript using the capabilities of the Truffle API as part of the GraalVM. The implemented fuzzer extended the standard behavior of the AST interpreter to extract the required information about the tested program at runtime. The study of this approach on several test programs contained an evaluation of code-coverage metrics and a consideration of runtime measurements. The fuzzer showed promising results on the testing scenarios, reaching more than 95% branch coverage on some programs. This result is accomplished without providing any information about the input structure of the programs. This testing approach is therefore completely automated. In addition, the fuzzer is tuned to the peculiarities of JavaScript through the special analysis targets and can thus detect common errors at an early stage. While the use of GraalVM and Truffle provides a highly efficient and fast execution speed of the tested program and the fuzzer, the implementation of the fuzzer was also work-intensive due to the high number of different data structures present in the AST. Therefore, a direction of future work is the modification of the language implementation to bring the requirements of execution speed and easy-to-use instrumentation closer together.

Acknowledgement

I thank my supervisors Falk Howar and Malte Mues for the fruitful discussions and feedback during the master thesis leading to this paper.

References

[Ba18] Baldoni, R. et al.: A Survey of Symbolic Execution Techniques. ACM Computing Surveys 51/3, 50:1–50:39, 2018.

[7] `JSArrayJoinNodeGen` corresponds to the JavaScript function `Array.prototype.join()`.

[GKS05] Godefroid, P.; Klarlund, N.; Sen, K.: DART: Directed Automated Random Testing. In: Proceedings of the 2005 ACM SIGPLAN Conference on Programming Language Design and Implementation. PLDI '05, ACM Press, Chicago, IL, USA, pp. 213–223, 2005.

[GLM08] Godefroid, P.; Levin, M.; Molnar, D.: Automated Whitebox Fuzz Testing. In: Proceedings of the Network and Distributed System Security Symposium. NDSS '08, The Internet Society, San Diego, CA, USA, pp. 151–166, 2008.

[Ka15] Kannavara, R. et al.: Challenges and opportunities with concolic testing. In: 2015 National Aerospace and Electronics Conference. NAECON '15, IEEE, Dayton, OH, USA, pp. 374–378, 2015.

[Li20] Lindholm, T. et al.: The Java® Virtual Machine Specification, Java SE 15 Edition, 2020.

[Ll21] LLVM Language Reference Manual, URL: `https://llvm.org/docs/LangRef.html`, visited on: 05/12/2021.

[Lu16] Luckow, K. et al.: JDart: A Dynamic Symbolic Analysis Framework. In: Tools and Algorithms for the Construction and Analysis of Systems. TACS '16, Springer, pp. 442–459, 2016.

[MS07] Majumdar, R.; Sen, K.: Hybrid Concolic Testing. In: 29th International Conference on Software Engineering. ICSE '07, IEEE, Minneapolis, MN, USA, pp. 416–426, 2007.

[OPZ11] Ocariza, F.; Pattabiraman, K.; Zorn, B.: JavaScript Errors in the Wild: An Empirical Study. In: Proceedings of the 2011 IEEE 22nd International Symposium on Software Reliability Engineering. ISSRE '11, IEEE, Hiroshima, Japan, pp. 100–109, 2011.

[PS15] Pradel, M.; Sen, K.: The Good, the Bad, and the Ugly: An Empirical Study of Implicit Type Conversions in JavaScript. In: 29th European Conference on Object-Oriented Programming. ECOOP '15, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl, Germany, pp. 519–541, 2015.

[Sa10] Saxena, P. et al.: A Symbolic Execution Framework for JavaScript. In: Proceedings of the 2010 IEEE Symposium on Security and Privacy. SP '10, IEEE, Washington D.C., USA, pp. 513–528, 2010.

[Su18] Sun, H. et al.: Efficient Dynamic Analysis for Node.js. In: Proceedings of the 27th International Conference on Compiler Construction. CC '18, ACM Press, Vienna, Austria, pp. 196–206, 2018.

[Vi03] Visser, W. et al.: Model Checking Programs. Automated Software Engineering 10/2, pp. 203–232, 2003.

[Wü13] Würthinger, T. et al.: One VM to Rule Them All. In: Proceedings of the 2013 ACM International Symposium on New Ideas, New Paradigms, and Reflections on Programming & Software. Onward! '13, ACM Press, New York, NY, USA, pp. 187–204, 2013.

Comparison of material models in modern physically based rendering pipelines

Franca Bittner[1]

Abstract: The appearance of materials results from a complex interaction of light, material properties and the geometric shape of an object. In computer graphics, various models were developed to describe these correlations. Modern rendering pipelines commonly adapt the philosophy of physically based rendering (PBR). This study examines if the reproduction of materials differs across modern PBR tools, and compares the intuitiveness of material design, the quality and range of reproducible materials. A sequential rendering framework was developed to evaluate the visual influences of four selected parameters on material appearance. The rendered images are qualitatively compared based on material charts, scanline plots and difference images. The examined rendering tools mostly yield similar results, with the main differences caused by disparate rendering methods. Still, subtle variations between the tools are noticable, indicating the individual strengths and flaws of each renderer in terms of intuitiveness and physical accuracy.

Keywords: Physically Based Rendering; Material Models; Material Appearance; Rendering Pipeline

1 Introduction

In many today's digital production workflows, physically based rendering (PBR) plays a central role in the working of a 3D artist. The main idea of PBR is to reproduce the virtual world based on the laws of physics. This applies to several topics such as the simulation of water or cloth movement, but especially to the shading of materials. While PBR is best known for creating photorealistic looks, various non-photorealistic looks can be achieved as well. To ensure creative workflows, it is crucial to provide an intuitive editing interface for users not being familiar with the underlying physical terminology. The Walt Disney Animation Studios played a leading role in establishing the creative principles of PBR in modern digital production workflows. Burley [Bu12] described a new material model, which compromises physical laws and the needs of artists for material design. Meanwhile, many other commonly used 3D rendering software, such as Maya, Blender, or the game engines Unity and Unreal Engine 4 (UE4), have adopted the PBR workflow. With software as the application suite Substance by Adobe, there even are tools purely dedicated to material appearance design. The study of material models is an active field of research. The topic addresses several interdisciplinary areas, such as “psychology, computer graphics, neuroscience [and] industrial design” [Dy17]. By now, a variety of models for describing

[1] Stuttgart Media University, Nobelstraße 10, 70569 Stuttgart, Germany; since October 2020 at Technische Universität Ilmenau, Ehrenbergstraße 29, 98693 Ilmenau, Germany, franca.bittner@mailbox.org

material appearance exists in terms of PBR. Yet, these models cannot be compared easily. Although there is already plenty of research on material models and their taxonomy, the quantity of material models can easily get confusing for the user, especially since every rendering tool uses its own parameter set. This becomes a problem as it is common to interchange 3D projects between different rendering tools in modern digital production pipelines, where each department may use its own preferred software [Gu20]. It cannot be assured that this exchange happens without loss of material description data. Certain parameters may not exist in another software or have a different impact on the rendered result, as they are included differently in the rendering equation. Since these dissimilarities are insufficiently studied, this study aims to compare the different material models in modern rendering software commonly used in digital productions. It is investigated how the material models in current PBR pipelines differ in terms of the intuitiveness of material design and the quality and range of reproducible materials.

2 Related Work

Guarnera et al. [Gu20] have dealt with the visual differences occurring when transferring parameter values from one parametric material model to another. The authors developed a genetic algorithm that finds a source parameter set matching the desired original parameters of a different material model. Two images of an identical object are compared: both rendered in the same scene, but with different material models. In an iterative process, two parental data sets are combined to generate new possibly fitting parameter sets. This way, the parameter sets of both material models are matched and evoke a nearly identical output image. While that work is more sophisticated than this paper, it does not examine material models commonly used in current digital production workflows in detail. This work is intended to fill this gap.

Burley [Bu12], Karis [Ka13], Lagarde [La11; La12; La14], Unity Technologies [UT14] and the Blender Foundation [BF17] provide material charts similar to the results in this work. However, those illustrations are only intended for the respective own rendering tool and do not provide any comparison among each other. Moreover, there are not material charts for every commonly used rendering tool. For example, the Arnold renderer only provides a few samples of individual material parameters, but no coherent overview over all material parameters [SA]. As reference images are actually in demand for testing different implementations and ensuring look consistency, this work aims to provide a rendering framework to generate such reference material charts across all examined renderers.

Another topic related to this work is the improvement of *material appearance design*. The report of Schmidt et al. [Sc14] deals with alternative ways of editing a virtual scene. The authors define the term "appearance design" and describe several concepts that connect lighting and material editing, as for instance editing the lighting scenario by clicking and dragging a specular highlight of an object [Sc14, pp. 2–5]. In accordance with this work, this study considers the appearance of a scene as an interplay of lighting and surface

materials, or global and local light transport respectively [Sc14, p. 2, 4, 6, 9]. Serrano et al. [Se16] developed an intuitive control space for editing captured bidirectional reflectance-distribution function (BRDF) data and proposed a list of fourteen attributes for material appearance [Se16, p. 6]. However, data-driven BRDFs are less relevant in the context of digital media productions, since in the commonly used tools the representation and editing of those data is not widely supported or feasible due to high memory requirements and expensive calculations. Gulbrandsen [Gu14] depicts an example for mapping unintuitive parameters of a physically plausible model to "artist friendly" [Gu14, pp. 64-65] parameters by decoupling the influences of two parameters on the appearance of the Fresnel curve. This paper serves as a favourable example for suiting the PBR workflow to the user when assessing the intuitiveness of material design in the chosen rendering tools.

3 Approach

3.1 Evaluation

Four renderers commonly used in the media industry are discussed: Arnold for Maya [Au20], Cycles and Eevee from Blender [BF20a], and the game engine Unreal Engine 4 (UE4) [EG20a]. Both, real-time and offline rendering pipelines are considered, representing different use cases like film production or game development. Arnold and Cycles are based on path tracing and thus considered as offline renderers. Although UE4 is capable of path tracing as well, only the real-time rasterization rendering pipeline is evaluated in this paper. Likewise, Eeve is a real-time rasterization renderer.

For the examination, several images depicting an object with different material properties in a certain lighting condition are generated. One parameter at a time is incremented with a specific step size to vary material appearance. A new image is rendered for each parameter change. This results in a row of images for each rendering tool, which are contrasted in a table, also referred to as *material chart*. The layout is inspired by previous related work [Bu12, p. 13][La11; La12; La14][UT14].

The script for this process was implemented individually for each rendering software in the scripting language Python (and additionally with Blueprints in UE4). The set of rendered images is then composed in a row for each renderer and plotted against the other rendering tools in a Jupyter Notebook. Moreover, difference images and scanline plots are generated. The source code is provided on the project website[2].

Four material parameters are examined: *Roughness*, *Metallic*, *Specular* and *Clearcoat*. While the latter is a rather advanced appearance feature, the other three parameters are the well-known basic parameters to influence materials in a PBR workflow. The used shader models are "Principled BSDF" in Blender, "Standard Surface" shader in Arnold

[2] `https://www.hdm-stuttgart.de/ifg/projects/2021/Bittner_2021_CMM`

and “Default Lit” and “Clear Coat” model in UE4. In some cases, the names of the parameters vary slightly among the tools. Here, the parameter with the more similar visual impact is selected. Arnold distinguishes between *specularRoughness* and *diffuseRoughness*. *specularRoughness* behaves analogous to *Roughness* in the other rendering tools, as both control the sharpness of the specular highlight. Also, *Metalness* in Arnold is used to evaluate the parameter *Metallic* in this paper, whereas *Coat* corresponds to *Clearcoat*.

Since a comprehensive quantitative investigation was beyond the scope of this paper, a qualitative comparison method is chosen. The images are rendered at a resolution of 500x500 pixels and compared by the author in three ways: by analysing and opposing the unchanged output images, by looking at the difference images between each parameter alteration step and by studying selected scanlines of an image.

3.2 Test environment

Four main aspects influence the final object appearance apart from material properties: camera, light, geometry, and the algorithm for emulating light transport [Sc14, p. 2]. As the latter differs in each rendering tool, certain visual differences will be inevitable. The implementation of the light transport is not manipulated, because this would not represent the general working environment of a 3D artist.

For this paper, two different lighting scenarios were chosen. The first one consists of a natural lighting situation, which is implemented using an environment map and image-based lighting (IBL). This setting addresses the natural human perception of material appearance, while at the same time it is an often-used technique in digital production workflows, e. g. in architecture design or virtual film production. IBL strongly depends on several factors such as the colour management of the texture inside the engine, the generated mipmaps, texture filtering, and the used method for unwrapping the texture. In contrast, the second lighting scenario is a very unnatural one, composed of only one directional light, to provide images that are generated with a more basic lighting feature. Directional lights are commonly controllable by two parameters, the power of the light source and the rotation or direction of light. This simplified lighting scenario is more likely to be implemented similarly in each software. Hence, the lighting scenario may be less error-prone with less causes for mismatch.

Four environment maps were chosen to evaluate different material appearance features. All high dynamic range images (HDRIs) are freely available and were taken from the website HDRIHaven[3][Za]. The chosen lighting scenarios range from high to medium dynamic range, as depicted in Table 1. Contrary to my initial expectations, the night scene “Moonless Golf” did not provide added value for recognizing the interplay of lighting and material appearance. The lighting scenario was still useful for determining sufficient lighting settings for the rendering process, as dark scenes are challenging due to the small amount of light.

[3] `https://hdrihaven.com/hdris/`

	Colorful Studio	Lebombo	Moonless Golf	Sunflowers
Dynamic Range	very high	medium	extremly high	extremly high
Exposure Values	15	9	23	23
Whitebalance	3700	4200	3942	6550

Tab. 1: Properties of used environment maps

The shape of an object can either be concave or convex and determines the observable material characteristics. For simplicity, a convex sphere is used. This has the advantage that important appearance features like the Fresnel effect can be observed towards the peripheral areas of the sphere. Also, the camera positioning is independent of the geometry of the object, as the 2D projection of a sphere appears to be a circle from any desired viewing angle. Other geometrical shapes are not examined, but may provide additional information about material appearances features. The resolution of a surface mesh is critical as well. A too low resolution of the polygonal mesh may result in artifacts like the *terminator problem* [WPO96, p. 22]. The distance of the camera to the sphere is set to 10m, while the focal length is set to 170mm. This was found to rule out perspective distortion effects. The sensor size is set to 36mm for width and height, resulting in a quadratic image.

4 Results

The results are described from two perspectives: First, the impact of a parameter within the same renderer allows for conclusions about the intuitiveness of material design. If the parameter causes different effects than expected, the descriptor may be classified as unintuitive. Secondly, the parameter impact is compared across all rendering tools to evaluate the impact of different material models on material appearance.

Two types of material charts depict the influence of a material parameter. The general material chart contrasts the rendered images unmodified. The difference image material charts oppose the difference image between two rendered images, where white areas remark unchanged areas, a red tone indicates an increase, and blue colouration a decrease in brightness. The colormap is normalized to the maximum negative and positive values.

The complete collection of figures and material charts is provided as supplementary material on the project website[4].

4.1 Roughness

The parameter *Roughness* has a great impact on the appearance of a material. Many characteristic features like specular highlights or reflections can only be observed if the

[4] https://www.hdm-stuttgart.de/ifg/projects/2021/Bittner_2021_CMM

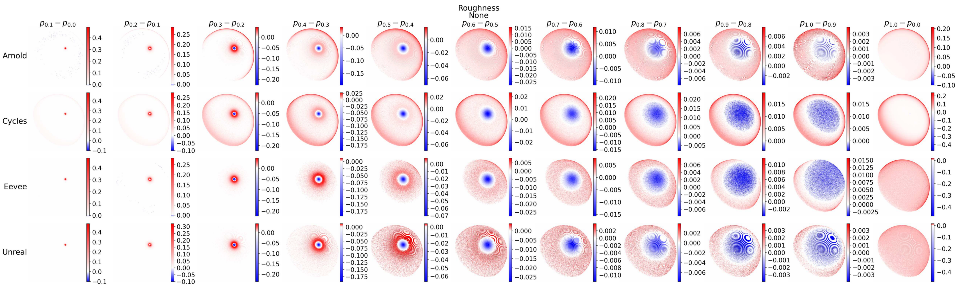


Fig. 1: Cross-renderer material difference chart for “Roughness” without environment map. The last image shows the difference image between value 0.0 and 1.0

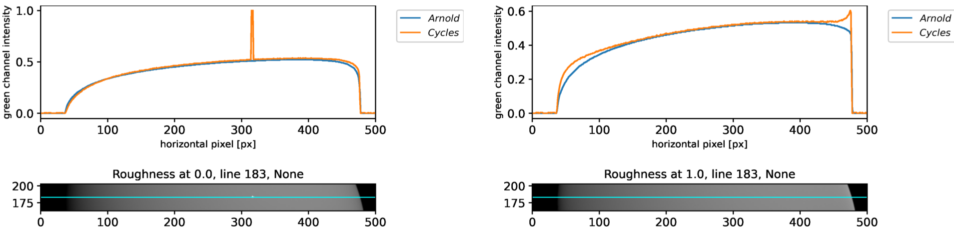


Fig. 2: Brightness plot for “Roughness” 0.0 (left) vs. 1.0 (right) in Arnold and Cycles. The image below shows the plotted image line 183 (cyan) in the context of the total image, rendered in Cycles.

surface is smooth. Lower values of *Roughness* correspond to smooth surfaces, while higher values correspond to a rougher surface. The behaviour of specular highlights in all rendering tools corresponds to the laws of physics for specular and diffuse reflection. Rough surfaces scatter light in many directions, resulting in a wider and less bright highlight, and vice versa. Hence, the general impact of the parameter *Roughness* behaves very intuitively.

When varying the parameter *Roughness*, the results are quite different across each renderer. The most striking disparity among the renderers is the handling of light at grazing angles and around the terminator. While both path tracers Arnold and Cycles show an increase in brightness towards grazing angles in Fig. 1, this appearance feature is not present in the two rasterizers. In a detailed look at the differences between Arnold and Cycles, Cycles considers retro-reflection, i. e. a higher reflectance at grazing angles than at normal incidence, whereas Arnold does not reproduce this effect. Instead, the maximum reflectance at grazing angles in Arnold seems to be limited to the diffuse colour at normal incidence (i. e. F_0). Fig. 2 contrasts the two renderers Arnold and Cycles for the *Roughness* values 0.0 and 1.0, which corresponds to an ideal smooth and purely rough surface. For an ideal rough surface, it is noticeable that Cycles returns higher brightness values at around 475 horizontal pixels, i. e. the edge of the sphere. The peak in brightness at grazing angles is even higher than the general diffuse colour of the sphere and might be interpreted as retro-reflection. Interestingly,

only Cycles appears to model this phenomenon correctly in the lighting situation with one directional light.

4.2 Metallic

The parameter *Metallic* influences whether a material is treated as conductor (i. e. metal) or dielectric (i. e. non-metal). This has a particular impact to the calculation of the Fresnel term and thus the specular reflectance of a material. Because conductors generally have a higher reflectance than dielectrics, a higher value of *Metallic* causes a more reflective material as shown in the rendering results in Fig. 3. The final appearance of a metallic material thus depends more on its environment than a dielectric material would. In general, the parameter *Metallic* causes expected intuitive results, since we are used to metallic surfaces being more reflective.

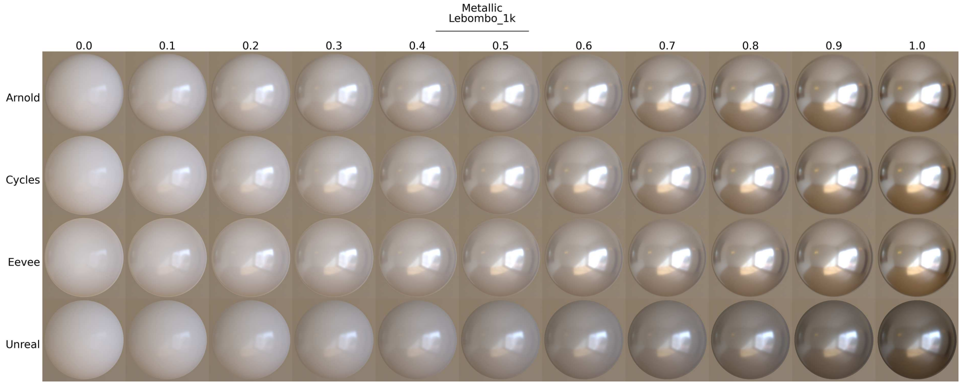


Fig. 3: Cross-renderer material chart for "Metallic" in environment "Lebombo"

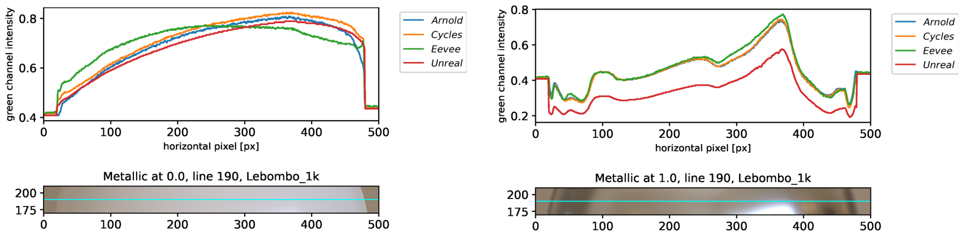


Fig. 4: Intensities for dielectric vs. metallic surface across all renderers with noticable drop-off in UE4. The image below shows the plotted image line 190 (cyan) in the context of the total image, rendered in Cycles.

The most noticeable difference for the parameter *Metallic* across all renderers was the significantly darker appearance of metallic materials in UE4. Fig. 3 shows the material chart for varying *Metallic* from 0.0 to 1.0 across the different renderer in the lighting scenario "Lebombo". The perceived colour of the sphere with a metallic surface in the last column

is remarkably darker in the results from Unreal compared with Arnold, Cycles and Eevee. To exclude the possibility that the phenomenon is caused by different light intensities of the environment map, Fig. 4 shows the intensity of the green channel in line 190. While Unreal has slightly lower intensity values in this lighting scenario than the other renderers, all rendering tools approximately start with the same intensity values. Nevertheless, there is a significant decrease in intensity of about 0.1 for metallic materials in UE4 compared with the other renderers in this environment. The other lighting settings "sunflowers" and "moonless_golf" produce similar results, only the environment map "colorful_studio" evoked consistent intensity values across all renderers. Unfortunately, I could not find a sufficient explanation for the deviating test results of Unreal in the mentioned light situations.

4.3 Specular

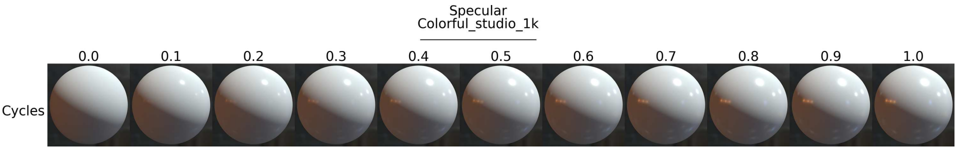


Fig. 5: Varying "Specular" in Cycles in environment "Colorful Studio"

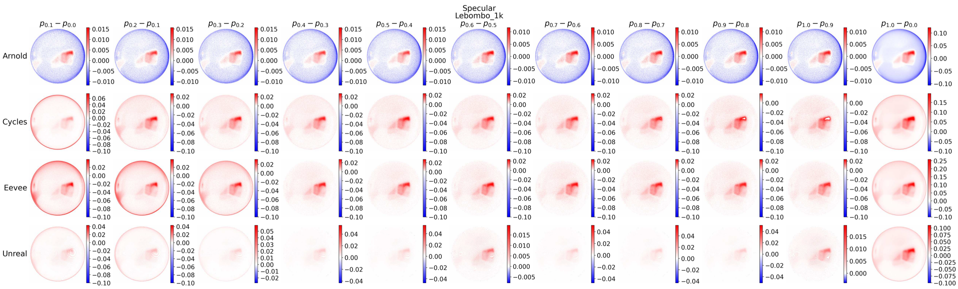


Fig. 6: Cross-renderer material difference chart for "Specular" in environment "Lebombo" . The last image shows the difference image between value 0.0 and 1.0.

The *Specular* value influences the intensity of all specular reflections (see Fig. 5). The parameter only has any impact if the material is non-metallic (i.e. dielectric). However, I claim that this parameter concept is not very intuitive and the term "specular" not self-explanatory. In fact, the reflection of pure metallic surfaces only consists of a specular component (i. e. no diffuse reflection), so it is misleading that the *Specular* parameter has no effect on metals.

Its visual effect can easily be mistaken for *Roughness*, however the parameter *Specular* does not affect the fuzziness of the specular highlight. Instead, it acts like a multiplier for the intensity of the highlight. According to the reference implementation of the "Autodesk Standard Surface" shader [Ge19, pp. l. 126–129], the *Specular* parameter indeed acts as a multiplier for the specular reflection intensity. The index of refraction (IOR) for calculating

the Fresnel term is exposed as a separate parameter in Arnold. In contrast, the *Specular* parameter in Blender and UE4 controls the IOR of the material directly. A conversion between *Specular*, F_0 and the IOR is provided by Burley [Bu15], based on the formulation for F_0 by Cook and Torrance [CT82], and adopted by the mentioned renderers [BF20b, pp. l. 101][EG20b, pp. ll. 76–97].

The physical term IOR is less intuitive than the parameter *Specular* [Bu15]. Hence, the material design in the renderers Cycles, Eevee and Unreal is more intuitive, as artists do not have to deal with setting the IOR. However, one could argue that the *Specular* parameter in Arnold allows for more traditional control over the amount of specular reflections. The specular highlight acts as separate layer and can be precisely blend in and out. While this might be less physically plausible, it allows for greater and easier artistic control, as this kind of interaction is commonly used in design software. The greatest issue in terms of intuitiveness is the non-consistent usage of the descriptor *Specular* across the examined tools. The parameters share the same descriptor but affect material features in a fundamental different manner. This semantic mismatch can easily get confusing for the user.

The different effects of *Specular* across all renderers is also observable in the material difference chart (see Fig. 6). While the image intensity at grazing angles linearly decreases in Arnold (i. e. less reflective), it increases non-linearly for the renderers UE4, Cycles and Eevee . It can be assumed that this is due to different calculations of the Fresnel term, since the grazing shadow of smooth surfaces can be derived from the Fresnel equations [Bu12, p. 6]. In the inner area of the sphere, all renderers behave linearly. Overall, the visual disparities among the renderers for *Specular* in the examined lighting situations are minimal and mostly noticable towards grazing angles.

4.4 Clearcoat

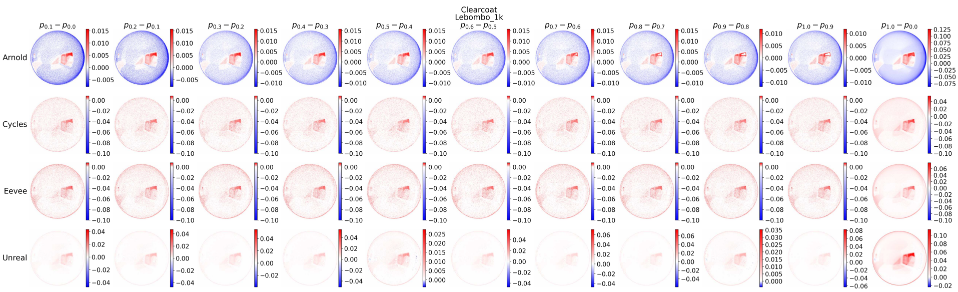


Fig. 7: Cross-renderer material difference chart for “Clearcoat” in environment “Lebombo”

The impact of *Clearcoat* is very subtle in the examined lighting situations and best observable with the renderer Arnold viewed in the lighting situation “Lebombo” or “Colorful Studio”. The effect can be mistaken for *Roughness*, as both influence the sharpness of the specular highlight. However, *Clearcoat* acts as a second layer on top of the base material. The

underlying blurry highlight typical for rough surfaces is still perceivable, although there is a specular layer on top of it. The intuitiveness of the parameter *Clearcoat* can only be assessed based on limited data. In my results, the effect is barely noticeable in the given lighting situations. Nevertheless, its descriptor is clear in its meaning and thus it is possible to define a material without seeing the effect directly in the viewport.

The difference image material charts in Fig. 7 revealed a deeper insight of the behaviour of this parameter. The overall impact of the parameter *Clearcoat* is greatest in Arnold and Unreal, while the results of Cycles and Eevee show almost no change in brightness. The increment of the parameter adds a grazing shadow in Arnold, while the other renderings show an increased reflectivity towards grazing angles instead. This again suggests that the Fresnel term for modelling the clearcoat layer is implemented differently in the material models. Due to the linearized parametrization, it can be assumed that the calculated Fresnel effect is faded in by linear interpolation, just like the impact of the parameter *Metallic*. Nevertheless, these are just assumptions based on little data and they should be understood with care.

5 Discussion

My findings show that the examined rendering software has mainly adopted the same underlying PBR workflow. The main differences occurr due to dissimilar methods for calculating light transport itself. For example, the renderers Arnold and Cycles rely on path tracing and generally produce more realistic results as the two rasterizers Eevee and Unreal. This was especially noticeable in the missing terminator in Eevee in the lighting scenarios “Colorful Studio” and “Lebombo”, but also in the lower visual quality of specular reflections in Eevee and Unreal, compared to Arnold and Cycles. Nevertheless, there are some minor disparities because of non-identical underlying material models, resulting in dissimilar intuitiveness of material editing and divergent scene appearance in each tool.

Although all listed disparities are subtle and mostly not perceptible when working within one rendering tool, they compound a seamless exchange of data between rendering tools among different departments in digital productions. The output renderings vary in the reproduction of special material features or might not be able to reproduce a material in another renderer at all, such as dielectric materials with tinted highlights.

6 Conclusion

This study showed that there is inconsistency among current rendering tools when it comes to capturing the material properties abstractly in a model. Every renderer has its own realisation, and it will likely remain a challenge to provide a seamless exchange between these pipelines for digital productions working with miscellaneous tools in different departments.

Nevertheless, the results of the examined rendering software are quite similar with some exceptions. The most significant visual disparities still arise from different rendering mechanisms rather than the underlying material model. In future works, a standardized material model may be developed to simplify the process of exchanging material data among renderers. Additionally, a material model adapted to the individual strengths of a renderer could be dedicated for working within a rendering tool.

References

[Au20] Autodesk: Maya, comp. software, version 2020, 2020.

[BF17] Blender Foundation: Principled BSDF — Blender Manual, material charts added by Alexander Gavrilov (https://developer.blender.org/rBM3686), July 2017, URL: https://docs.blender.org/manual/en/latest/render/shader_nodes/shader/principled.html, visited on: 08/09/2020.

[BF20a] Blender Foundation: Blender, comp. software, version 2.83.1, 2020.

[BF20b] Blender Foundation: Cycles - node_principled_bsdf.osl, Medium: Source Code, Mar. 2020, URL: https://github.com/blender/blender/blob/master/intern/cycles/kernel/shaders/node_principled_bsdf.osl, visited on: 08/07/2020.

[Bu12] Burley, B.: Physically Based Shading at Disney, Conference session updated version 3, Aug. 2012, URL: http://disney-animation.s3.amazonaws.com/uploads/production/publication_asset/48/asset/s2012_pbs_disney_brdf_notes_v3.pdf, visited on: 06/13/2020.

[Bu15] Burley, B.: Extending the Disney BRDF to a BSDF with Integrated Subsurface Scattering, course notes, Los Angeles, California, Aug. 2015, URL: https://blog.selfshadow.com/publications/s2015-shading-course/burley/s2015_pbs_disney_bsdf_notes.pdf, visited on: 08/22/2020.

[CT82] Cook, R. L.; Torrance, K. E.: A Reflectance Model for Computer Graphics. ACM Transactions on Graphics 1/1, pp. 7–24, Jan. 1982, ISSN: 0730-0301.

[Dy17] DyViTo Project, Oct. 2017, URL: https://dyvito.com/dyvitoproject/, visited on: 08/04/2020.

[EG20a] Epic Games: Unreal Engine 4, comp. software, version 4.25.0, May 2020.

[EG20b] Epic Games: Unreal Engine 4 - BRDF, Medium: Source Code for BRDF.ush, July 2020, URL: https://github.com/EpicGames/UnrealEngine, visited on: 08/07/2020.

[Ge19] Georgiev, I.; Portsmouth, J.; Andersson, Z.; Herubel, A.; King, A.; Ogaki, S.; Servant, F.: Standard Surface Shader, Medium: Source Code of reference implementation, Sept. 2019, URL: https://github.com/Autodesk/standard-surface, visited on: 08/07/2020.

[Gu14] Gulbrandsen, O.: Artist Friendly Metallic Fresnel. Journal of Computer Graphics Techniques 3/4, p. 9, 2014, ISSN: 2331-7418.

[Gu20] Guarnera, D.; Guarnera, G. C.; Toscani, M.; Glencross, M.; Li, B.; Hardeberg, J. Y.; Gegenfurtner, K. R.: Perceptually Validated Cross-Renderer Analytical BRDF Parameter Remapping. IEEE Transactions on Visualization and Computer Graphics 26/6, pp. 2258–2272, June 2020, ISSN: 1941-0506.

[Ka13] Karis, B.: Real Shading in Unreal Engine 4, Conference session updated version 2, July 2013, URL: https://blog.selfshadow.com/publications/s2013-shading-course/karis/s2013_pbs_epic_notes_v2.pdf, visited on: 08/12/2020.

[La11] Lagarde, S.: Feeding a physically based shading model, Aug. 2011, URL: https://seblagarde.wordpress.com/2011/08/17/feeding-a-physical-based-lighting-mode/, visited on: 08/12/2020.

[La12] Lagarde, S.: DONTNOD specular and glossiness chart, Apr. 2012, URL: https://seblagarde.wordpress.com/2012/04/30/dontnod-specular-and-glossiness-chart/, visited on: 08/12/2020.

[La14] Lagarde, S.: DONTNOD Physically based rendering chart for Unreal Engine 4, Apr. 2014, URL: https://seblagarde.wordpress.com/2014/04/14/dontnod-physically-based-rendering-chart-for-unreal-engine-4/, visited on: 08/12/2020.

[SA] Solid Angle S.L.: Standard Surface - Arnold for Maya User Guide - Arnold Renderer, URL: https://docs.arnoldrenderer.com/display/A5AFMUG/Standard+Surface, visited on: 08/13/2020.

[Sc14] Schmidt, T.-W.; Pellacini, F.; Nowrouzezahrai, D.; Jarosz, W.; Dachsbacher, C.: State of the Art in Artistic Editing of Appearance, Lighting, and Material. In (Lefebvre, S.; Spagnuolo, M., eds.): Eurographics 2014 - State of the Art Reports. The Eurographics Association, 2014.

[Se16] Serrano, A.; Gutierrez, D.; Myszkowski, K.; Seidel, H.-P.; Masia, B.: An Intuitive Control Space for Material Appearance. ACM Trans. Graph. 35/6, Nov. 2016, ISSN: 0730-0301, URL: https://doi.org/10.1145/2980179.2980242.

[UT14] Unity Technologies: Unity - Manual: Material charts, 2014, URL: https://docs.unity3d.com/Manual/StandardShaderMaterialCharts.html, visited on: 08/12/2020.

[WPO96] Woo, A.; Pearce, A.; Ouellette, M.: It's really not a rendering bug, you see. IEEE Computer Graphics and Applications 16/5, pp. 21–25, Sept. 1996, ISSN: 1558-1756.

[Za] Zaal, G.: HDRI Haven, URL: https://hdrihaven.com/hdris/, visited on: 08/11/2020.

Business Informatics

Data-based Transparency and Leadership in Small and Medium-sized Enterprises

Carmen Mayer [1]

Abstract: Based on the increasing usage of Information Systems (IS), the amount of employee-specific data in companies is rising. As this data is more often used for leadership, referred to as data-based leadership, the question about complete transparency and its consequences in companies needs consideration. This work therefore, aims to analyze the amount of gathered employee-specific data, the resulting data-based leadership, and the exercised control and transparency in small and medium enterprises (SME) which have limited experience in using digital leadership approaches. The applied case study provides qualitative insights into these aspects. This case study is based on five selected SMEs from different industries. With my study, I enhance control theory and derive practical recommendations for a sustainable handling of employee-data for leadership.

Keywords: Data-based Transparency, Leadership, Control Theory, SMEs, Information Systems.

1 Introduction

Due to the increasing availability and exploitation of digital technologies in companies, the amount of produced data is growing. This enhances digital innovations and consequently transforms businesses [WH20]. Particularly, the appearance of human resource information systems (HRIS) illustrates the transformation of employee management into a data-driven direction [GH20]. These systems are defined as cross-sectional systems used for workforce planning and performance management [La18].

The increasing transparency that results from data-driven human resource management (HRM) creates a higher level of control and provides insightful learning about the employee from a manager's perspective [Be17]. On the employee level, data-based transparency, meaning transparency that is generated by rising amounts of employee data, favor employee empowerment [MMH15]. However, increased data-based transparency also entails significant risks for managers and employees [Be17]. These include, for instance, a loss of trust and growing pressure on the employee level and, on the management level, problems due to the complexity of data evaluation [MMH15]. It is therefore necessary to balance the opportunities and risks of increased transparency.

Currently, the number of publications in the area of data-based leadership is growing constantly [CBZ19]. Prior contributions investigate people analytics, workforce analytics, talent analytics, HR analytics and HRIS which are mostly used as synonyms [Hu18]. Qualitative works focus mainly on advantages and disadvantages of transparent leadership

[1] Ludwig-Maximilians-Universität München, Fakultät für Betriebswirtschaft, Institut für Wirtschaftsinformatik und Neue Medien, Ludwigstraße 23, 80539 München, carmen.nadine.mayer@gmail.com

without taking into account the size of the analyzed firm and resulting characteristics [VVV19]. Moreover, researchers state that “there is still a lack of theorization about the impact of technology on leadership” [CBZ19]. The focus of this paper is on data-based transparency in small and medium- sized enterprises (SMEs) as they usually have fewer financial resources and less hierarchies than large enterprises [ID19]. Thus, the impact of data on leadership is even harder to estimate for these firms. In this paper, I will therefore address the research gap of the availability and use of data in leadership and the resulting data-based transparency in SMEs. Thus, I pose the following research question:

RQ1: Which data can be used in the context of data-based leadership in SMEs?
RQ2: How are data-based leadership, control and transparency related in SMEs?

To answer these research questions, the concepts of data-based transparency, leadership and control theory are explained in the theoretical background. Based on this theoretical background, a case study research, with ten semi-structured interviews from various sectors, is conducted. These findings are then analyzed and discussed. Finally, the limitations of the work are outlined and implications for theory and practice are provided.

2 Theoretical Foundation

2.1 Information Systems (IS), data-driven Leadership and Transparency

In companies in general, data is created, processed and displayed in different IS. For the aim of this paper, only the IS in which employee-specific data are gathered and evaluated are considered. Hence the following section outlines three operational and one strategic IS. Knowledge Management Systems (KMS) pertain to cross-sectional systems. Their function is the facilitation of access to and storage of tacit and explicit knowledge [AM17]. Besides KMS, Office Management Systems (OMS) are also classified as cross-sectional systems. OMS support individual office work, such as word processing, spreadsheet calculation, email services or direct messaging systems which support communication and collaboration in enterprises [AM17]. The last systems which are allocated in the area of cross-sectional systems are HRIS which are used for workforce planning and development, performance management and human resource development [La18]. Strategic IS, namely Management Information Systems (MIS) are positioned one level above operational IS. MIS supply information to managers and therefore support their data-driven decision-making [HMN15]. Examples of MIS are planning and control systems, cross-functional software and corporate planning systems [Me17]. Overall, if employee data are used for leadership, this is referred to as data-based leadership. Data-based leadership may result in a higher level of transparency which is recently being analyzed more frequently in literature. In a broad sense, synonyms to transparency are openness, freedom of information and clarity [Be17]. According to Hofmann, transparency is generally defined as a "state of more or less clarity of structures and processes [... where] information and communication are the medium [...] with which this

state can be created and maintained" [Ho07]. Further, transparency can be analyzed across and within companies, whereby this paper focusses on the latter. Persons involved in the transparency process are the observed and the observer. The observed are the employees who send information, and the ones who receive and process it. Whereas, the observer is the manager who has access to this information [Be17]. In sum, data-based transparency is an observability of structures and processes by managers and employees that is created with data from HRIS, and that is used for data-based leadership. As transparency can either lead to increased efficiency, empowerment and participation or to privacy concerns and mistrust [Be17], enterprises have to find their optimal degree of transparency by balancing overall transparency and privacy [MMH15].

2.2 Control Theory

Control in companies is defined as any attempt to align the individual behavior of employees with the firm's goals [Wi19]. Control theory is a well-established theory which originates from sociology and is frequently applied in the context of IS to explain situations where control is used in organizations [CW18]. Thus, it combines IS and data-based transparency, as well as leadership in SMEs and serves as a valuable basis for this work. Subsequently, five dimensions of control theory are described.

The first dimension, the *control relationship*, defines the parties involved in the control process, the controlled (the employee) and the controller (the manager) [CW18]. The second dimension examines the *control environment*, which consists of the corporate strategy, the organizational structure, the process characteristics, the corporate culture and the members of the organization [CBG16]. The third dimension, the *control mode* is divided in formal and informal control. For formal control we can further differentiate between behavioral and outcome control. In behavioral control, the controller implements rules and influences the process and the attitude of the controlled through observation [CS03]. Outcome control targets the results rather than the process of a project [CS03]. Informal control can also be divided into two dimensions: self-control and clan control. Self-control occurs if the controlled determine their goals and behavior through formal control mechanisms [CS03]. In contrast, clan control relies on shared norms and values that reduce differences between the controlled and the controller [CS03]. A fourth dimension is referred to as *control style* and can be subdivided in coercive control and enabling control. The coercive style is hierarchical and is exerted unilaterally by the manager, whereas empowering control is collaborative and bidirectional [Wi19]. The fifth dimension is the *degree of control* which can either be relaxed or strict [CW18].

In order to analyze data-based transparency in leadership, the classification of IS serves as a structuring method to cluster different sources of employee data and possible data points. Taking the definition of data-based leadership as a basis, the situation in SMEs can be examined. The concept of control theory with its descriptive dimensions is used to determine the exerted control in SMEs.

3 Method: A Case-Study Approach

In this work an exploratory case study research has been conducted, based on Yin's ideas [Yi14]. The method of the case study research has been applied in this context because of its possibility to precisely map a current complex social phenomenon in a holistic and real context. Furthermore, the method aims to provide insights into a largely unexplored area where the boundaries between the object of study and the context are not clearly evident. To provide insights into different company sizes, cultures and industries, the multiple case study research approach has been utilized [Yi14]. To ensure quality, the approach follows the guidelines by Yin [Yi14].

3.1 Data collection

Subsequent to the definition of the research design, a semi-structured interview guideline has been formulated. Simultaneously, suitable SMEs have been selected based on the following criteria. The interviewed SMEs have to be classified as an SME under article 267 (2) HGB [Fl17]. According to this article, quantitative characteristics of SMEs are a balance sheet of less than 20 million euros per year, a turnover of less than 40 million euros per year, and fewer than 250 employees on average per year. Furthermore, SMEs have been sampled according to their industry, location in Rhineland-Palatinate and Baden-Württemberg in Germany, and private ownership. Thus, 15 suitable SMEs have been selected and contacted with a standardized e-mail. From all contacted SMEs, 8 responses were received. A total of three pre-tests were carried out. Finally, interviews with 7 SMEs were conducted. The interviews were held between 2020-09-21 and 2020-10-19. In order to gather insights on the management and employee perspective in each SME, one member of the management or head of HR and one employee from a related division was interviewed for one 30-minute telephone or video-call. As with two firms only one interview with the managing director could be arranged, these interviews were not further analyzed due to missing data from an employee. For the analysis, the firm's websites were also used as sources. The following table 1 shows the final selection of SMEs. To ensure anonymity, the number of employees is listed in increments of fifty, and total assets and sales are rounded up or down.

	Bakery	Furniture	Car Trade	Wholesale	Pharma
Employees	200-250	150-200	100-150	150-200	<50
Balance sheet total (2018) in mio €	10	10	20	10	5
Revenue in mio €	15	20	40	30	30
Industry	Confection-ery and bakery	Office furniture	Car trade sector	Wholesale and retail	Pharma-ceutical industry

Table 1: Overview of the cases

3.2 Data analysis

Subsequent to the data collection, the interviews were transcribed and later coded using the software ATLAS.ti following Miles and Huberman [MH94]. The coding scheme constitutes of 15 keys, which were clustered accosting to the theoretical foundation into corresponding topics (cf. table 2).

Topic	Code
Characteristics of SMEs	Characteristics of SMEs
Information Systems and resulting employee-specific data	Knowledge Management Systems Human Resource Information Systems Office Management Systems Management Information Systems Employee-specific data from IS
Utilization of employee-specific data for management	Data-based leadership
Degree of Transparency in the SME	Degree of Transparency
Control Theory	Control Relationship Control Environment Control Mode Control Style Degree of Control

Table 2: Coding Scheme

4 Findings

In a first step, the interviewed SMEs are analyzed following the within-case analysis after Yin [Yi14]. Therefore, four main topics are examined, the particularities of the surveyed SME, their data form IS and data-based leadership, control and the degree of transparency.

4.1 Results Within-Case Analysis

Bakery: In the 1991 founded bakery, data that is used for leadership primarily derives from their workforce planning software and enterprise resource planning (ERP) system. In addition, data occurs in MS Word and Excel files which are used for employee performance and knowledge management as well as internal communication. In sum, this firm primarily collects master data about employees and operational company-related employee data. Thus, data-based leadership is sparsely developed.

Their control environment is characterized by structured processes, flat hierarchies and strong bilateral trust between employees and managers. Notably, the manager mentions that the *"assessed openness to share data is 50/50"* (Bmanager). With regard to the control

modes, the behavioral control takes place with dashboards about absences of employees: *"these are just performance things, like the amount of working hours.* (Bmanager). Moreover, they apply outcome control with performance-based bonuses for managers. Self-control takes place through jointly developed goals in annual meetings, and clan control with in-depth trainings of employees: *"when new staff is employed, they are trained with onboarding modules"* (Bemployee). The control style tends to be authoritarian, and the degree of control is moderate to low in comparison with other researched SMEs. Thus, the observed control is moderate to low.

Overall, the transparency is one-directional as the interviewed employee has no knowledge about it. Further, the potential of bidirectional transparency is not seen by the manager: *"if they had transparency, I do not see the added value"* (Bmanager). The degree of transparency from the manager about the employee is low, and vice versa very low.

Furniture firm: In the furniture firm, founded in 1893, the usage of data for leadership varies largely across its functional areas. The agile departments work with a task management system (TMS), which is used for workforce planning and team performance measurement. Employee-specific data thereby results from task distribution and status. In all other departments, the manager mentioned that MS Excel and Word files are used as HRIS. Further, a time recording software tracks master data about employees. In sum, the manager states that they "*have certain proliferations of software"* (Fmanager). Thus, data-based leadership in agile departments is high, in the other departments it is low to medium.

Their control environment is two-folded with strict hierarchy levels and agile departments. According to the employee, the corporate culture is open in agile organized departments: *"it results from the method [Scrum] that employees are transparent"* (Femployee). In contrast, the culture in non-agile departments is non-transparent. Regarding the control mode, behavioral control takes place through measuring key performance indicators (KPIs). Outcome control occurs through a strong focus on results as *"at the end of the day, what counts is the result*" (Femployee). Self-monitoring in agile teams takes place in their TMS. The control style is empowering in agile divisions and authoritarian in non-agile departments. The degree of control is moderate. In sum, the perceived control is moderate.

When assessing the degree of transparency, it has also a two-sided manner. In hierarchically organized divisions, data-based transparency about the employee is created by the measurement of KPIs as *"there are certain key figures, and we have an IS support therefore"* (Fmanager). However, the extent of transparency about the employee is in comparison low and one-directional. In agile departments transparency between the manager and the employee is bidirectional and high as *"data in the TMS is published even beyond the team."* (Fmanager). Thus, transparency between the manager and the employee is moderate and vice versa rather low.

Car trade: The family-owned and 1970-founded car trade has implemented a workforce planning software with an integrated HR development tool whose employee-specific data are used for data-based leadership. Besides master data about the employee, accomplished trainings and training propositions are stored. In addition, they use a performance

management system and a MIS in which employee-centric data are evaluated. Overall, data-based leadership is above average, thus medium to high.

The control environment is characterized by flat hierarchies and highly optimized work processes due to the managers personal characteristics: *"I am a process-freak"* (Cmanager). The company pursues a culture of trust with regard to collected data and according to the manager *"(we have) a performance oriented culture"* (Cmanager). The control mode is characterized by behavioral control through established and controlled processes, by outcome control through monthly performance evaluations, and informal self-control through the usage of these evaluations for self-optimization. In addition, the control style tends to be authoritarian and the degree of control through monthly evaluations tends to be strict as *"the KPIs are monitored, we are doing this very intensively"* (Cmanager). Overall, the perceived control is moderate to high.

The degree of transparency can be observed by their workforce development system, in which employee-related evaluations are conducted and accessible to the manager and the employee. However, the interviewed employee mentions that she has *"honestly never asked herself what is basically processed or how is it used"* (Cemployee), in contrast to the manager. Therefore, the overall transparency level between the manager and the employee is medium to high, but between the employee and the manager it is medium.

Wholesale: Founded in 1880, the wholesale uses a HRM, in which modules for workforce development are integrated and master data about the employee is collected. In addition, they implemented an ERP system, which also serves as a KMS, communication and project management system where performance data about the employee is compiled. Therefore, data-based leadership is medium compared to the other analyzed SMEs.

The control environment is characterized by flattening hierarchies and *"[they] have a very open culture [...] because the employees have been with the company for a very long time"* (Wemployee). In addition, a work council ensures security against misuse of data by executives. With regard to the control modes, the boundary between behavioral control and outcome control is blurred, as employee-specific KPIs in the sales department are calculated daily and included in staff appraisals in case of underperformance: *"I can see who my strong sellers are."* (Wmanager). Self-control is partially feasible in their ERP system as "*the employee has an overview of his processes in the ERP system"* (Wemployee) and clan control is implemented by shared norms, values and experience due to the long company affiliation of employees. The overall control style is authoritarian and the level of control rather strict. In sum, the control is moderate to high.

Data-based transparency is achieved by the mutual access and usage of their HRM system and the access to employee-specific KPIs in their ERP system as *"the employee can make an evaluation in the ERP system and can see directly: I have received so many orders in total today"* (Wmanager). These KPIs are also visible for managers and colleagues. Unlike their HRM and ERP system, their internal skill database and workforce planning software is only accessible by the management. Overall, the degree of transparency from the manager to the employee is medium and in reverse low to medium.

Pharma: In the pharmaceutical firm founded in 2011 a TMS is used as a performance management, communication, project planning and KMS for data-based leadership. Gathered data is evaluated at the team level. In addition, the MS Planner serves as a project management and KMS, where employee-specific data are processed. Particularly for the laboratory employee-specific capacity and performance metrics are collected and analyzed. In general, the use of data for leadership is practiced intensively, but with the primary goal of process optimization as *"the data from laboratory evaluation are not used for [performance monitoring] but process improvement"* (Pemployee).

The control environment is characterized by flat organizational structures with agile and cross-functional teams. The corporate culture is transparent "*but the employees also do not know it differently"* (Pmanager). Furthermore, employees stand out due to their high level of personal responsibility. The control mode at P contains all forms of control: behavioral control, which is primarily applied in the laboratory with KPIs for performance. Outcome control takes place in staff appraisals with target agreements in the employment phase. Self-monitoring is ensured by access to all measured KPIs and *"everyone can see what his/ her priorities are"* (Pmanager). Clan control is observed through structured and intensive training of new employees. Thus, the control is moderate.

The degree of transparency is very high, as all analyzed data can be viewed by anyone in the company: *"All our evaluations are public in the company"* (Pemployee) and "*the aspiration of our executives is basically that they say transparency is good and we want transparency*" (Pemployee). Therefore, data-based transparency from both the manager to the employee and the employee to the manager is very high.

4.2 Results cross-case analysis

Data from IS and leadership: Overall, the analyzed companies have implemented numerous different software solutions. Thereby, IS often have several functions, e.g. TMS at the furniture firm. Also, many SMEs are using MS Word and Excel files instead of complex HRIS, e.g. Bakery, Furniture and Pharma. Similar to the heterogeneous IS, the employee-related data also varies greatly across SMEs and their departments. Overall, all SMEs record master data about employees and many use an ERP system where data, such as employee-specific sales figures, are tracked. In agile organized SMEs, data on current tasks and their progress is collected via TMS. Many SMEs also have documents for personnel development. On average, employees listed less IS where their data is gathered than their managers and they do not seem to be fully aware of their data traces. In summary, the five SMEs implemented IS for HRM and leadership, but their intensity of use and their purpose for management differ significantly. For instance, the bakery sparsely uses data from IS for leadership, whereas the pharmaceutical firm and the car trade make more extensive use of them.

Dimensions of control: Similar to the data-based leadership, the extent of control varies greatly. In the bakery, for example, employees are little controlled by their superiors, at the furniture firm, the pharmaceutical firm and the wholesale moderately, and at the car

trade very often. In figure 1 the relationship between data-based leadership and control is illustrated. Thereby a tendency can be observed between a low level of data-based leadership and low level of control and vice versa. One exception is the pharmaceutical firm, where the extent of data-based management is very high, but employee control is only moderate.

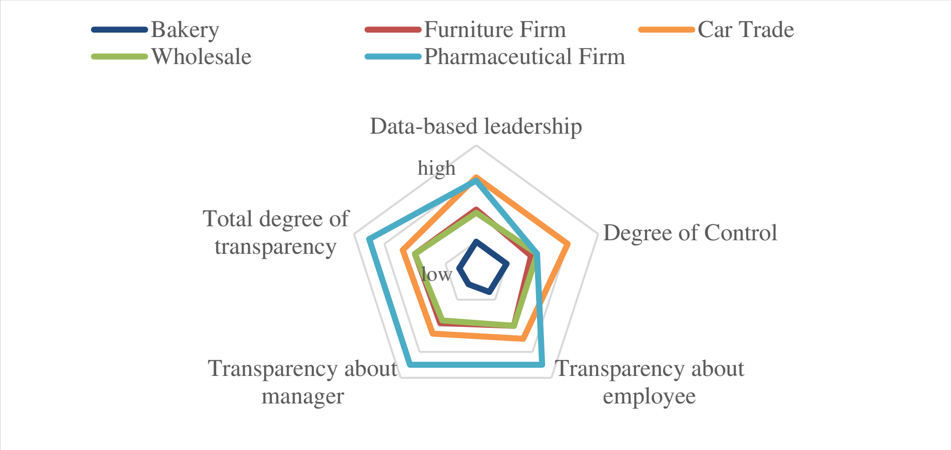


Figure. 1: Results of the cross-case analysis

Degree of transparency: Regarding the transparency between the employee and the manager, with the exception of the pharma firm, data-based transparency of the manager about the employee is always higher than vice versa. Thus, we observe information asymmetries between both stakeholder groups. Moreover, when transparency in general is increased, data-based management is also applied more frequently. With regard to data-based transparency and the degree of control, it can be seen that for the pharmaceutical firm, for instance, the degree of control is moderate, but transparency is very high. For the car trade, control is very pronounced, but transparency is only moderate. Overall, however, there is a tendency that a higher transparency is related to a higher degree of control.

5 Discussion

Addressing RQ 1 on data that can be used in the context of data-based leadership in SMEs, the following section discusses the results from the case-study research and existing literature in this field. Taking a look at prior works, Taylor and Taylor [TT14] also confirm outdated and partially improvised IT infrastructure in SMEs (e.g. Bakery, Furniture, Wholesale). Their listed reasons for this infrastructure, informal information and communication paths, flat hierarchies, scarcity of resources and limited financial capabilities can also be affirmed (e.g. Furniture). Regarding the OMS, the assumption by Lindner et al. [Li19] about the frequent use of e-mail, video conferencing and office

applications can be supported with this study. Overall, in all surveyed SMEs, different IS were implemented to different extents, which is also reflected in the literature. Reasons for this heterogeneity might be diverse industries and differences in the number of employees, turnover and IT expertise. Inevitably, the employee-centric data required for data-based management also varies greatly. This case study research therefore shows that the limited presence of IS restricts the extent of data-based leadership.

In order to discuss RQ2, the following section analyzes the extent of control and transparency in the surveyed SMEs without a strong literal support due to the scarcity of publications in this research field. In sum, the extent of control with its five dimensions varies greatly across the surveyed SMEs. These differences can be explained by diverse leadership styles and corporate cultures. However, it is noticeable that in almost all companies, the transparency of the manager about the employee is significantly higher than vice versa. One reason for this might be the classic understanding of leadership, according to which transparency about the employee is necessary for the purpose of control, but not vice versa. In the agile-structured pharmaceutical firm, a new leadership understanding can be observed where both the manager and employee is transparent in his or her actions. In summary, data-based leadership, control, and transparency is low in all areas at the bakery, moderate at the furniture firm and the wholesale, and moderate to high at the pharmaceutical firm, and the car trade.

6 Conclusion

6.1 Theoretical and Practical Implications

Due to the research gap on the role of data for leadership and transparency in SMEs, this paper aims to provide insights in this field with an explorative case study research. This works holds multiple theoretical contributions. First, I shed light on the very heterogeneous IT landscapes and correspondingly diverse employee-related data from IS. Thus, I deepen the understanding on technical framing conditions in SMEs and their chances for digital leadership approaches. Second, I connect the concepts of data-based management, the extent of control, and the degree of transparency in this study. Thereby, a positive relation between data-based management, the degree of control, and transparency is derived from gathered data. Third, I extent control theory by taking the aspect of company sizes as a crucial framing condition into account. From a practice-oriented point of view, this work can serve as a guideline on how to integrate employee-data in leadership with a special focus on SMEs. Enabling control styles in contrast to coercive, hierarchical ones and offering transparency to employees to establish sustainable novel leadership concepts at these firms can be derived as practical guidelines.

6.2 Limitations and Outlook

Limitations of this research result from the qualitative research design. As no quantitative data was collected, only tendencies and no generalizations can be deduced. Furthermore, the scope of the study is narrow due to its goal to provide insights into a largely unexplored field. But no holistic picture of data-based transparency and its influencing factors can be drawn. Another limitation of the applicability of these results is the restricted insights of the surveyed employees, as in each SME only one has been interviewed. Therefore, I suggest the following research avenue to overcome the mentioned concerns: First, further clarification about accruing data in IS in SMEs for employees as well as managers is needed. Second, the link between characteristics of SMEs and data-based leadership, control, and transparency should be analyzed more deeply, either by conducting case studies with more companies or by investigating the topic with a quantitative research approach.

6.3 Acknowledgement

Many thanks to Maren Gierlich-Joas for her support for the conception of this paper.

7 References

[AM17] Abts, D.; Mülder, W.: Grundkurs Wirtschaftsinformatik - Eine kompakte und praxisorientierte Einführung. Springer Vieweg, Wiesbaden, 2017.

[Be17] Bernstein, E.: Making Transparency transparent: The evolution of observation in management theory. In: Harvard Business Review, pp. 217-266, 2017.

[CBG16] Cram, W.; Brohman, M. K.; Gallupe, R.: Information Systems Control: A Review and Framework for Emerging Information Systems Processes. In: Journal of AIS, pp. 216-266, 2016.

[CBZ19] Cortellazzo, L.; Bruni, E.; Zampieri, R.: The Role of Leadership in a Digitalized World: A Review. In: Frontiers in Psychology, pp. 1-21, 2019.

[CS03] Choudhury, V.; Sabherwal, R.: Portfolios of Control in Outsourced Software Development Projects. In: Information Systems Research, pp. 291-314, 2003.

[CW18] Cram, W. A.; Wiener, M.: Perceptions of control legitimacy in information systems development. In: Information Technology & People, pp. 712-740, 2018.

[Fl17] Fleischer, H.: Handelsgesetzbuch PublizitätsG HandelsregisterV, 61. Auflage. Beck-Texte im dtv, München, 2017.

[GH20] Gierlich, M.; Hess, T.: Towards an Understanding of Data's Influence on Leadership. In: Proceedings of the 15th International Conference on Wirtschaftsinformatik, Potsdam, 2020.

[HMN15] Hansen, H. R.; Mendling, J.; Neumann, G.: Wirtschaftsinformatik Grundlagen und Anwendungen. De Gruyter, Berlin, München, Boston, 2015.

[Ho07] Hofmann, Y. E.: Transparenz in Unternehmen. In: Zeitschrift für Management, pp. 6-27, 2007.

[Hu18] Huselid, M. A.: The science and practice of workforce analytics: Introduction to the HRM special issue. In: Human Resource Management, pp. 679-684, 2018.

[ID19] Ihlau, S.; und Duscha, H.: Besonderheiten bei der Bewertung von KMU-Planungsplausibilisierung, Steuern, Kapitalisierung. Springer Gabler, Wiesbaden, 2019.

[La18] Lackes, R.; Siepermann, M.; Engelhard, J.; Wichert, J.: Personalinformationssystem, https://wirtschaftslexikon.gabler.de/definition/personalinformationssyst em-44178/version-267495, accessed: 09/08/2020.

[Li19] Lindner, D.: KMU im digitalen Wandel, Ergebnisse empirischer Studien zu Arbeit, Führung und Organisation. Springer Gabler, Wiesbaden, 2019.

[Me17] Mertens, P.; Bodendorf, F.; König, W.; Schumann, M.; Hess, T.; Buxmann, P.: Grundzüge der Wirtschaftsinformatik. Springer Gabler, Berlin, 2017.

[MH94] Miles, M. B.; Huberman, A. M.: Qualitative data analysis: An expanded sourcebook. Sage Publications, Thousand Oaks, London, New Delhi, 1994.

[MMH15] Morlok, T.; Matt, C.; Hess, T.: Privatheit und Datenflut in der neuen Arbeitswelt – Chancen und Risiken einer erhöhten Transparenz. In: Forum Privatheit und selbstbestimmtes Leben in der digitalen Welt, Peter Zoche, Regina Ammicht Quinn, Marit Hansen, Jessica Heesen, Thomas Hess, Jörn Lamla, Christian Matt, Alexander Roßnagel, Sabine Trepte, Michael Waidner, pp. 1-48, 2015.

[TT14] Taylor, A.; Taylor, M.: Factors influencing effective implementation of performance measurement systems in small and medium-sized enterprises and large firms. In: International Journal of Production Research, pp. 847-866, 2014.

[VVV19] Valentinov, V.; Verschraegen, G.; Van Assche, K.: The limits of transparency: A systems theory view. In: Systems Research and Behavioural Science, pp. 289-300, 2019.

[Wi19] Wiener, M.; Mähring, M.; Remus, U.; Saunders, C.; Cram, W. A.: Moving IS Project Control Research into the Digital Era: The “Why” of Control and the Concept of Control Purpose. In: Information Systems Research, pp. 1387-1401, 2019.

[WH20] Wiesböck, F.; Hess, T.: Digital innovations. In: Electronic Markets, pp. 75-86, 2020.

[Yi14] Yin, R. K.: Case Study Research: Design and Methods. SAGE Publications, Los Angeles, London, New Delhi, Singapore, Washington DC, 2014.

IT-Qualitätsmanagement im Rahmen des Informationsmanagements – Eine State-of-the-Art Betrachtung

Benjamin Zenth und Majeed Malik

Abstract: Die Anzahl an zu verarbeitenden Unternehmensdaten steigt stetig an, in diesem Kontext stellt das Informationsmanagement eine zentrale Disziplin dar. Dem IT-Qualitätsmanagement kommt hierbei eine wichtige strategische Rolle zu, da es die Qualitätssicherung der einzelnen Teilbereiche des Informationsmanagements zum Ziel hat. Obwohl dieses folglich eine zentrale Managementaufgabe darstellt, fehlt es einer aktuellen Betrachtung zum Stand der Wissenschaft. Mit dem vorliegenden Beitrag schließen wir diese Forschungslücke und zeigen den aktuellen Stand der Wissenschaft zum IT-Qualitätsmanagement im Rahmen des Informationsmanagements und decken möglichen Forschungsbedarf auf. Hierfür wurden 38 Teilthemen zum IT-Qualitätsmanagement im Rahmen des Informationsmanagements definiert und je Teilthema eine Literaturanalyse im Zeitraum 2016 bis 2021 durchgeführt. Dabei wurde je Teilthema der aktuelle Stand der Wissenschaft aufgezeigt und eine Einordnung hinsichtlich des zukünftigen Forschungsbedarfs vorgenommen. Hierbei konnte in einem Teilthema ein hoher, in 20 ein mittlerer und in 16 ein niedriger Forschungsbedarf identifiziert werden.

Keywords: Informationsmanagement, IT-Qualitätsmanagement, State-of-the-Art, Literaturanalyse, Forschungsbedarf

1 Einleitung

Die Anzahl an zu verarbeitenden Unternehmensdaten ist nach [Re20] in den vergangenen Jahren exponentiell gestiegen. Als Ursache hierfür nennt [Re20] unter anderem die zunehmende Digitalisierung und unternehmensübergreifende Kommunikation, welche neue Herausforderungen an die Informationserstellung, -verwaltung, -verteilung und -archivierung definieren. In diesem Kontext nennt [Ti20 S.226f] das Informationsmanagement als zentrale Disziplin eines Unternehmens, welches nach [Kr15 S.107] „im Hinblick auf die Unternehmensziele [den] bestmöglichen Einsatz der Ressource Information [zum Ziel hat]".

Die Autoren [e14] benennen das IT-Qualitätsmanagement als eine wichtige (strategische) Teildisziplin des Informationsmanagements, welches die Qualitätssicherung der einzelnen Bereiche des Informationsmanagements zum Ziel hat. Durch das von [Kr15 S.107] beschriebene Modell des Informationsmanagements wird dessen Vielfältigkeit ersichtlich, weshalb es einer individuellen Betrachtung dieser Teilbereiche im Hinblick auf den aktuellen Stand der Wissenschaft zum IT-Qualitätsmanagement bedarf.

Das Ziel des vorliegenden Beitrags ist es, mittels einer Literaturanalyse den aktuellen Stand der Wissenschaft (engl. state-of-the-art) zum IT-Qualitätsmanagement im Rahmen des Informationsmanagements aufzuzeigen. Durch eine individuelle Betrachtung der Teilbereiche des Informationsmanagement im Hinblick auf das IT-Qualitätsmanagement sollen zentrale Arbeiten identifiziert und der state-of-the-art zu diesen Teilbereichen dargestellt werden. Ferner soll mittels der identifizierten Arbeiten der aktuelle Stand der Wissenschaft bewertet und mögliche Forschungslücken aufgezeigt werden.

Im Rahmen dieses Beitrags sollen folgende Forschungsfragen beantwortet werden:

- **Q1:** Welche Teilbereiche lassen sich im Rahmen des strategischen IT-Informationsmanagement im Hinblick auf das IT-Qualitätsmanagement identifizieren?
- **Q2:** Welche aktuellen Herausforderungen und Schlüsselarbeiten lassen sich in den einzelnen Teilbereichen des IT-Informationsmanagement hinsichtlich des IT-Qualitätsmanagement finden?
- **Q3:** In welchen Teilbereichen des IT-Informationsmanagement lässt sich im Hinblick auf das IT-Qualitätsmanagement ein Forschungsbedarf identifizieren?

2 Grundlagen und Begriffe

2.1 Informationsmanagement

Zur Identifikation der Teilbereiche des IT-Qualitätsmanagements im Rahmen des Informationsmanagements ist es notwendig, den Begriff „Informationsmanagement“ im Kontext von diesem Beitrag zu definieren. Nach [Kr15 S.107] ist das Informationsmanagement „das Management der Informationswirtschaft, der Informationssysteme, der Informations- und Kommunikationstechniken sowie der übergreifenden Führungsaufgaben. Das Ziel des Informationsmanagements ist es, den im Hinblick auf die Unternehmensziele bestmöglichen Einsatz der Ressource Information zu gewährleisten“. [Ti20 S. 235] führt ferner aus, dass das „Informationsmanagement [sicherstellt], dass die richtige Information zur richtigen Zeit am richtigen Ort dem richtigen Adressaten zu angemessenen Kosten zur Verfügung steht. Aufgabe des Informationsmanagements ist es, dafür zu sorgen, dass alle Informationshandlungen im Unternehmen durch eine aufgabenadäquate Informationsversorgung effizient vollzogen werden können“. Diese Definitionen des Informationsmanagements werden im vorliegenden Beitrag angewendet, um den Stand der Wissenschaft zum IT-Qualitätsmanagement aufzuzeigen.

Analog zu den Definitionen zum Informationsmanagement finden sich in der Literatur ebenfalls unterschiedliche Modelle zum Informationsmanagement. [Ti20 S. 256]

beschreibt jedoch, dass sich das Model nach [Kr15 S.107] zum Standard innerhalb des Informationsmanagements durchgesetzt hat.

2.2 IT-Qualitätsmanagement

Die ISO 9001:2015 [Di15] bezeichnet das Qualitätsmanagement als Teilbereich, der alle Tätigkeiten umfasst, die dazu beitragen, dass die Qualitätsziele des Unternehmens erreicht werden. Im Rahmen des strategischen Informationsmanagements wird es als wichtige Teildisziplin mit dem Ziel der Qualitätssicherung innerhalb der einzelnen Bereiche gesehen [He14].

Im Rahmen des Informationsmanagements wird es von [Ti20 S. 665f] als strategische Aufgabe beschrieben. Er ordnet das IT-Qualitätsmanagement dem Teilbereich des IT-Controllings zu, welcher im Modell des Informationsmanagements nach [Kr15 S.107] den Führungsaufgaben des Informationsmanagements zugeordnet ist. Dies verdeutlicht die Notwendigkeit, das IT-Qualitätsmanagement auf sämtliche Bereiche des Informationsmanagements anzuwenden.

3 Ergebnisse

Zur Darstellung des Stands der Wissenschaft wurde eine fünfstufige Literaturanalyse nach [Fe06] durchgeführt. Dabei wurden ausgewählte Literaturdatenbanken (AIS Electronic Library, Association for Computing Machinery Digital Library, IEEE Xplore Digital Library, ScienceDirect, Springer Link) auf Beiträge im Zeitraum 2016 – 2021 untersucht. Dieser Zeitraum wurde definiert, da durch die Arbeit von [Kr15] aus dem Jahr 2015 das Modell des Informationsmanagements beschrieben wurde und in diesem Beitrag der Stand der Wissenschaft entlang dieses Modells dargestellt werden soll. Die Suche wurde dabei in der deutschen und englischen Sprache durchgeführt. Zur übersichtlichen Darstellung wird in den folgenden Kapiteln nur der deutsche Suchbegriff genutzt.

Auf Basis dieser Literaturanalyse wurden die Erkenntnisse zu aktuellen Forschungen und Forschungslücken je Teilbereich des Informationsmanagements gewonnen. Zu jedem Teilbereich wird außerdem der Forschungsbedarf in einer der unterhalb dargestellten Dimensionen angegeben.

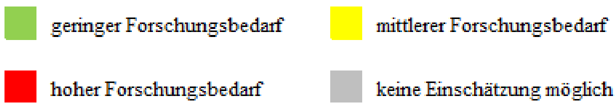


Abb. 1: Einordnungsmöglichkeiten des Forschungsbedarfs

Auf Grundlage des Modelles zum Informationsmanagement nach [Krcm15, S.107] werden nachfolgend die Teilbereiche des IT-Qualitätsmanagements im Rahmen des

strategischen Informationsmanagements identifiziert (Forschungsfrage Q1), Schlüsselarbeiten dargestellt (Forschungsfrage Q2) und eine Einordung hinsichtlich der Forschungsrelevanz (Forschungsfrage Q3) vorgenommen. Dadurch sollen die zuvor spezifizierten Forschungsfragen beantwortet werden. Nachfolgend werden die gewonnenen Erkenntnisse zum IT-Qualitätsmanagement je Teilbereich im Modell des Informationsmanagements beschrieben.

3.1 Management der Informationswirtschaft

[Krcm15, S.139ff.] beschreibt als zentrale Aufgabe des IT-Qualitätsmanagements im Teilbereich Management der Informationswirtschaft die Sicherstellung der Informationsqualität. Die Autoren [Po19 S. 12], [Kr05 S.78] und [BD08] nennen folgende Kriterien zur Informationsqualität, welche nachfolgend hinsichtlich Schlüsselarbeiten (Forschungsfrage Q2) und Forschungslücken (Forschungsfrage Q3) untersucht werden.

- Datenverfügbarkeit
- Datenzugänglichkeit
- Datennachverfolgbarkeit
- Datenqualität
- Zusammenfügen von Daten
- Datenrelevanz
- Datenvertraulichkeit

Auf Basis dieser Kriterien zur Informationsqualität wurde eine Literatursuche durchgeführt, wodurch insgesamt 290 Beiträge identifiziert wurden, 55 Beiträge wurde als relevant zur Darstellung des aktuellen Forschungsstandes betrachtet.

	Datenverfügbarkeit	Datenzugänglichkeit	Datennachverfolgbarkeit	Dateneignung	Zusammenfügen von Daten	Datenrelevanz	Datenvertraulichkeit
Anzahl Quellen	86	34	22	100	17	14	17
Anzahl relevanter Quellen	6	4	8	1	9	9	18
Forschungsbedarf							

Abb. 2: Forschungsbedarf Management der Informationswirtschaft

Im Teilbereich Datenverfügbarkeit werden hinsichtlich der Verfügbarkeit von Daten wenig Bedenken geäußert, jedoch stellt deren Verarbeitung die Wissenschaft und Unternehmen vor Herausforderungen [DZE17]. Um die Datenverarbeitung zu optimieren, wird der Einsatz von künstlicher Intelligenz [Le20, Za20] und Machine Learning [Roec20] untersucht. [GK18] erarbeiteten eine Backup Strategie im Falle des Datenverlustes.

Hinsichtlich der Zugänglichkeit von Daten thematisieren die Arbeiten vorrangig die Sicherstellung des ungewollten Zugriffs auf Unternehmensdaten [Gh20, SS20]. Im Teilbereich Nachverfolgbarkeit wird aktuell in allen identifizierten Arbeiten die Distributed Ledger Technologie als mögliche Lösung zur Sicherstellung der Datenrückverfolgbarkeit untersucht [Xu20]. [We20] beschreiben einen Ansatz der Rückverfolgung von Daten aus unternehmensübergreifender Kommunikation.

Die im Bereich der Dateneignung relevante Publikation beschäftigt sich mit der Auswahl geeigneter Daten mittels Machine Learning [JD18]. Aufgrund der Ergebnisse kann die Hypothese aufgestellt werden, dass der aktuelle Fokus der Wissenschaft auf der Verarbeitung großer Datenmengen ist und dieser Teilbereich folglich noch nicht relevant ist. Im Bereich des Zusammenfügens von Daten beschäftigen sich aktuelle Publikationen mit der Zusammenführung auf Basis des Ähnlichkeitsprinzips oder der Verbesserung korrupter Daten durch Nutzung mehrerer Datenquellen [Ch16].

Hinsichtlich der Datenrelevanz wird der Einsatz von Machine Learning untersucht [DR19]. Andere Arbeiten beschäftigen sich mit der Identifikation relevanter unstrukturierter Daten [Li20].

[Br17] entwickelten ein siebenstufiges Framework zur Bewertung der Datenrelevanz. Gemessen an der Anzahl an Beiträgen kann erkannt werden, dass die Zuverlässigkeit von Daten eine hohe Relevanz aufweist. Die Arbeiten von [Hi17], [Ra19] und [Sa17] konnten hierzu als Schlüsselarbeiten, welche sich mit den Grundlagen beschäftigen, identifiziert werden. Zusammenfassend lässt sich festhalten, dass in diesem Teilbereich wenig Forschungsbedarf identifiziert werden konnte. Im Bereich der Datennachverfolgbarkeit sehen [We20] und [Ei17] einen Bedarf nach Fallstudien und praktischer Umsetzung in Unternehmen.

3.2 Management der Informationssysteme

[Krcm15, S.108] unterteilt den Teilbereich Management der Informationssysteme in die Aufgaben Management der Daten, der Prozesse, des Anwendungslebenszyklus sowie der Systemlandschaft. Basierend auf dieser Unterteilung wird der aktuelle Stand der Wissenschaft im Hinblick der definierten Forschungsfragen zum IT-Qualitätsmanagement im Teilbereich Management der Informationssysteme untersucht. Dabei wurden auf Basis von [Krcm15 S.315ff] folgende Suchbegriffe je Aufgabe definiert:

- **Daten:** Datenmanagementstrategie, Datenbankarchitektur, Enterprise Content Management, Datenmodellierung und Stammdatenmanagement

- **Prozesse:** Prozessidentifikation, Prozessmodellierung, Prozessstrukturierung, Prozessstandardisierung, Prozesswiederverwendung und Prozessevaluierung
- **Anwendungslebenszyklus:** Softwareeinführung, Softwarewartung, Softwareablösung, Anforderungsmanagement und Softwareauswahl
- **Systemlandschaft:** Multi-Projektmanagement, IS-Portfoliomanagement

	Daten					Prozesse						Anwendungslebenszyklus					Systemlandschaft	
	Datenmanagementstrategie	Datenbankarchitektur	Enterprise Content Management	Datenmodellierung	Stammdatenmanagement	Prozessidenzifikation	Prozessmodellierung	Prozessstrukturierung	Prozessstandardisierung	Prozesswiederverwendung	Prozessevaluierung	Softwareeinführung	Softwarewartung	Softwareablösung	Anforderungsmanagement	Softwareauswahl	Multi-Projektmanagement	IS-Portfoliomanagement
Anzahl Quellen	246	141	13	270	27	71	85	36	16	15	102	106	33	39	15	146	13	77
Anzahl relevanter Quellen	20	28	12	20	21	1	11	0	1	1	0	5	6	3	6	2	9	0
Forschungsbedarf																		

Abb. 3: Forschungsbedarf Management der Informationssysteme

Im Bereich der Datenmanagementstrategie [Th19] sowie im Stammdatenmanagement [Iq19, Pr18] wird das Themengebiet Big Data anhand von Fallstudien untersucht. Hier liegt der Fokus auf kontextabhängiger und nicht auf allgemeiner Forschung. Im Bereich des Enterprise Content-Management hingegen liegt der Fokus auf den Grundlagen [Dr19, Ta18]. Hier konnte ein großer Forschungsbedarf identifiziert werden. In den Bereichen Datenbankarchitekturen [Ug19] und Datenmodellierung [Ve20] liegt der Fokus auf dem Wandel von SQL zu NoSQL Datenbanken. Innerhalb der Datenmodellierung liegt der Fokus der Wissenschaft ebenfalls auf Fallstudien zu den durch Big Data auftretenden Herausforderungen [Ra19].

Im Teilbereich der Prozesse zeigt die Arbeit von [RM16], dass hier seit 2013 ein Abwärtstrend an Publikationen erkennbar ist. Diese Ergebnisse decken sich mit denen aus der Literaturanalyse, dieser Teilbereich ist weitestgehend untersucht. Einzig in der Prozessmodellierung versuchen Arbeiten wie die von [WS18] die steigende Komplexität zu minimieren.

Im Bereich der Softwareeinführung sind gängige Methoden vorhanden, welche weiterhin genutzt werden [WZ18]. Die Forschungen zur Softwareablösung beschäftigten sich damit, wann [HKK17] und wie [Vi17] ein System abgelöst werden kann. Innerhalb des Anforderungsmanagements werden keine konkreten Anforderungen, sondern Ansätze wie Gamification [Ga17] untersucht, um passende Anforderungen zu stellen. Was die Wartung von Software angeht, werden die Kostenuntersuchung [GS16] und Qualitätsmessung [Ra19] erforscht.

Zum IS-Portfoliomanagement konnten keine relevanten Quellen gefunden werden. Dieser Bereich sollte folglich genauer untersucht werden, um den Forschungsbedarf festlegen zu

können. Das Multi-Projektmanagement hingegen ist gut untersucht. Abschließend lässt sich festhalten, dass im Bereich der Datenmodellierung ein hoher Forschungsbedarf hinsichtlich qualitativ hochwertiger Modellierung der steigenden Menge an Daten identifiziert werden konnte. Auch im Bereich des Anwendungslebenszyklus konnte, was die Umsetzung angeht, eine Lücke identifiziert werden.

3.3 Management der Informations- und Kommunikationstechnik

In seiner Arbeit [Kr15 S.317] unterteilt er diesen Fachbereich in die Teilbereiche Management der Verarbeitung, Speicherung und Kommunikation. Auf Basis der Teilaufgaben innerhalb der einzelnen Bereiche wurden diese in folgende Suchbegriffe unterteilt:

- **Verarbeitung**: Datentransformation, Datenaggregation, Datenspezifikation und Informationsverarbeitung
- **Speicherung**: Datenspeicherung, Datenverlust, Datenwiederherstellung, Datenredundanz, Speichernetze und Information Lifecycle Management
- **Kommunikation**: Kommunikationsstandards, Kommunikationsnetzwerke und Trends der Kommunikationstechnik

Der ebenfalls von [Kr15 S.318f] definierte Teilbereich Technikmanagement wurde nicht untersucht, da es sich hierbei um das unternehmensindividuelle Management der genutzten Technologien handelt.

	Verarbeitung				Speicherung						Kommunikation		
	Datentransformation	Datenaggregation	Datenspezifikation	Informationsverarbeitung	Datenspeicherung	Datenverlust	Datenwiederherstellung	Datenredudanz	Speichernetze	Information Lifecycle Management	Kommunikationsstandards	Kommunikationsnetze	Trends der Kommunikationstechnik
Anzahl Quellen	74	131	6	317	346	57	165	412	56	2	39	135	37
Anzahl relevanter Quellen	7	8	0	10	41	9	11	20	18	1	4	9	2
Forschungsbedarf													

Abb. 4: Forschungsbedarf Management der Informations- und Kommunikationstechnik

In den Bereichen Datentransformation und Datenaggregation konnte der Bedarf nach automatisierter Verarbeitung, welcher aktuell bereits untersucht [Sa18] wird, identifiziert werden. Andere Arbeiten verfolgen den Ansatz der Reduktion durch Aggregation [DN19, SRT18]. Alle Arbeiten zielen darauf ab, die von [Gu17] beschriebenen 5V zu beheben.

Diese fassen ebenfalls aktuelle Probleme für zukünftige Arbeiten zusammen. Was die Datenspezifizierung angeht, lässt sich die Hypothese aufstellen, dass die Definition von Datentypen bereits abgehandelt ist.

Innerhalb der Datenspeicherung liegt der aktuelle Fokus auf Datensicherheit [Li20], Internet of Things [Tc20], Zuverlässigkeit [KEA17], Big Data [Bh18] und der Blockchain [SIV20]. Ebenfalls werden Datenverlust und Datenwiederherstellung z.B. mit Hilfe von Machine Learning [Ni20] untersucht, jedoch bedarf es hier nach [Fe19] einer Fokussierung auf die Prävention des Datenverlustes. Im Bereich der Datenredundanz wird hauptsächlich der Redundant Array of Independent Disks untersucht [So20]. Die Verfeinerung von Network Attached Storage [TC19] ist der aktuelle Fokus der Speichernetze. Ebenfalls lassen sich erste Arbeiten zur Nutzung der Blockchain finden [Ya20]. Weitere hierzu sollten in Zukunft durchgeführt werden. Innerhalb des Information Lifecycle Managements besteht ein geringer Forschungsbedarf, das aktuelle Modell hat weiterhin Bestand.

Im Zentrum der Forschung zur Kommunikation liegt die 5G-Technologie. Hierzu wird deren Einfluss auf die Kommunikation zwischen Menschen [Mo17], sowie Maschinen [ZS20] untersucht. Fokus liegt jedoch auf der Maschine – Maschine Kommunikation [Re19], welche ebenso wie die Datenqualität der von Maschinen erhaltenen Informationen untersucht wird [Ol17].

Zusammenfassend lässt sich festhalten, dass in diesem Teilbereich speziell die von [Gu17] beschriebenen 5V und deren Probleme angegangen werden sollten. Innerhalb der Kommunikation stellt die 5G-Technologie als möglicher zukünftiger Standard einen Forschungsbedarf dar, welcher möglicherweise starken Einfluss auf die Qualität der Kommunikation nehmen könnte.

4 Diskussion und Ausblick

Zusammenfassend lässt sich festhalten, dass durch die durchgeführte Literaturanalyse der aktuelle Stand der Wissenschaft zum IT-Qualitätsmanagement im Rahmen des Informationsmanagements dargestellt werden konnten. Insgesamt wurden 38 Themen, aufgeteilt auf die drei Teilbereiche des Informationsmanagements, identifiziert, welche eine Relevanz für das IT-Qualitätsmanagement aufweisen. Im Rahmen des Beitrags wurden alle Teilbereiche des Informationsmanagements hinsichtlich des aktuellen Stands der Wissenschaft zum IT-Qualitätsmanagement untersucht. Insgesamt wurden 341 relevante Publikationen identifiziert, welche Themen des IT-Qualitätsmanagements im Rahmen des Informationsmanagements adressieren. Durch die anschließend durchgeführte Analyse dieser Beiträge konnte der aktuelle Stand der Wissenschaft zu den einzelnen Teilbereichen dargestellt und ein Forschungsbedarf innerhalb einzelner Teilbereiche identifiziert werden.

Für zukünftige Arbeiten wäre es interessant, den identifizierten Forschungsbedarf weiter zu untersuchen und diesen durch wissenschaftliche Beiträge zu erfüllen. Hierdurch würde ein wesentlicher Beitrag zum IT-Qualitätsmanagement im Informationsmanagement geleistet werden. Zudem wäre es interessant, die durchgeführte Literaturanalyse zur Erhebung des aktuellen Stands der Wissenschaft zum IT-Qualitätsmanagement im Rahmen des Informationsmanagements in der Zukunft zu wiederholen. Dadurch kann zum einen die Entwicklung in diesem Themenbereich weiter aufgezeigt und ferner die Erkenntnisse aus diesem Beitrag hinsichtlich dem identifizierten Forschungsbedarf, validiert werden.

5 Literatur

[BD08] Dragen, Bernhard: Bewertung der Informationsgüte in der Informationsgewinnung für die modellgestützte Analyse großer Netze der Logistik Bewertung der Informationsgüte in der Informationsgewinnung für die modellgestützte Analyse großer Netze der Logistik (2008)

[Bh18] Bhat: Is a Data-Capacity Gap Inevitable in Big Data Storage? (2018)

[Br17] Bridges et al.: Framework for the quantitative weight-of-evidence analysis of 'omics data for regulatory purposes, Elsevier Ltd (2017)

[Ch16] Chen et al. An Efficient Privacy-Preserving Ranked Keyword Search Method, IEEE (2016)

[DN19] Das namboodiri: A Quality-Aware Multi-Level Data Aggregation Approach to Manage Smart Grid AMI Traffic, IEEE (2019)

[Di15] Din: DIN ISO 9001:2015 Qualitätsmanagementsysteme - Anforderungen (2015)

[DZE17] Dieye Zhani Elbiaze: On achieving high data availability in heterogeneous cloud storage systems (2017)

[Dr19] Draheim: On the Radical De- and Re-Construction of Today's Enterprise Applications, Elsevier B.V. (2019)

[Ei17] Eickhoff: What do they mean? Using Media Richness as an Indicator for the Information Value of Stock Analyst Opinion regarding post-earnings Firm Performance (2017)

[Fe06] Fettke: State-of-the-Art des State-of-the-Art Eine Untersuchung der Forschungsmethode „Review" innerhalb der Wirtschaftsinformatik (2006)

[Fe19] Feng et al: Data Loss Prevention and Storage Utilization Improvement of the Hidden Volume on Mobile Devices., 2019, S. 1–6

[GK18] Gawade Kadu: Secure Data Storage and Efficient Data Retrieval Over Cloud Using Sensitive Hashing., 2018, S. 87–92

[Gh20] Ghaffar et al.: An improved authentication scheme for remote data access and sharing over cloud storage in cyber-physical-social-systems, IEEE (2020)

[Gu17] Guan et al.: A survey on big data pre-processing (2017)

[Ga17] Garcia et al.: A framework for gamification in software engineering (2017)

[GS16] Guo Seaman da Silva: Costs and obstacles encountered in technical debt management – A case study, Elsevier Inc. (2016)

[Hi17] Hinrichs et al.: Analyzing qualitative data (2017)

[HKK17] Henttonen Kääriäinen Kylmäaho: Lifecycle management in government-driven open source projects - practical framework (2017)

[He14] Heinrich et al.: Informationsmanagement: Grundlagen, Aufgaben, Methoden : De Gruyter, 2014

[Iq19] Iqbal et al A.: Master Data Management Maturity Assessment: Case Study of XYZ Company., 2019, S. 133–139

[JD18] Jia Ding: Reliability Assessment of Data Storage in Cyber Physical Systems., 2018, S. 62–66

[KEA17] Kishani Eftekhari Asadi: Evaluating impact of human errors on the availability of data storage systems., 2017, S. 314–317

[Kr05] Krcmar: Informationsmanagement : Springer Berlin Heidelberg, 2005

[Kr15] Krcmar: Informationsmanagement. 6. Auflage. : Springer Gabler, 2015

[Li20] Liang et al.: Secure Data Storage and Recovery in Industrial Blockchain Network Environments (2020)

[Le20] Leyer et al.: Decision-making with artificial intelligence: Towards a novel conceptualization of patterns (2020)

[Li20] Liu et al.: Semantic-aware data quality assessment for image big data, Elsevier B.V. (2020)

[Mo17] Mowla et al.: An energy efficient resource management and planning system for 5G networks (2017)

[Ni20] Nijma et al.: Deep Learning Based Data Recovery for Localization (2020)

[Ol17] Olsen et al .: On the use of information quality in stochastic networked control systems (2017)

[Pr18] Pratama et al.: Master Data Management Maturity Assessment: A Case Study of Organization in Ministry of Education and Culture., 2018, S. 1–6

[Po19] Pospiech: Aufgabengerechte bereitstellung in Zeiten von Big Data - Konsequenzen für ein Informationsmanagement, 2019

[Ra19] Ramzan et al Intelligent Data Engineering for Migration to NoSQL Based Secure Environments (2019)

[Re20] Reinartz: Geschäftsmodelle in die Zukunft denken (2020)

[RM16] Recker Mendling: The state of the art of business process management research as published in the BPM conference: Recommendations for progressing the field, Springer Fachmedien Wiesbaden (2016)

[RR19] Reddivari Raman: Software quality prediction: An investigation based on machine learning, IEEE (2019)

[Ro20] Roeck: The Foundation of Distributed Ledger Technology for Supply Chain Management (2020)

[Ra19] Rabcam et al.: Analysis of Data Reliability based on Importance Analysis, IEEE (2019)

[Re19] Reiff et al.: Model for the Client-Oriented Selection of Addititve Manufacturing Infrastrcuture based on Information gathered from Production Networks (2019)

[SR18] Salam Rehman Tao: Cooperative Data Aggregation and Dynamic Resource Allocation for Massive Machine Type Communication (2018)

[SS20] Schwade Schubert: A Semantic Data Lake for Harmonizing Data from Cross-Platform Digital Workspaces using Ontology-Based Data Access (2020)

[So20] Son et al.: Design and Implementation of SSD-Assisted Backup and Recovery for Database Systems (2020)

[Sa17] Sairanen et al.: A novel measure of reliability in Diffusion Tensor Imaging after data rejections due to subject motion, Elsevier (2017)

[Sa18] Saez et al.: A Data Transformation Adapter for Smart Manufacturing Systems with Edge and Cloud Computing Capabilities, IEEE (2018)

[SIV20] Sosin Ivanova Vasilyeva: Prospects for Implementing Blockchain Data Storage Technology as a Process of Digital Transformation of Society., 2020, S. 1–5

[Ta18] Tarkhanov: Access control model for collaborative environments in ECM, IEEE (2018)

[Tc20] Tchernykh babenko et al.: Scalable Data Storage Design for Nonstationary IoT Environment With Adaptive Security and Reliability (2020)

[Th19] Thomas et al Data management maturity assessment of public sector agencies, Elsevier (2019)

[TC19] Tebbi Chan: Multi-Rack Distributed Data Storage Networks (2019)

[Ti20] Tiemeyer: Handbuch IT-Management: Konzepte, Methoden, Lösungen und Arbeitshilfen für die Praxis. 7. Aufl. : Hanser, 2020

[Ug19] Ugli: Electronic invoice system database architecture issues and the possible solutions, IEEE (2019)

[Ve20] Vega et al.: Mortadelo: Automatic generation of NoSQL stores from platform-independent data models (2020)

[Vi17] Villeneuve.: Decision-Support Methodology to Assess Risk in End-of-Life Management of Complex Systems (2017)

[WS18] Wang Sun: Business Process Modeling Abstraction Based on Semi-Supervised Clustering Analysis, Springer Fachmedien Wiesbaden (2018)

[We20] Wei: Blockchain-based Data Traceability Platform Architecture for Supply Chain Management (2020)

[WZ18] Wyrwicka Zasadan: Implementation of Information Management System: Human Factors Lessons Learned from Industrial Company (2018)

[Xu20] Xu et al.: Blockchain: A new safeguard for agri-foods, Elsevier B.V. (2020)

[Ya20] Yang et al.: Topology-Aware Cooperative Data Protection in Blockchain-Based Decentralized Storage Networks., 2020, S. 622–627

[Za20] Zapadka et al.: Leveraging „AI-as-a-Service“ - Antecedents and Consequences of Using Artificial Intelligence Boundary Resources (2020)

[ZS20] Zhao Song: A Reinforcement Learning Method for Joint Mode Selection and Power Adaptation in the V2V Communication Network in 5G, IEEE (2020)

Künstliche Intelligenz im Requirements Engineering

Eine Systematische Literaturrecherche

Judith Breuninger[1], Franziska Kücher[2] und Natali Misic[3]

Abstract: Dem Einsatz von Künstlicher Intelligenz (KI) im Requirements Engineering (RE) wird ein hohes Potenzial zugeschrieben. Der Stand der Forschung gestaltet sich jedoch unübersichtlich. Im Rahmen einer Systematischen Literaturrecherche werden 27 wissenschaftliche Publikationen aus drei Datenbanken identifiziert und analysiert. Anschließend werden diese in die RE-Phasen der Anforderungserhebung, -analyse, -spezifikation und -validierung eingeordnet und zusammengefasst. Die Ergebnisse zeigen, dass KI in den vier Phasen eingesetzt wird, allerdings ist die Anwendung unterschiedlich stark ausgeprägt. Weitere tiefgehende Forschungsarbeit, insbesondere zum Einsatz von KI in der Anforderungsvalidierung, ist notwendig. Die vorliegende Arbeit stellt dafür einen wesentlichen Ausgangspunkt dar, indem sie einen strukturierten Überblick der verschiedenen KI-Methoden zum Einsatz im RE aufzeigt und diskutiert.

Keywords: Künstliche Intelligenz, Requirements Engineering, Systematische Literaturrecherche

1 Einleitung

Obwohl die Wurzeln der Künstlichen Intelligenz (KI) bis in die Mitte des zwanzigsten Jahrhunderts zurückreichen, besitzt das Thema insbesondere heute hohe Relevanz in Wissenschaft und Praxis [Er08]. Die Zahl der qualitätsgeprüften Publikationen stieg bspw. im Jahr 2018 um knapp die Hälfte im Vergleich zum Vorjahr [Ac20].[4] Im Requirements Engineering (RE) leistet KI Unterstützung aufgrund der steigenden Aufgabenkomplexität. Zukünftig könnte der Problemlösungsfähigkeit durch KI eine relevante Rolle bzgl. der Automatisierung von Prozessen im RE zukommen [Ba20], [Er08], [De15].

Das RE beschäftigt sich mit Anforderungen während des gesamten Lebenszyklus des Softwareprodukts. Ziel ist es, ein grundlegendes Verständnis der Anforderungen bei allen Projektbeteiligten zu schaffen. Bereits in den frühen Phasen der Entwicklung können Fehler aufgezeigt, behoben und damit Kosten reduziert werden, um eine erfolgreiche Softwareentwicklung zu erreichen [BF14], [Eb19], [IE90], [JIA18], [Ng12], [Pa10].

In der wissenschaftlichen Literatur wird KI zwar intensiv erforscht, gilt jedoch zugleich als schwer definierbar. Grundsätzlich wird ein System als künstlich intelligent verstanden,

[1] Universität Stuttgart, Keplerstr. 17, 70174 Stuttgart, judith.breuninger7@gmail.com
[2] Universität Stuttgart, Keplerstr. 17, 70174 Stuttgart, franzi.kuecher@gmail.com
[3] Universität Stuttgart, Keplerstr. 17, 70174 Stuttgart, natali.misic@web.de
[4] Im Jahr 2018 wurden knapp 90.000 Beiträge in Europa, China und den USA veröffentlicht, was einen Anstieg von ungefähr 45 Prozent im Vergleich zum Vorjahr bedeutet [Ac20].

wenn es Aufgaben oder Konzepte selbstständig bzw. effizient erlernen und Herausforderungen in der realen Welt lösen kann [Ba20], [Er08], [Ma16]. Trotz des hohen Potenzials von KI im RE weist die Fachliteratur darauf hin, dass sich Forschung noch am Anfang befindet. Darüber hinaus existiert kein übersichtlicher Forschungsstand zum Einsatz von KI im RE [Ba20], [De15]. Zur Lösung dieses Problems soll die vorliegende Arbeit einen Beitrag leisten. Im Rahmen einer Systematischen Literaturrecherche (SLR) wird der Status quo hinsichtlich des Einsatzes von KI im RE erarbeitet und für weitere Forschung dargelegt. Aufgrund dessen lautet die Forschungsfrage: "Wie gestaltet sich der Einsatz von KI-Methoden in den einzelnen Phasen des RE nach aktuellem Stand der wissenschaftlichen Literatur?".

2 Methodik

Zur strukturierten Erhebung und Analyse der wissenschaftlichen Literatur orientiert sich diese Forschungsarbeit am Ansatz zur Durchführung einer SLR nach Kitchenham [Ki09].

Der Suchterm der SLR lautet: „artificial intelligence" AND „requirements engineering". Dieser wird in allen durchsuchbaren Feldern, insbesondere im Titel, Abstract und Volltext, für den Zeitraum von 2010 bis 2020 innerhalb der folgenden Datenbanken angewandt: Association for Computing Machinery Digital Library (ACM DL), ScienceDirect sowie EBSCOhost. Die Datenbanken bieten Zugang zu umfassenden Volltext-Artikeln, Journalen und Fachliteratur [As15], [EB20], [El20].

Der in Abbildung 1 dargestellte Exklusionsprozess der relevanten Literatur erfolgt in fünf Schritten im Zeitraum von Mai bis Juni 2020. Die anfänglichen Suchtreffer belaufen sich auf 746 Publikationen. In drei Schritten werden die Suchergebnisse eingegrenzt. Schließlich bleiben nach dem vierten Schritt 30 Publikationen bestehen. Durch eine identische Publikation in zwei Datenbanken kann die Anzahl auf 29 Publikationen reduziert werden.[5] Des Weiteren werden zwei Publikationen ausgeschlossen, die ebenfalls eine SLR durchführen und damit für diese Arbeit Sekundärquellen darstellen. Zudem verwenden die ausgeschlossenen SLR eine abweichende Phaseneinteilung des RE und beziehen sich lediglich auf jeweils eine KI-Methode. Im Rahmen der Datensynthese werden 27 Publikationen analysiert und den Phasen des RE inhaltlich zugeordnet.

Die Phaseneinteilung des RE ist essentiell für das methodische Vorgehen, da diese die Grundlage der Strukturierung und Auswertung der Ergebnisse bildet. Die vorliegende Arbeit orientiert sich an der Definition des IEEE, welche das RE in die folgenden vier Phasen unterteilt [IE90]: Die Anforderungserhebung als erste Phase im RE hat den Zweck, Anforderungen an ein Softwaresystem zu ermitteln oder aus Systemanforderungen abzuleiten. Dabei wird ein umfassendes Verständnis der Problemstellung erarbeitet, welche das

[5] Die Publikation von [IAA18] entstammt den Datenbanken ScienceDirect und EBSCOhost.

zu entwickelnde Softwaresystem lösen soll. In der zweiten Phase, der Anforderungsanalyse, wird die Vollständigkeit, Korrektheit, sowie die technische und finanzielle Durchführbarkeit auf Basis der erhobenen Anforderungen untersucht und bewertet. Außerdem werden Grenzen des Softwaresystems identifiziert, indem dieses von seinem Kontext abgegrenzt wird. In der Anforderungsspezifikation werden im dritten Schritt die erhobenen und analysierten Anforderungen im Spezifikationsdokument präzisiert festgehalten. Das Dokument wird dabei systematisch überprüft, bewertet und genehmigt. In der vierten Phase, der Anforderungsvalidierung, wird das Spezifikationsdokument auf Vollständigkeit und Korrektheit überprüft [BF14], [DD10].

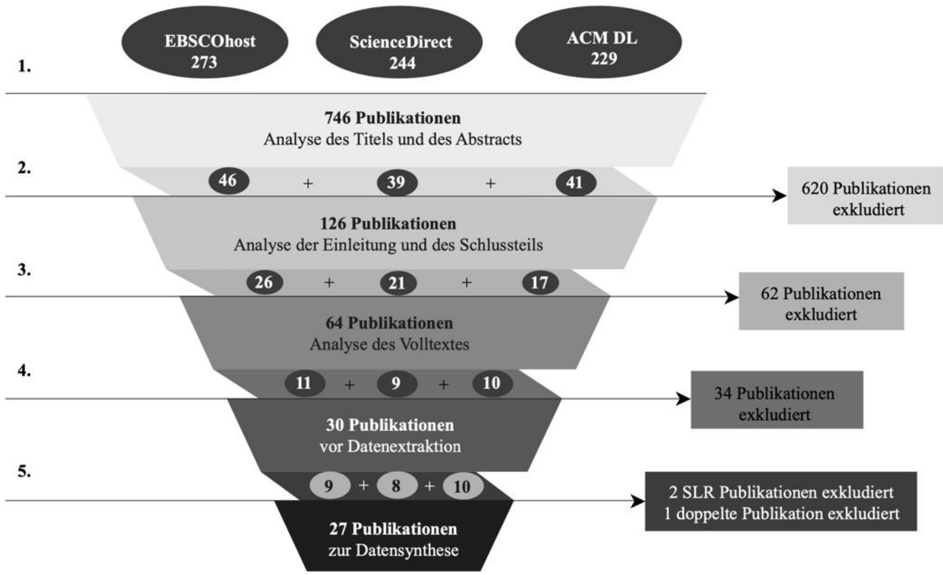


Abb. 1: Exklusionsprozess im Rahmen der SLR

3 Ergebnisse

Aus der SLR können zum Einsatz von KI im RE sieben Publikationen der Anforderungserhebung zugeordnet werden. Für die Anforderungsanalyse und Anforderungsspezifikation werden jeweils neun Publikationen identifiziert. Zwei Publikationen werden in die Anforderungsvalidierung eingeordnet. In wenigen Publikationen werden KI-Methoden phasenübergreifend thematisiert, was eine konkrete Zuordnung erschwert. In diesen Fällen wird die Publikation in jene Phase eingeordnet, die vordergründig behandelt wird. Eine zusammenfassende Übersicht aller relevanten Publikationen und den darin thematisierten KI-Methoden enthält Tabelle 1.

Nr.	Literatur	KI-Methode
Anforderungserhebung		
1	[DR20]	Algorithmus basierend auf NLP
2	[GAS17]	ML
3	[Ma19]	KNN, Worteinbettung, ML, Inferenzmaschine
4	[MX19]	NLP, NLU
5	[Ng12]	Ontologien
6	[Ro19]	Data Mining
7	[We12]	Tool basierend auf NLP
Anforderungsanalyse		
8	[BMS20]	Neuronales Worteinbettungsmodell
9	[Ca14]	NLP, Data Mining
10	[DS11]	Unified Modeling Language Modellgenerator mit NLP
11	[GG18]	Particle Swarm Optimization basierend auf ML
12	[Ka19]	KNN, Deep-Learning Algorithmus
13	[Lu17]	Tool basierend auf NLP
14	[MSP16]	NLP, ML
15	[Pa15]	Regelinduktiontsechniken basierend auf ML
16	[Vi10]	Multiagentensysteme
Anforderungsspezifikation		
17	[BLG17]	Ontologien
18	[ENT18]	Tool basierend auf NLP
19	[HK16]	ML, Association Rule Mining
20	[IAA18]	KNN, NLP
21	[JIA18]	NLP, Semantic Role Labeling
22	[JN14]	NLP, Inferenzmaschine, Natural Language Generation
23	[Na20]	Standford CoreNLP
24	[Sa12]	NLP, Semantik
25	[Zh10]	Algorithmus basierend auf ML
Anforderungsvalidierung		
26	[Al15]	Logikbasiertes Lernen
27	[Mo16]	Tool basierend auf NLP

Tab. 1: Ergebnisse der SLR

3.1 Anforderungserhebung

In [GAS17] werden zwei Ansätze Maschinellen Lernens (ML) vorgestellt. Durch ML kann eine Software auf Basis bestehender Datensätzen trainiert werden, um darauf aufbauend Vorhersagen zu treffen. Mit Hilfe von Algorithmen können Anforderungen an eine

Software aus Twitter Tweets erhoben werden. Diese können relevante Informationen für Requirements Engineers enthalten, wie bspw. Softwarefehler oder Wünsche für neue Funktionen. Auf Basis manuell ausgewerteter Tweets werden fünf Algorithmen trainiert, um den Text der Tweets zu analysieren. Als Ergebnis des ersten Ansatzes zeigt sich, dass Tweets nach ihrer Relevanz für technische, nicht-technische und allgemeine Stakeholder klassifiziert werden können. Somit können Requirements Engineers die für sie relevanten technischen Tweets für die Erhebung von Anforderungen nutzen. Im zweiten Ansatz lassen sich durch die Verwendung der Algorithmen von Bots generierte Tweets filtern. Dadurch können sich Stakeholder und Requirements Engineers auf relevante, manuell verfasste Tweets konzentrieren [BS19]. [Ma19] beschreibt zudem die Anwendung von ML in Form von Künstlichen Neuronalen Netzen (KNN). Dabei handelt es sich um ein Netzwerk, welches das biologische Nervensystem zur Informationsverarbeitung nachahmt. KNN tragen dazu bei, unerwünschte Verhaltensweisen eines späteren Produkts frühzeitig zu erkennen und zu reduzieren. Bspw. werden diese eingesetzt, um bei der Generierung von Modellen Elemente aus Use Cases zu filtern [IAA18].

[Ng12] thematisiert ein wissensgestütztes Framework, das auf Ontologien basiert. Letztere beschreiben Relationen zwischen Objekten. Das Domänenwissen und die Semantik von Anforderungen werden in einer Ontologie kodiert, sodass implizite Konsequenzen von expliziten Merkmalen in Anforderungen enthüllt werden. Des Weiteren wird eine Beschreibungslogik als Regel in der Ontologie verwendet, die zur Erfassung von Anforderungen sowie deren Einschränkungen und Beziehungen notwendig ist. Diese eignet sich, um Domänenkonzepte sinnvoll zu repräsentieren. Folglich lassen sich Inkonsistenzen und fehlende Elemente in Anforderungen identifizieren und beheben. Darüber hinaus werden zur Realisierbarkeit einer Anforderung automatisch Analysen durchgeführt [Er08], [Ng12].

Natural Language Processing (NLP) ist ein Prozess der Verarbeitung natürlicher Sprache zu digitalisiertem Text. Eine bedeutende Herausforderung stellt nach [MX19] das Verständnis und die Interpretation der natürlichen Sprache dar. Dabei handelt es sich um Natural Language Understanding (NLU), welches ein Teilbereich des NLP ist. Der Einsatz betrifft Sprachaktivierung, Textkategorisierung, Inhaltsanalyse und Beantwortung von Kundenfragen. Mit Hilfe von Techniken des NLU wird maschinelles Lesen und die Interpretation natürlicher Sprache ermöglicht. Ziel ist es, die Formulierung von Anforderungen qualitativ zu verbessern, indem Redundanzen, Mehrdeutigkeiten und Unvollständigkeiten identifiziert werden. In [DR20] wird eine interaktive Chatbot-Anwendung für Android eingesetzt, die mit Hilfe eines Algorithmus basierend auf NLP eine Nutzeranfrage aus einem Chat aufbereitet. Aus einer auf KI basierenden Wissensdatenbank generiert der Chatbot anschließend die Antwort auf eine Nutzeranfrage. Dadurch gelingt die Schließung der Kommunikationslücke zwischen Entwicklern und Nutzern eines Softwaresystems. Durch die Interaktion mit den Nutzern können neue Anforderungen aufgezeigt und Feedback berücksichtigt werden.

In [Ro19] wird ein auf Data Mining gestützter Ansatz vorgestellt. Dieser soll den Kontext spezifizieren, in dem ein zu entwickelndes Softwaresystem eingesetzt wird. Data Mining

umfasst einen Prozess, welcher Wissen aus Daten zieht und dieses anschließend darstellt und anwendet. Der Ansatz ist in der Lage, Kontextvariablen und Kombinationen solcher Variablen zu identifizieren und die Unsicherheiten im operativen Umfeld zu reduzieren. Mit Hilfe des Ansatzes wird somit die kontextbezogene Anforderungserhebung optimiert und detailliert dargestellt [Er08].

3.2 Anforderungsanalyse

[Lu17] und [Ca14] zeigen, dass NLP über die Anforderungserhebung hinausgeht. Die extrahierten Anforderungen werden in der Anforderungsanalyse mit NLP-Tools analysiert und auf ihre Qualität geprüft. Nach [Lu17] können empirische Heuristiken bei der Generierung konzeptueller Modelle unterstützen, um relevante Objekte oder Sachverhalte und deren Beziehungen zu erheben. Ein weiteres Tool generiert spezielle Ansichten des konzeptuellen Modells durch Clustering oder Filtern. [Ca14] beschreibt eine Methode, die NLP und Data Mining einsetzt, um Requirements Engineers bei der Extraktion und Modellierung von Zielen aus Textdokumenten zu unterstützen. Bei der Entwicklung komplexer Systeme besteht häufig die Notwendigkeit, eine große Anzahl von veralteten Anforderungsspezifikationen, Forschungspublikationen oder Standards zu analysieren. Data-Mining-Techniken können helfen, den manuellen Aufwand für die Analyse dieser Ressourcen zu reduzieren, indem die Abstracts der veralteten Dokumente automatisch gesammelt werden. Durch NLP werden anschließend Sätze aus dem Abstract extrahiert, die zielbezogene Schlüsselwörter enthalten. Der Ansatz ist eine effektive Möglichkeit, um Ziele für komplexe und forschungsintensive Systeme zu modellieren. In [DS11] wird eine Technik mit NLP vorgestellt, die Entwickler bei der Generierung von Unified Modeling Language-Modellen aus natürlichsprachlichen Anforderungen unterstützt. Es lassen sich Use Case Diagramme und verschiedene Analysen generieren. Aus [Ca14], [DS11] und [Lu17] kann entnommen werden, dass NLP in dieser Phase, ebenso wie in der Anforderungserhebung, zur Verbesserung von Anforderungsformulierungen und zur Reduktion des manuellen Aufwands eingesetzt wird.

Durch die Kombination von NLP und ML werden laut [MSP16] manuell interpretierbare Cluster aus semantisch verwandten Anforderungen in Textform generiert. Domänenrelevante Begriffe lassen sich extrahieren und die semantische Nähe zwischen Begriffen und Anforderungen analysieren. Für ein besseres Verständnis werden die entstandenen Cluster mit Erklärungen versehen. Laut [Pa15] werden Regelinduktionstechniken, welche auf ML basieren, zur Nachbildung von Bewertungen eingesetzt, die ein Qualitätsexperte eines Projekts vornehmen würde. Als Input erhalten diese einen Datensatz, woraus Lernbeispiele gezogen werden mit dem Ziel, einen Entscheidungsbaum oder eine Zusammenstellung von Regeln zu erstellen. Dadurch können neue Stichproben aus dem Datensatz bestimmt und klassifiziert werden. Durch die frühzeitige Qualitätsüberprüfung der Daten lassen sich fehlerhafte Anforderungen sowie Kosten oder Lieferungsverzögerungen bei der Erreichung der vorgegeben Ziele reduzieren. ML in Form von KNN wird nach [Ka19] genutzt, um Anforderungen aus unvollständigen Produktbeschreibungen auf bspw. Internetseiten zu erhalten und zu modellieren. Die KNN gleichen Anforderungsbeschreibungen

verwandter Softwaresysteme miteinander ab, extrahieren mit Hilfe eines Deep-Learning-Algorithmus relevante Merkmale und entwickeln diese weiter. Hierdurch wird die Qualität der Anforderungen an ein Softwaresystem verbessert.

Die Generierung semantischer Ähnlichkeitsmaße in der Anforderungsanalyse wird durch die Nutzung eines neuronalen Modells zur Worteinbettung unterstützt. So lassen sich Glossarbegriffe aus umfangreichen Anforderungsdokumenten gruppieren [BMS20].

3.3 Anforderungsspezifikation

Für die Anforderungsspezifikation ist die Extraktion von Aktionen und Akteuren aus natürlichsprachlichen Anforderungen entscheidend. Dieser manuelle Vorgang gilt als anspruchsvoll und zeitaufwendig. [IAA18] stellen die Kombination aus NLP und KNN als eine selbst entwickelte intelligente Technik des RE vor, welche zwei Arten der Verarbeitung, Syntax und Semantik, umfasst. Während der Generierung von Aktionen und Akteuren erfordert der halbautomatisierte Prozess Reaktionen des Benutzers. Jede Interaktion des Benutzers mit dem System wird als Use Case formuliert, aus welchem die Akteure und Aktionen mittels NLP erfasst werden. KNN werden zur semantischen Verarbeitung eingesetzt, um die aus der Anwendung von NLP hervorgegangenen Wörter auf die Semantik des Use Cases zu beziehen. [JIA18] zeigt die Kombination von NLP mit Semantic Role Labeling auf, die zur Extraktion von Aktionen und Akteuren sowie zur Erstellung von Use Case Diagrammen eingesetzt wird. Zum Verständnis der natürlichen Sprache wird durch Semantic Role Labeling Wörtern oder Phrasen eine Beschriftung zugewiesen, welche die semantische Rolle im Satz angibt. Laut [Sa12] kann das Spezifikationsdokument mit einem System, das NLP mit Semantik kombiniert, qualitativ verbessert werden. So lassen sich Mehrdeutigkeiten und Inkonsistenzen automatisch erkennen.

Im Kontext agiler Methoden kann NLP das Schreiben funktionaler Anforderungen unterstützen. [ENT18] beschreibt die automatisierte Umwandlung von User Stories in Use Case Diagramme, die in Unified Modeling Language abgebildet werden. Zunächst wird eine Textdatei mit User Stories durch einen Algorithmus um alle verzichtbaren Wörter reduziert. Anschließend wird mit Hilfe von NLP die neue Datei zerlegt und Syntaxbäume für jede User Story erstellt. Durch die Einbindung von NLP wird der Bezug von Anforderungen zum Design und die Generierung der Ergebnisse in Unified Modeling Language erleichtert. Die Anwendung von NLP im agilen RE lässt sich zudem in [Na20] finden. Ebenso werden dort die Spezifikationen von User Stories in ein Unified Modeling Language Diagramm umgewandelt.

In [Zh10] wird ein Algorithmus basierend auf ML in der Anforderungsspezifikation verwendet. Dieser Algorithmus kann Handlungsmodelle mit Quantoren und logischen Implikationen erlernen. Ausgewählte Formeln werden dabei in gelernte Handlungsmodelle umgewandelt, welche sich durch Experten optimieren lassen. Der Nutzen besteht darin, den manuellen Aufwand zu verringern.

Um die Eigenschaften und Bedeutung eines Softwaresystems zu bewerten und festzuhalten, lassen sich Ontologien zur Dokumentation eines Qualitätsschemas einsetzen. Dadurch sollen Entwickler Qualitätsaspekte leichter verstehen und messen können. Ontologie und Semantik tragen dazu bei, den Inhalt eines Softwaresystems darzustellen und die Qualitätsattribute mit den Metriken zu verknüpfen [BGL17].

3.4 Anforderungsvalidierung

Der kombinierte Einsatz von Modellprüfung und logikbasiertem Lernen soll nach [Al15] die Validierung, Diagnose und Reparatur von Anforderungsspezifikationen automatisiert unterstützen. Dadurch wird der manuelle Aufwand reduziert, sodass eine robustere Software entwickelt werden kann. Mittels Modellprüfung sollen Fehler automatisch in der formalen Beschreibung von Anforderungen erkannt und ein Gegenbeispiel erzeugt werden, welches den Fehler veranschaulicht. Da dieses Gegenbeispiel nicht erschöpfend ist, können weitere Beispiele von den Softwareentwicklern manuell ermittelt oder durch eine weitere automatisierte Analyse generiert werden. Nachdem die Gegenbeispiele erfasst sind, führt ein Algorithmus mittels logikbasierten Lernens automatisch Diagnose- und Reparaturaufgaben durch, woraus korrekt überarbeitete Beschreibungen resultieren.

Die Anforderungsvalidierung wird ebenso durch NLP unterstützt. In [Mo16] wird ein automatisiertes kollaboratives Tool basierend auf NLP vorgestellt, welches die Zusammenarbeit zwischen Stakeholder und Requirements Engineers unterstützt. Durch das Tool, das im webbasierten Texteditor Etherpad integriert ist, lassen sich Anforderungen virtuell in Echtzeit validieren und kollaborativ bearbeiten. Darüber hinaus werden Requirements Engineers durch den Einsatz von Vorlagen darin unterstützt, vollständige und korrekte Testanforderungen zu verfassen. Diese werden mit Test Cases aus einer Bibliothek mit Testmustern überprüft. Die Verarbeitung natürlicher Sprache sichert die Korrektheit und Genauigkeit der Testanforderungen und Test Cases.

4 Implikation

Auf Basis der Ergebnisse dieser SLR wird gezeigt, wie sich der Forschungsstand zum Einsatz von KI-Methoden in den einzelnen Phasen des RE gestaltet. Die Phasen der Anforderungserhebung, -analyse und -spezifikation werden durch unterschiedliche KI-Methoden unterstützt. In der Anforderungsanalyse treten NLP und ML dominant auf. Die Phase der Anforderungsspezifikation ist insbesondere durch den Einsatz von NLP geprägt. Darüber hinaus kommen in geringem Maße ML und Ontologien zur Anwendung. Grundsätzlich können im Rahmen der SLR lediglich zwei Publikationen der Anforderungsvalidierung zugeordnet werden [Al15], [Mo16]. Demnach ist die Phase der Anforderungsvalidierung vergleichbar geringfügig in der Literatur behandelt. Um weitere Aussagen diesbezüglich treffen zu können, sollte diese Phase zukünftig stärker in den Fokus von Forschungsarbeiten, insbesondere auf normativer Ebene, gerückt werden.

Im Rahmen der SLR zeichnen sich Techniken des NLP als häufig verwendete Methode ab, die in allen RE-Phasen Anwendung findet. Daraus lässt sich schließen, dass NLP in den speziellen Aufgaben im Anforderungsprozess der Softwareentwicklung vielfältig einsetzbar ist. Folglich ergibt sich ein hohes Potenzial zur automatisierten Sprachverarbeitung von Anforderungen in Textform. Allerdings zeigen sich hier Schwierigkeiten, die es bei der Integration von KI im RE zu bewältigen gibt. Die fehlerhafte und mehrdeutige Interpretationsweise der natürlichen Sprache ist Gegenstand der Forschungen. Mit Hilfe von NLU wird insbesondere die Verringerung von Redundanz, Mehrdeutigkeit, Inkonsistenz und Unvollständigkeit in Anforderungen erreicht. Die Errungenschaften in der Sprachverarbeitung können zur korrekten Extraktion der Anforderungen von Stakeholdern in verschiedenen Kontexten beitragen. Allerdings ist die automatische und fehlerfreie Extraktion von Anforderungen aus natürlichsprachlichen Beschreibungen bisher nicht vollständig gelungen. Es ist zu erwarten, dass NLP weiterentwickelt und nahezu alle Sprachen weltweit abdecken wird, wodurch ein großes Potential in der vollautomatischen Verarbeitung von natürlicher Sprache besteht. Diesbezüglich bedarf es Verbesserungen der Interpretationsfähigkeit von Worten und Phrasen, die dem natürlichsprachlichen Ursprung entstammen, um eine zuverlässige Verarbeitung der Anforderung von Stakeholdern zu gewährleisten [DS11], [Ma19], Mo16], [MX19], [Ng12], [Sa12].

Nach aktuellem Stand der wissenschaftlichen Literatur herrscht kein Konsens darüber, welche KI-Methode für welche Aufgaben im Anforderungsprozess genutzt wird [Ba20], [SP13]. So dienen in der Anforderungsanalyse NLP oder ML mehrfach zur Überprüfung der Qualität der erhobenen Anforderungen [Ca14], [Ka19], [Lu17], [Pa15]. Die SLR deutet jedoch ebenso darauf hin, dass für ähnliche Aufgabenstellungen im RE dieselbe KI-Methode angewendet wird. In der Anforderungsspezifikation trägt die Kombination von NLP mit Semantik zur Extraktion von Aktionen und Akteuren und zur Erstellung von Use Case Diagrammen bei [IAA18], [JIA18]. Darüber hinaus kann NLP im Rahmen von agilen Methoden User Stories automatisiert in Use Case Diagramme umwandeln [ENT18], [Na20]. Wie die Beispiele zeigen, könnte ein gesamtheitliches Framework zum Einsatz von KI-Methoden bei speziellen Aufgaben im Anforderungsprozess förderlich sein, um Stakeholder und Requirements Engineers adäquat zu unterstützen. Diesbezüglich ist weitere Forschung notwendig, um fundierte Aussagen darüber zu treffen, welche KI-Methoden in den vorgestellten RE-Phasen bereits möglich sind. So können weitere SLR das breite Spektrum der Forschung näher eingrenzen und zukünftigen Forschungsbedarf noch präziser ableiten. Außerdem sollten Gründe für die unterschiedlich ausgeprägte Nutzung von KI in den RE-Phasen weiter ausgearbeitet werden. Ferner ist zu berücksichtigen, dass der SLR ein breiter KI-Begriff zugrunde liegt. Dies hat einerseits den Effekt, dass das Thema weitläufig erfasst wird. Andererseits bietet es zukünftig die Möglichkeit, einzelne KI-Methoden oder RE-Phasen spezifischer zu betrachten.

Seit 1950 stellt KI ein Forschungsgebiet dar, welches sich stetig weiterentwickelt und neue Erkenntnisse hervorbringt. Die weitere Erforschung des Einsatzes von KI im RE kann auch zukünftig die komplexen Prozesse vereinfachen und die neuen Erkenntnisse im Bereich KI nutzen [Bu20]. Durch diese Forschungsarbeit wird ein Beitrag geleistet, um den

unübersichtlichen Forschungsstand zu strukturieren und diesbezüglich eine Ausgangsbasis für weitere Forschungsaktivitäten zu schaffen. Darüber hinaus werden Einsatzmöglichkeiten von verschiedenen KI-Methoden in den einzelnen Phasen des RE aufgezeigt. Dadurch trägt diese Forschungsarbeit zu einem KI-Methoden übergreifenden Überblick bei, was die Bildung einer Forschungsagenda unterstützt.

5 Literaturverzeichnis

[Ac20] Acatech, https://www.plattform-lernende-systeme.de/infografiken.html, Stand: 19.06.2020.

[Al15] Alrajeh, D. et.al.: Automated Support for Diagnosis and Repair. Communications of the ACM 02/15, S. 65-72, 2015.

[As15] Association for Computing Machinery Digital Library, https://dl.acm.org/search/advanced, Stand: 05.06.2020.

[Ba20] Barenkamp, M.: Künstliche Intelligenz in der Softwareentwicklung. Wirtschaftsinformatik & Management 12/20, S.120-129, 2020.

[BF14] Bourque, P.; Fairley, R. E.: SWEBOK V3.0: Guide to the Software Engineering Body of Knowledge. IEEE Computer Society Press 01/14, S.1-335, 2014.

[BGL17] Blas, M. J.; Gonnet, S.; Leone, H.: An ontology to document a quality scheme specification of a software product. Expert Systems 05/17, 21 Seiten, 2017.

[BMS20] Bhatia, K.; Mishra, S.; Sharma, A.: Clustering Glossary Terms Extracted from Large-Sized Software Requirements using FastText. In (ACM Hrsg.): Proc.13th Innovations in Software Engineering Conference on Formerly known as India Software Engineering Conference, S.1-11, 2020.

[BS19] Buxmann, P.; Schmidt, H.: Künstliche Intelligenz: Mit Algorithmen zum wirtschaftlichen Erfolg. Springer, Berlin, 2019.

[Bu20] Bundesregierung, https://www.bundesregierung.de/breg-de/aktuelles/meilensteine-der-entwicklung- kuenstlicher-intelligenz-1708150, Stand: 31.01.2020.

[Ca14] Casagrande, E. et.al.: NLP-KAOS for Systems Goal Elicitation: Smart Metering System Case Study. IEEE Transactions on Software Engineering 10/14, S. 941-956, 2014.

[DD10] Dube, R. R.; Dixit, S. K.: Process-oriented complete requirement engineering cycle for generic projects. In (ACM Hrsg.): Proc. 5th International Workshop on Realizing Artificial Intelligence Synergies in Software Engineering, S. 194-197, 2010.

[De15] Dermeval, D. et.al.: Applications of ontologies in requirements engineering: A systematic review of the literature. Requirements Engineering 04/16, S. 405-437, 2015.

[DR20] Dwitam, F.; Rusli, A.: User stories collection via interactive chatbot to support requirements gathering. TELKOMNIKA Telecommunication, Computing, Electronics and Control 02/20, S. 890-898, 2020.

[DS11] Deeptimahanti, D. K.; Sanyal, R.: Semi-automatic generation of UML models from natural language requirements. In (ACM Hrsg.): Proc. 4th India Software Engineering Conference, S. 165-174, 2011.

[Eb19] Ebert, C.: Systematisches Requirements Engineering: Anforderungen ermitteln, dokumentieren, analysieren und verwalten. dpunkt.verlag, 6.Auflage, Heidelberg, 2019.

[EB20] EBSCO Industries, https://connect.ebsco.com/s/article/What-Field-Codes-are-available-when-searching-EBSCO-Discovery-Service-EDS?language=en_US, Stand: 19.06.2020.

[El20] Elsevier, https://www.elsevier.com/solutions/sciencedirect, Stand: 10.06.2020.

[ENT18] Elallaoui, M.; Nafil, K.; Touahni, R.: Automatic transformation of user stories into UML use case diagrams using NLP techniques. Procedia computer science 04/18, S. 42-49, 2018.

[Er08] Ertel, W.: Grundkurs Künstliche Intelligenz: Eine praxisorientierte Einführung. Springer Vieweg, Wiesbaden, 2008.

[GAS17] Guzman, E.; Alkadhi, R.; Seyff, N.: An exploratory study of Twitter messages about software applications. Requirements Engineering 03/17, S. 387-412, 2017.

[GG18] Gupta, A.; Gupta, C.: CDBR: A semi-automated collaborative execute-before-after dependency-based requirement prioritization approach. Journal of King Saud University - Computer and Information Sciences, wird veröffentlicht, 12 Seiten, 2018.

[HK16] Husain, M. S.; Khanum, M. A.: Word Sense Disambiguation in Software Requirement Specifications Using WordNet and Association Mining Rule. In (ACM Hrsg.): Proc. 2nd International Conference on Information and Communication, S. 1-4, 2016.

[IAA18] Imam, A. T.; Al-Hroob, A.; Al-Heisa, R.: The Use of Artificial Neural Networks for Extracting Actions and Actors from Requirements Document. Information and Software Technology 09/18, S. 1-15, 2018.

[IE90] IEEE Standards Coordinating Committee: IEEE Standard Glossary of Software Engineering Terminology (IEEE Std 610.12-1990). IEEE Computer Society 12/90, S. 1-84, 1990.

[JIA18] Jebril, E. M.; Imam, A. T.; Al-Fayuomi, M.: An algorithmic approach to extract actions and actors (AAEAA). In (ACM Hrsg.): Proc. of the International Conference on Geoinformatics and Data Analysis, S. 13-17, 2018.

[JN14] Jurkiewicz, J.; Nawrocki, J.: Automated events identification in use cases. Information and Software Technology 02/15, S. 110-122, 2018.

[Ka19] Kang, Y. et.al.: A transfer learning algorithm for automatic requirement model generation. Journal of Intelligent & Fuzzy Systems 02/19, S. 1183-1191, 2019.

[Ki09] Kitchenham, B. et.al.: Systematic literature reviews in software engineering – A systematic literature review. Information and Software Technology 01/09, S. 7-15, 2009.

[Lu17] Lucassen, G. et.al.: Extracting conceptual models from user stories with Visual Narrator. Requirements Engineering 03/17, S. 339-358, 2017.

[Ma16] Mainzer, K.: Künstliche Intelligenz - Wann übernehmen die Maschinen?. Springer, Berlin und Heidelberg, 2016.

[Ma19] Madala, K.: An artificial intelligence-based model-driven approach for exposing off-nominal behaviors. In (IEEE Press Hrsg.): Proc. 2019 IEEE/ACM 41st International Conference on Software Engineering: Companion Proceedings, S. 214-217, 2019.

[Mo16] Moketar, N. A. et.al.: An automated collaborative requirements engineering tool for better validation of requirements. In (ACM Hrsg.): Proc. 31st IEEE/ACM International Conference on Automated Software Engineering, S. 864-869, 2016.

[MSP16] Misra, J.; Sengupta, S.; Podder, S.: Topic cohesion preserving requirements clustering. In (ACM Hrsg.): 2016 IEEE/ACM 5th International Workshop on Realizing Artificial Intelligence Synergies in Software Engineering (RAISE), S. 22-28, 2016.

[MX19] Memon, K. A.; Xiaoling, X.: Deciphering and Analyzing Software Requirements employing the techniques of Natural Language Processing. In (ACM Hrsg.): Proc. 4th International Conference on Mathematics and Artificial Intelligence, S. 153-156. 2019.

[Na20] Nasiri, A. et.al.: Towards a Generation of Class Diagram from User Stories in Agile Methods. Procedia Computer Science 04/20, S. 831-837, 2020.

[Ng12] Nguyen, T. H. et.al.: REInDetector: a framework for knowledge-based requirements engineering. In (ACM Hrsg.): Proc. 27th IEEE/ACM international conference on automated software engineering, S. 386-389, 2012.

[Pa10] Patsch, H.: Requirements-Engineering systematisch. Springer, 2. Auflage, Berlin und Heidelberg, 2010.

[Pa15] Parra, E. et.al.: A methodology for the classification of quality of requirements using machine learning techniques. Information and Software Technology 11/15, S. 180-195, 2015.

[Ro19] Rodrigues, A. et.al.: Enhancing context specifications for dependable adaptive systems: A data mining approach. Information and Software Technology 08/19, S. 115-131, 2019.

[Sa12] Sateli, B. et. al.: ReqWiki: a semantic system for collaborative software requirements engineering. In (ACM Hrsg.): Proc. 8th Annual International Symposium on Wikis and Open Collaboration, S. 1-4, 2012.

[SP13] Sharma, S.; Pandey, S. K.: Integrating AI techniques in requirements phase: a literature review. In (Institution of Engineering and Technology Hrsg.): Proc. 4th International IT Summit Confluence 2013, S.21-25, 2013.

[Vi10] Villata, S.: A normative multiagent approach to requirements engineering. Logic Journal of the IGPL 01/10, S. 245-274, 2010.

[We12] Wen, K. et.al.: A model based transformation paradigm for cross-language collaborations. Advanced Engineering Informatics 01/13, S. 27-37, 2012.

[Zh10] Zhuo, H. H. et.al.: Learning complex action models with quantifiers and logical implications. Artificial Intelligence 12/10, S. 1540-1569, 2010.

Machine Learning

The Impact of Domain Knowledge on Applying Machine Learning Methods to Exoplanet Detection

The-Gia Leo Nguyen [1]

Abstract: Exoplanets do not emit electromagnetic waves which makes it challenging to detect them. Based on transit photometry, we trained a neural network on NASA Kepler space telescope data to detect exoplanets based on light intensity curves. We showcase, that with a well designed data pipeline, a small neural network is sufficient to achieve state-of-the-art performance, saving both computation time and hardware cost. The strongest improvement in performance could only be achieved by adding domain specific processing steps to the data pipeline. Domain knowledge was essential in selecting the appropriate machine learning concepts that are beneficial to solving the problem and have a higher impact on the performance than the actual classification method itself. We encourage to consider the data pipeline as an additional component, besides the classification model, that can potentially improve the overall performance.

Keywords: Machine Learning; Neural Networks; Deep Learning; Astrophysics; Exoplanets; Artificial Intelligence

1 Introduction

The idea of exoplanets was proposed early on. Dating back to the 16th century, Giordano Bruno suggested that fixed stars are similar to the Sun and likewise would have planets orbiting them. However, exoplanets do not emit light on their own and given the proximity to their star, it is hardly possible to detect them directly. The first confirmed exoplanet detection only occurred in 1992. According to NASA [2], 4375 exoplanet detections have been confirmed of which over 80% were detected in the last seven years.

Given the enormous amount of stars in the universe, we want to use machine learning methods to recommend interesting stars for further investigation by astrophysicists. Our model will be trained on data from the NASA Kepler missions. This approach is rationalized by two main observations. First, the NASA Kepler space telescope is specifically designed for detecting exoplanets with transit photometry and has detected over 60% of all confirmed exoplanets. Second, 76% of all exoplanet detections have been made by transit photometry. This indicates that training a machine learning model to make predictions based on light intensity curves used for transit photometry might be the most promising approach.

[1] Heidelberg University, Germany
[2] as of 7th of April 2021

Besides achieving state-of-the-art performance, we also analyze in depth the effectiveness of commonly used machine learning approaches to boost performance in the application of exoplanet detection. Our investigation shows the importance of domain knowledge when applying machine learning methods to real world problems. In most cases, machine learning methods will not provide the desired results *out of the box*.

2 Related Works

The idea of applying machine learning and deep learning methods to the problem of exoplanet detection is not new and has been proposed by several researchers. Zucker et al. [ZG18] conducted a feasibility study on the application of Convolutional Neural Networks on exoplanet detection with simulated data. Further works that applied Convolutional Neural Networks to exoplanet detection have been done by [SV18], [YJ21] and [CJ19]. Ensemble models for exoplanet detection have been introduced by Priyadarshini et al. [PP21] with Ensemble-CNNs and Malik et al. [MMO20] with gradient boosting classifiers. Sturrock et al. [SMR] and Schanche et al. [Sc19] analyzed and compared the application of several machine learning methods on exoplanet detection, including Support Vector Machines, K-Nearest neighbor, Random Forest Classifier, Logistic Regression Classifier and Convolutional Neural Networks. Ofman et al. [Of21] applied ThetaRay Fintech algorithms, originally used for anomaly detection in financial institutions, on exoplanet detection. Tsang et al. [TS19] proposed a recurrent neural network autoencoder for unsupervised feature extraction combined with an estimation network for supervised classification and novelty detection. Jara-Maldonado et al. [Ja20] surveyed different machine learning methods for exoplanet detection and propose a model to generate synthetic light curves which can be used as training data. Mislis et al. [MPA18] introduced an unsupervised method for exoplanet detection using the DBSCAN clustering algorithm.

Most of previous research has been focusing on the classification model itself. However, not much attention has been given to the data pipeline that comes before the classification model. In this work, we analyze different pipelines and their effect on the overall performance. In particular, we will investigate the effect of different oversampling methods, data augmentation and domain-specific processing steps. Additionally, we show that most machine learning methods do not work *out of the box* and are highly problem dependent. This makes domain knowledge essential for selecting the appropriate data pipeline for machine learning based methods.

3 Theory

3.1 Transit Photometry

Transit photometry uses the fact that a planet passing directly between its star and the observer would dim the star's light periodically resulting in measurable dips of brightness.

Our goal is to design an algorithm using machine learning techniques to recognizes specific patterns in the light intensity curve of stars accompanied by exoplanets and classify them as such. The probability of a grazing transit or full transit to be observable by a spectator for randomly oriented orbits is given by:

$$P\left(cos(i) < \frac{R_s + R_p}{a}\right) = \frac{1}{2}\int_{-\frac{R_s+R_p}{a}}^{\frac{R_s+R_p}{a}} = \frac{R_s + R_p}{a} \quad (1)$$

where R_p is the radius of the exoplanet, R_s the radius of the hosting star, a the semi-major axis of the orbit and i the angle between the line of sight and the angular-momentum vector of the exoplanet's orbit. Despite transit photometry being the most successful method for detecting exoplanets, the probability of an exoplanet actually orbiting its star in such a way, that it is observable to the spectator, is not exceptionally high. This makes exoplanets one of the most difficult objects to detect in the universe. As an example the transit of Earth is only visible from 0.46 percent of the celestial sphere [BMS16].

3.2 Imbalanced Dataset

Since exoplanets are hard to detect and only 4375 confirmed exoplanet detections [3] have been made so far, we expect our dataset to be highly imbalanced. Assuming a dataset with only 1% of the data points being stars with confirmed exoplanets, an algorithm that classifies every star to not have any exoplanets would already achieve a prediction accuracy of 99%. Even though, accuracy is commonly used as a performance metric in machine learning classification tasks, it will not be very indicative of the real performance in our application. For imbalance datasets, precision [4] and recall [5] are often used instead. In our experiments, we will consider the F1 score as our main performance metric which is the harmonic mean between precision and recall:

$$\text{F1 score} = 2 \cdot \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (2)$$

Choosing an appropriate performance metric alone, will not prevent the classification model from always predicting the majority class. A typical approach to deal with an imbalanced dataset, is to oversample the minority class. The most naive approach is random oversampling. By creating copies of random samples from the minority class, our classification model will see those samples more often during the training process. A more sophisticated way to oversample the minority class was introduced by Chawla et al. [Ch11] with the SMOTE algorithm. The main idea is to adjust the class distribution of the dataset by generating new data points of the minority class. To generate a new data point x_{new}, we randomly pick a

[3] as of 7th of April 2021
[4] ratio of true positives against all examples that were classified as true
[5] ratio of true positives against all positive examples

data point $x_i \in \{x_1, ..., x_n\}$, where $\{x_1, ..., x_n\}$ is the set of all data points which belongs to the minority class. Then, we compute the k-nearest neighbors of x_i in the feature space and randomly select one of the k-nearest neighbor as our data point x_{zi}. The last step includes an interpolation, where we generate the new data point x_{new} such that it is located on an imaginary line connecting the data points x_i and x_{zi}:

$$x_{new} = x_i + \lambda \cdot (x_{zi} - x_i) \tag{3}$$

where λ is randomly generated to be in the range $[0, 1]$. Two variants of the original SMOTE algorithm propose to generate new points at the border of the minority class, since those points will make the most impact on the decision boundary. One of the variants proposed by Han et al. [HWM05] uses K-nearest neighbors to identify borderline points. In the following, this method will be referred to as Borderline-SMOTE. The other variant proposed by Nguyen et al. [NCK11] uses support vector machines instead of K-nearest neighbors and will be referred to as SVM-SMOTE. The ADASYN oversampling algorithm proposed by He et al. [He08] follows the same idea as SMOTE. It also emphasizes borderline points, but the main difference is the noise added to the generated points. The idea is that removing the linear dependency from existing data points allows for generation of more realistic samples.

3.3 Data Augmentation

For machine learning methods, the quality and amount of data available plays an important role. Data augmentation is a common approach to boost performance by generating more synthetic data from the existing dataset. For time series data, Guennec et al. [GMT16] have shown in their experiments that *window sliding* and *time warping* are viable data augmentation methods for time series classification tasks using neural networks. The main idea of *window sliding* is to slide a window of smaller size over the time series creating new shorter time slices. These slices carry the same label as its source time series. At testing time, we will slide a window over the testing time series and perform classification on each window. The final prediction is then decided by majority vote. The main idea of *time warping* is to select random slices which will be either stretched or compressed. Additionally, Gaussian noise will be added to randomly selected data points.

3.4 Fully-connected Neural Network

The goal of our neural network is to learn an approximated mapping f between the light curves x from NASA Kepler Telescope and the labels y denoting if the star has an exoplanet. A neural network can be interpreted a directed acyclic graph, where each node n is a computational unit often referred to as a *neuron*. Each node n takes an input vector I and

performs the following computation with its internal weights (W, b) [6] to output a single output value o:

$$o = \phi(W^T \cdot I + b) \tag{4}$$

where ϕ is an arbitrary non-linear function called *activation function*. In our neural network architecture the ReLU [NH10] is used as an activation function for all neurons. Each layer in a neural network is constructed by stacking multiple neurons. As a result, the output of each layer is a vector, where every entry is the output scalar value o_i of the corresponding i-th neuron in the layer. This output vector can then be given as input for the next neural network layer. Using *softmax* and *cross entropy* as our loss function, we can compare the neural network output $\hat{y} = f(x, \theta)$ [7] with the ground truth label y. Adam optimization [KB14] will be used to minimize our loss function by optimizing the weights θ and is often referred to as *training*. At testing time, our model is evaluated on previously unseen data, not used during training, on both accuracy and F1 score to determine its generalization ability.

4 Methodology

4.1 Dataset

For our experiments, we will be using a Kaggle dataset [8] which contains data from NASA Kepler space telescope Campaign-3 mission taken from Mikulski Archive for Space Telescopes [9]. According to the creator of the Kaggle dataset, Campaign-3 was picked because it is unlikely to contain any undiscovered exoplanet-stars and therefore would minimize the probability of wrongly labeled data points. Additionally, the number of stars with exoplanets was boosted by including data from other Kepler Campaigns. Even after boosting, the dataset is still highly unbalanced with over 99% of the data being observations of stars without exoplanets. The trainset consists of 5078 flux diagrams of which only 37 are confirmed exoplanet-stars. The testset consists of 570 flux diagrams of which only 5 are confirmed exoplanet-stars. However, we will not use the train-test-split provided by Kaggle for our experiments, but instead perform stratified k-fold cross-validation. The main reason is that the train-test split provided on Kaggle with a ratio of 90-10 is prone to overfitting and selection bias, which might not represent the true generalization ability of our model. Instead with stratified k-fold cross-validation, we will split our dataset into 10 equal subsets with approximately the same ratio between confirmed exoplanet stars and non-exoplanet stars. The training and testing procedure will then be executed 10 times, where each time another subset is chosen as the testset and the remaining 9 subsets will be used as training data. The performance is then calculated by taking the mean of all 10 runs.

[6] W is a vector and b is a scalar
[7] θ are the weights of the whole neural network consisting of all (W, b) of all neurons combined
[8] https://www.kaggle.com/keplersmachines/kepler-labelled-time-series-data
[9] https://archive.stsci.edu/

4.2 Architecture and Training

The classification model will be kept relatively small with a 3-layer neural network [GBC16] of hidden size 100 with dropout [Sr14], batch normalization [IS15] and ReLU activation functions [NH10]. This model was design such that it can be trained in reasonable time on a CPU [10] without the need for any GPUs. As a comparison the AstroNet by Shallue et al. [SV18] consists of 14 convolutional layers, 5 fully-connected layers and 7 pooling layers. A smaller network has a lower representation ability, but requires significantly less resources to train and evaluate. We initialize the weights with Xavier initialization. For training we use Adam optimization [KB14] with a learning rate of 0.01 and a batch size of 1024 trained for a total of 100 epochs. We employ adaptive learning by reducing the learning rate by a factor 0.5 at epoch 20, 40, 55, 70, 80, 90 and 95. Further implementation details can be deducted from our code [11].

4.3 Domain specific pre-processing

On every flux diagram, we will independently apply normalization and standardization, since we are only interested in the periodically reoccurring dips and not in the light curves amplitude. Note that the normalization is applied per flux diagram and not over the whole dataset. Besides normalization and standardization, we will apply domain specific pre-processing steps using domain knowledge. Analyzing its performance impact can give insight into the significance of domain knowledge for applying machine learning methods to real world problems.

In transit photometry, we are looking for periodically reoccurring dips whenever an exoplanet passes the line of sight between star and telescope. We propose to apply a Fourier transform [Ob07] to the flux curves. For an input vector $(x_0, ..., x_{2n-1})$ with a dimension of $2n$ the discrete Fourier transformed output vector $(f_0, ..., f_{2n-1})$ is given by:

$$f_m = \sum_{k=0}^{2n-1} x_k \cdot exp\left(-\frac{2\pi i}{2n} mk\right) \tag{5}$$

where $m \in \{0, ..., 2n-1\}$. This means that the light curves, which are functions of time, will be transformed into functions of frequency. In other words, those Fourier transformed functions measure how prominent certain frequencies appear in the original light curve. We argue that this frequency domain makes it much easier for the classification model to detect the periodical dips, since periodicity and frequency are inherently connected. Furthermore, by only considering the absolute value of the Fourier transform, we also compensate for possible time shifts between data from different stars, since phase information is ignored.

[10] under 2 minutes on a AMD Ryzen 7

[11] https://github.com/Xenovortex/exoplanet_detection

Based on transit photometry, we expect exoplanets to create periodically occurring dips of flux in the negative direction. Therefore, we will only remove outliers in the positive direction with a deviation higher than 5σ to avoid removing sensible effects introduced by the presence of an exoplanets. Note that the outlier removal only applies to points within a flux diagram. The total number of flux diagrams stays the same and all flux diagrams in the dataset were used in the conducted experiments. Since the obtained data after Fourier transform is still quite noisy, we will apply a Gaussian filter afterwards.

5 Results

5.1 Domain Knowledge and Oversampling

We evaluate the performance of all oversampling techniques with different pre-processing settings as presented in table 1. The best performing oversampling method depends on the pre-processing steps performed on the data. However, we find that overall the performance depends more on the pre-processing pipeline than the actual oversampling method.

The main improvements can be observed through pre-processing, where Fourier Transform plays an essential part. Without Fourier Transform, the F1 score will fall under 0.2 irrespective of other included pre-processing steps. However, for optimal performance both Fourier Transform and Gaussian Smoothing are required independent from the oversampling method used. These results are aligned with our expectations. Based on transit photometry, we were looking for periodic dips in the flux curves. This makes the Fourier frequency space the perfect feature space to train our model on and at the same time acts as a normalization on the flux curves.

In table 2, we replace our neural network classification model by other machine learning methods such as K-nearest Neighbor, Support Vector Machine and Random Forest Classifiers and show that our proposed inclusion of Fourier transform and Gaussian Smoothing into the data pipeline can significantly improve performance for a variety of machine learning based classification models. With a performance increase of about 25% to 60% in F1 Score, the classical machine learning methods show less improvements than for the fully-connected neural network model. We hypothesize that our neural network model has a higher representation power than the classical machine learning methods and thereby is more prone to overfitting by picking up unimportant details in the data, that are not related to the classification task. Both Fourier Transform and Gaussian Smoothing will help remove those unrelated details by changing the feature space, since the frequency domain directly represent the periodicity we are looking for.

Oversampling Technique	Test Accuracy	Test F1 Score
Without Pre-processing		
Without Oversampling	0.993	0.000
Random	0.987	0.100
SMOTE	0.986	0.141
Borderline-SMOTE	0.988	0.165
SVM-SMOTE	0.987	0.102
ADASYN	0.987	0.110
Normalization + Outlier Removal		
Without Oversampling	0.992	0.040
Random	0.992	0.033
SMOTE	0.991	0.123
Borderline-SMOTE	0.991	0.135
SVM-SMOTE	0.992	0.062
ADASYN	0.992	0.124
Normalization + Outlier Removal + Fourier Transform		
Without Oversampling	0.996	0.655
Random	0.995	0.510
SMOTE	0.995	0.470
Borderline-SMOTE	0.995	0.492
SVM-SMOTE	0.995	0.573
ADASYN	0.995	0.528
Normalization + Outlier Removal + Gaussian Smoothing		
Without Oversampling	0.992	0.000
Random	0.993	0.100
SMOTE	0.991	0.089
Borderline-SMOTE	0.991	0.000
SVM-SMOTE	0.993	0.089
ADASYN	0.993	0.186
Normalization + Outlier Removal + Fourier Transform + Gaussian Smoothing		
Without Oversampling	0.998	0.810
Random	0.999	0.843
SMOTE	0.999	0.857
Borderline-SMOTE	0.998	0.824
SVM-SMOTE	0.998	0.790
ADASYN	**0.999**	**0.905**

Tab. 1: Mean Test Accuracy and Mean Test F1 Score for different pre-processing pipelines and oversampling techniques using stratified k-fold cross-validation with $k = 10$

Method	With FT and GS		Without FT and GS	
	Test Accuracy	Test F1 Score	Test Accuracy	Test F1 Score
K-nearest Neighbor	0.995	0.725	0.970	0.581
Support Vector Machines	0.998	0.868	0.992	0.535
Random Forest	0.996	0.731	0.895	0.494

Tab. 2: Mean Test Accuracy and Mean Test F1 Score comparison on different machine learning methods with and without Fourier transform (FT) + Gaussian Smoothing (GS) using stratified k-fold cross-validation with $k = 10$

5.2 Data Augmentation

Table 3 shows our results for the window sliding experiment. We observe a significant performance decrease, when using window sliding irrespective of the chosen window size. Despite window sliding being a valid data augmentation technique in the context of machine learning, we have to consider domain knowledge when applying machine learning methods to real world problems. From the physics perspective, we know through transit photometry that we are looking for periodical dips in flux when an exoplanet is between the observer and the star. This requires the star to be observed for at least a multiple of the exoplanet's orbital period. For context, the orbital period of Earth is roughly 1 year and for Neptune 165 years. Using window sliding, the neural network will only look at a small window, making all exoplanets with orbital periods larger than the window size undetectable.

The results for time warping are summarized in table 4. Time warping will be applied with a given probability to randomly selected segments during training. We observe a performance decrease for higher probabilities. Therefore, we conclude that there is no apparent advantage in using time warping, since it only worsen the performance. From the physics perspective, we assume that randomly stretching and compressing parts of the time series will interfere with the periodicity of the dips.

Window Size	Test Accuracy	Test F1 Score
Without WS	**0.999**	**0.905**
100	0.976	0.003
500	0.993	0.000
1000	0.993	0.000

Tab. 3: Mean Test Accuracy and Mean Test F1 Score for Window Sliding (WS) using stratified k-fold cross-validation with $k = 10$ and the full pre-processing pipeline (Normalization + Outlier Removal + Fourier Transform + Gaussian Smoothing)

5.3 Comparison to other Kaggle solutions

For the comparison with other Kaggle solutions, we will employ normalization, outlier removal, Fourier transform and Gaussian smoothing as pre-processing steps with ADASYN

Probability	Test Accuracy	Test F1 Score
0	**0.999**	**0.905**
0.1	0.998	0.813
0.3	0.997	0.649
0.8	0.995	0.378

Tab. 4: Mean Test Accuracy and Mean Test F1 Score for Time Warping using stratified k-fold cross-validation with $k = 10$ and the full pre-processing pipeline (Normalization + Outlier Removal + Fourier Transform + Gaussian Smoothing)

for oversampling. We do not use any data augmentation, since our experiments show that they only negatively impinge the performance. Since most of the solutions on Kaggle do not use stratified k-fold cross-validation, but use the train-test split provided on Kaggle, we additionally evaluate our model on the given Kaggle split. Table 5 shows our test performance in comparison with other top solutions retrieve from Kaggle. Our method beats every solution on both accuracy and F1 score achieving state-of-the-art performance. Even our stratified k-fold cross-validation results still beat all other methods. This showcase that using domain knowledge to design appropriate pre-processing pipelines can greatly improve the performance of machine learning models.

Approach	Test Accuracy	Test F1 Score
Juan Felipe (XGBoost)	–	0.88
Travillion (SVC)	–	0.73
Gabriel Garza (SVC)	–	0.71
Alexei D. (SVC)	–	0.67
Rahul Singh (Random Forest)	0.991	0.5
Peter Grenholm (CNN)	0.998	–
Rahul Misal (Naive Bayes)	0.991	–
Amajid Sinar (CNN)	0.991	–
Keyur Paralkar (MLP)	0.990	–
Ours (stratified k-fold)	**0.999**	**0.905**
Ours (Kaggle split)	**1.000**	**1.000**

Tab. 5: Performance Comparison of our method and top solutions on Kaggle

6 Conclusion

Through this work, we have demonstrated the importance of domain knowledge when applying machine learning methods to real world problems. Most of the time, established machine learning methods are very case dependent and can not be applied *out of the box* to every problem. This was apparent in our data augmentation experiments. Even though, *window sliding* and *time warping* are sensible approaches in the machine learning context, they are not constructive when considering the astrophysics behind exoplanet detection with transit photometry. In contrast, having domain knowledge on transit photometry allows us to

identify effective processing steps such as Fourier transform that allows for state-of-the-art performance with a smaller neural network architecture. In summary, we successfully applied machine learning techniques on the recent research topic of exoplanet detection. Our model achieves state-of-the-art performances beating all current approaches on the same dataset. As a use case, we see your model as a pre-selection algorithm that can recommend interesting star for further inspection by an expert astrophysicist.

7 Acknowledgment

We thank Paul Hill, Ullrich Köthe and Tsungnan Lin for the inspiration to this project. We thank Holger Fröning for bringing the SKILL conference to our attention. Furthermore, we thank the reviewers for the constructive criticism and recommendations.

Bibliography

[BMS16] Bozza, Valerio; Mancini, Luigi; Sozzetti, Alessandro: Methods of Detecting Exoplanets: 1st Advanced School on Exoplanetary Science. 01 2016.

[Ch11] Chawla, N. V.; Bowyer, K. W.; Hall, L. O.; Kegelmeyer, W. P.: SMOTE: Synthetic Minority Over-sampling Technique. arXiv e-prints, p. arXiv:1106.1813, June 2011.

[CJ19] Chintarungruangchai, Pattana; Jiang, Ing-Guey: Detecting Exoplanet Transits through Machine-learning Techniques with Convolutional Neural Networks. Publications of the Astronomical Society of the Pacific, 131(1000):064502, June 2019.

[GBC16] Goodfellow, Ian; Bengio, Yoshua; Courville, Aaron: Deep Learning. MIT Press, 2016. `http://www.deeplearningbook.org`.

[GMT16] Guennec, Arthur Le; Malinowski, Simon; Tavenard, Romain: , Data Augmentation for Time Series Classification using Convolutional Neural Networks, 2016.

[He08] He, Haibo; Bai, Yang; Garcia, Edwardo; Li, Shutao: ADASYN: Adaptive Synthetic Sampling Approach for Imbalanced Learning. pp. 1322 – 1328, 07 2008.

[HWM05] Han, Hui; Wang, Wen-Yuan; Mao, Bing-Huan: Borderline-SMOTE: A New Over-Sampling Method in Imbalanced Data Sets Learning. volume 3644, pp. 878–887, 09 2005.

[IS15] Ioffe, Sergey; Szegedy, Christian: Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. arXiv e-prints, p. arXiv:1502.03167, February 2015.

[Ja20] Jara-Maldonado, Miguel; Alarcon-Aquino, Vicente; Rosas-Romero, Roberto; Starostenko, Oleg; Ramirez-Cortes, Juan: Transiting Exoplanet Discovery Using Machine Learning Techniques: A Survey. Earth Science Informatics, 09 2020.

[KB14] Kingma, Diederik P.; Ba, Jimmy: Adam: A Method for Stochastic Optimization. arXiv e-prints, p. arXiv:1412.6980, December 2014.

[MMO20] Malik, Abhishek; Moster, Benjamin P.; Obermeier, Christian: Exoplanet Detection using Machine Learning. arXiv e-prints, p. arXiv:2011.14135, November 2020.

[MPA18] Mislis, D.; Pyrzas, S.; Alsubai, K. A.: TSARDI: a Machine Learning data rejection algorithm for transiting exoplanet light curves. Monthly Notices of the Royal Astronomical Society, 481(2):1624–1630, December 2018.

[NCK11] Nguyen, Hien M.; Cooper, E.; Kamei, K.: Borderline over-sampling for imbalanced data classification. Int. J. Knowl. Eng. Soft Data Paradigms, 3:4–21, 2011.

[NH10] Nair, Vinod; Hinton, Geoffrey: Rectified Linear Units Improve Restricted Boltzmann Machines Vinod Nair. volume 27, pp. 807–814, 06 2010.

[Ob07] Oberst, Ulrich: The Fast Fourier Transform. SIAM J. Control and Optimization, 46:496–540, 01 2007.

[Of21] Ofman, Leon; Averbuch, Amir; Shliselberg, Adi; Benaun, Idan; Segev, David; Rissman, Aron: Automated identification of transiting exoplanet candidates in NASA Transiting Exoplanets Survey Satellite (TESS) data with machine learning methods. arXiv e-prints, p. arXiv:2102.10326, February 2021.

[PP21] Priyadarshini, Ishaani; Puri, Vikram: A convolutional neural network (CNN) based ensemble model for exoplanet detection. Earth Science Informatics, 14:1–13, 06 2021.

[Sc19] Schanche, N.; Collier Cameron, A.; Hébrard, G.; Nielsen, L.; Triaud, A. H. M. J.; Almenara, J. M.; Alsubai, K. A.; Anderson, D. R.; Armstrong, D. J.; Barros, S. C. C.; Bouchy, F.; Boumis, P.; Brown, D. J. A.; Faedi, F.; Hay, K.; Hebb, L.; Kiefer, F.; Mancini, L.; Maxted, P. F. L.; Palle, E.; Pollacco, D. L.; Queloz, D.; Smalley, B.; Udry, S.; West, R.; Wheatley, P. J.: Machine-learning approaches to exoplanet transit detection and candidate validation in wide-field ground-based surveys. Monthly Notices of the Royal Astronomical Society, 483(4):5534–5547, March 2019.

[SMR] Sturrock, George Clayton; Manry, Brychan; Rafiqi, Sohail: Machine Learning Pipeline for Exoplanet Classification. SMU Data Science Review, 2(1 , Article 9).

[Sr14] Srivastava, Nitish; Hinton, Geoffrey; Krizhevsky, Alex; Sutskever, Ilya; Salakhutdinov, Ruslan: Dropout: A Simple Way to Prevent Neural Networks from Overfitting. Journal of Machine Learning Research, 15(56):1929–1958, 2014.

[SV18] Shallue, Christopher J.; Vanderburg, Andrew: Identifying Exoplanets with Deep Learning: A Five-planet Resonant Chain around Kepler-80 and an Eighth Planet around Kepler-90. Astronomical Journal, 155(2):94, February 2018.

[TS19] Tsang, Benny T. H.; Schultz, William C.: Deep Neural Network Classifier for Variable Stars with Novelty Detection Capability. Astrophysical Journal Letters, 877(2):L14, June 2019.

[YJ21] Yeh, Li-Chin; Jiang, Ing-Guey: Searching for Possible Exoplanet Transits from BRITE Data through a Machine Learning Technique. Publications of the Astronomical Society of the Pacific, 133(1019):014401, January 2021.

[ZG18] Zucker, Shay; Giryes, Raja: Shallow Transits - Deep Learning. I. Feasibility Study of Deep Learning to Detect Periodic Transits of Exoplanets. Astronomical Journal, 155(4):147, April 2018.

Bicycle Detection from Top View Perspective in Surveillance System using Convolutional Neural Network

Sanal Darshid Ramkumar[1]

Abstract: Bicycle detection and tracking with cameras from top view perspectives using deep learning is a highly active research area for video surveillance and automatic ticket generation in Advanced Public Transportation System (APTS). People detection using conventional cameras has received massive attention for video surveillance inside public transportation systems but inattentive towards bicycle detection. Experimentation is performed on You Only Look Once (YOLO), Faster Regional-Convolutional Neural Network (Faster R-CNN) and Single Shot Multibox Detector (SSD). Due to the sparse availability of dataset for this work, a customized dataset was recorded in the Media Computing lab, Junior Professorship of Media Computing, TU Chemnitz, Germany. True positive and false positive analysis are done to find the best case solutions to reduce the problems affecting the performance of deep learning models due to occlusions and view point variations. This paper provides best case solution for bicycle detection from a top view perspective and has achieved mean average precision of 92 %.

Keywords: YOLO, Faster R-CNN, SSD

1 Introduction

Humans and computers have inborn conditions suited for performing different types of tasks. For example, calculating the cube root of a large number is very easy for a computer, but difficult for humans. The human visual system is fast and accurate, allowing us to perform complex tasks like driving with little conscious thought [Fe10]. On the other hand, a task such as recognizing the objects in an image is easy for humans but has traditionally been very difficult for an automated learning algorithm. The term object detection refers to object localization and classification [Fe10]. Generic object detection methods are based on region proposal and regression/ classification [Zh19].
Closed Circuit Television (CCTV)/ Internet Protocol (IP) cameras and network infrastructure have become cheaper and more affordable. Thus providing a new surveillance technology that applies to a wide range of end-users in retail sectors, schools, homes, office, industrial/ transportation systems, and government sectors. Over the years, with the advances of deep learning, more concretely Convolutional Neural Networks (CNN) [HS06], image recognition and object detection have been progressing at a rapid pace. Most of this progress is not just the result of powerful hardware, bigger datasets and models, but mainly a consequence

[1] Chemnitz University of Technology, Junior Professorship Media Computing, Straße der Nationen 62, 09111 Chemnitz, Germany sanal-darshid.ramkumar@informatik.tu-chemnitz.de

of new ideas, algorithms and improved network architectures. The real-world detection and counting of a bicycle are hard, which involves locating bicycle in different scenes irrespective of orientation, scale, environment and type of view.

The paper is structured in the following sections as Related work 2, Dataset preparation 3, Experimental setup 4, Implementation 5 and Results 6.

2 Related Work

This section provides overview of object detection in public transportation systems, as well as the challenges for bicycle detection.

An APTS is an application of Intelligent Transportation Systems (ITS) [Zh18] which aims to provide services relating to different modes of transport and traffic management resulting in a safer, more coordinated and smarter use of transport networks. Research works are focused on fields like calling for emergency services when an accident occurs, using cameras to enforce traffic laws or signs that mark speed limit changes depending on conditions, usage of CCTV for surveillance. The directive of the European Union (EU) 2010/40, on July 7, 2010, defined ITS [Li18] as a system that applies information and communication technologies in the field of road transport, infrastructure, vehicle users, traffic and mobility management [Le89]. High demand is seen in today's surveillance systems in APTS for detection and counting of people and bicycle for surveillance, automatic ticket generation and to display the availability of free slots for parking bicycles in public transport systems such as bus and train [VM19].

Many research works are focused on people detection but countable in terms of bicycle detection for applications in public transportation systems. Cameras deployed in top view perspective to monitor a large scene with limited resources for people and bicycle detection are crucial. Nevertheless, the problem has received only limited attention, and state of the art lags behind bicycle detection in conventional top view perspectives. A research work based on traditional machine learning [Sh15] is available but limited to viewpoint variation. Research work in APTS for bicycle and motorcycle detection discussed on effect of different climate, clearly this paper focuses on traffic monitoring from the data gathered from CCTV [De21]. Another work towards safety of cyclist focuses on orientation using deep learning in the field of autonomous vehicles [GMC20]. The previous research works lags in identifying effects of radial distortion, elevated viewpoint and deformation from the top view, extensive variability of bicycle appearance from top view leads to difficulty in obtaining features for accurate detection. These features are sensitive to noise, occlusion of multiple bicycles, illumination changes and viewpoint variations. Multiple cameras deployed in the top view perspective might provide a better solution for bicycle detection.

In this work experimentations are done on YOLOv2 using Darknet and PyTorch [Sa18], [Op18], YOLOv3 using Darknet and PyTorch [RF18], [Op18], SSD using Inception v2 [Li16], Faster R-CNN using Inception v2, ResNet-50 and ResNet-101[Re16] architecture.

3 Dataset Preparation

For surveillance systems, privacy and comfort are key factors to be considered, as they should not impose pressure on anyone. This research work demands a dataset from a top view perspective for applications in indoor and outdoor surveillance systems for bicycle detection and tracking. As there is a shortage of publicly available dataset, a customised dataset was created. This customised dataset is created concerning laboratory and bicycle parking scenarios where humans take bicycle along with them to the respective areas. The customised dataset (DS1 and DS2) are recorded in the Media Computing lab, Junior Professorship of Media Computing, TU Chemnitz, Germany. Visual information is retrieved by recording videos in High Definition (HD) resolution from 10 optical smart embedded stereo sensors (S2000, Intenta GmbH). To further enhance the quantity of the dataset two more datasets (DS3 and DS4) recorded using mobile camera were included. Data augmentation like flipping, scaling and cropping are performed on DS3 and DS4.
Videos were converted to images and labelled using the Computer Vision Annotation Tool (CVAT) [BM]. Sample images available in the dataset is depicted in figure 1. Key facts and typical differences are presented in table 1.

- DS1 - Lab dataset consisting of 207 videos recorded using a wide-angle smart stereo sensor (S2000, Intenta GmbH) mounted in top view perspective with bicycle and human in an indoor scene. Sample images are shown in the first row of figure 1
- DS2 - Lab dataset consisting of 232 videos recorded using a wide-angle smart stereo sensor (S2000, Intenta GmbH) mounted in top view perspective with bicycle and human in an outdoor scene. Sample images are shown in the second row of figure 1
- DS3 - An indoor dataset consisting of 15 videos recorded using a mobile camera in top view perspective with bicycle and human in a bicycle garage. Sample images are shown in the third row of figure 1
- DS4 -An outdoor dataset consisting of 15 videos recorded using a mobile camera in top view perspective with bicycle and human near the TU Chemnitz university area. Sample images are shown in the fourth row of figure 1

Dataset	**Application**	**Field of view (degree)**	**Position of Camera from ground (meter)**	**Total Camera Deployed**	**Videos Generated**	**Average Video length**
1	Indoor Lab Scene	97	5	10	207	20 sec
2	Outdoor Lab Scene	97	5	1	232	31 sec
3	Indoor Bicycle garage	64	6.5	1	15	50 sec
4	Outdoor University area	64	6.5	1	15	50 sec

Tab. 1: Dataset Overview

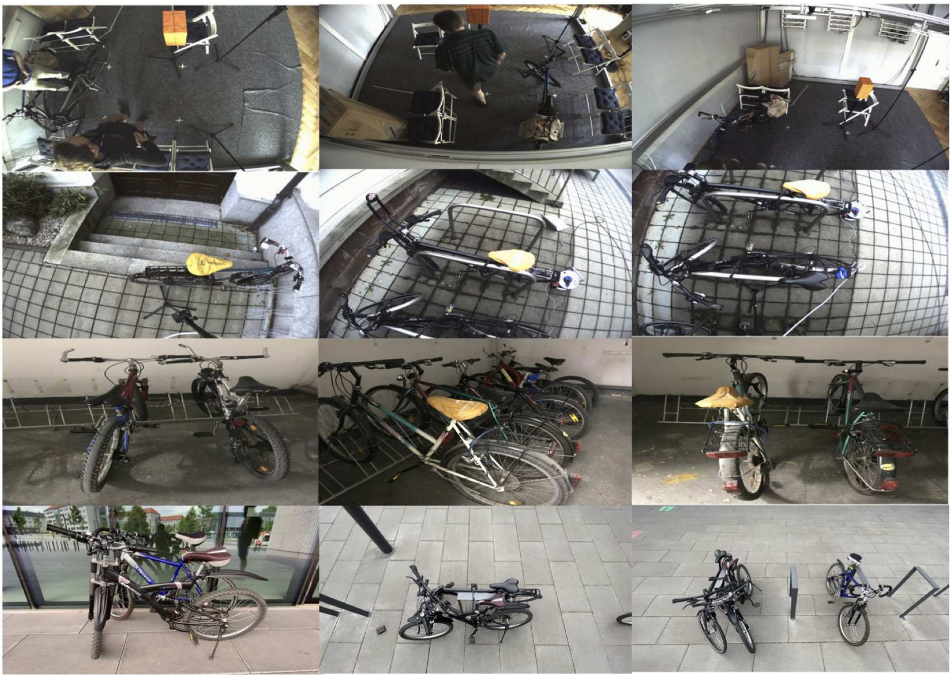


Fig. 1: The images in the first row shows DS1 dataset, the second row shows the DS2 dataset, the third row shows the DS3 dataset and the fourth row shows DS4 dataset

Combinations of dataset for training

- The dataset is split into train (80%), validation (10%) and test (10%)
- For applications related to indoor surveillance system, DS1 and DS3 were combined
- For applications related to outdoor surveillance system, DS2 and DS4 were combined
- For general purpose surveillance system application, all four dataset were combined

4 Experimental Setup

For this research work the following hardware and software configurations are used.

Hardware Configuration Intel Core i7 with 8GB RAM, CPU (4 cores at 3.60 GHz) with Ubuntu Bionic Beaver as operating system is used. NVIDIA GeForce RTX 2070 SUPER 8GB is the GPU used.
Software Configuration Python 3 with cuda 10.0 tool kit for GPU accelaration and cuDNN 7.5 libraries are used. TensorFlow, Darknet and PyTorch are the framework used.

Dependencies such as numpy 1.14.2, scipy 1.0.0, python-openCV, matplotlib 2.2.0 and cython 0.29.2 are used.

5 Implementation

This section describes in detail the implementation of all the eight models introduced in section 2. Furthermore, this facilitates to benchmark the obtained results in bicycle detection.

5.1 Configuration setup

The configuration details of the eight models, YOLOv2 (Darknet and PyTorch), YOLOv3 (Darknet and PyTorch), SSD (Inception v2), Faster R-CNN (Inception v2, ResNet-50 and ResNet-101) are shown in the table 2. From the table 2, the Faster R-CNN models used a batch size of 1 and this restriction was due to the hardware configuration limiting the conversion of tensor shape. Table 3 shows the details of the number of training and validation images used for the different dataset combinations used in this research work. For this research work, the default settings of hyperparameters worked well.

S.no	Model	Architecture	(Height,Width, Channel)	Batch size	Epochs
1	Faster R-CNN [Re16]	Inception v2	(600,1024,3)	1	200
2	Faster R-CNN [Re16]	ResNet-50	(600,1024,3)	1	200
3	Faster R-CNN [Re16]	ResNet-101	(600,1024,3)	1	200
4	SSD [Li16]	Inception v2	(300,300,3)	24	200
5	YOLO v2 [Sa18]	Darknet-19	(416,416,3)	64	250
6	YOLO v2 [Op18]	Darknet-19 PyTorch	(416,416,3)	64	80
7	YOLO v3 [RF18]	Darknet-53	(416,416,3)	64	250
8	YOLO v3 [Op18]	Darknet-53 PyTorch	(416,416,3)	64	80

Tab. 2: Configuration

Dataset	Training Images	Validation Images
DS1 and DS3	2983	200
DS2 and DS4	3647	300
DS1,2,3,4	6630	600

Tab. 3: Overview of Training and Validation

6 Results and discussion

Analysis of Mean Average Precision (mAP)
The goal of an object detector is to predict the location of a given class in an image or video with a high confidence value. This is done by placing bounding boxes to identify the position of objects. This is done by a set of three attributes: 1) Object class, 2) Corresponding bounding box and 3) Confidence score usually given by a value between 0 and 1. The evaluations are done based on :

- A set of ground-truth bounding boxes representing the rectangular areas of an image containing objects of the class to be detected
- A set of detections predicted by a model, each one consisting of a bounding box, a class, and a confidence value

Average precision (AP) is a popular metric in measuring the accuracy of object detectors. It computes the average value of the precision over the recall r interval from $r = 0$ to $r = 1$. The mean average precision (mAP) is the average of AP, which quantifies how good the model is at performing the query, results are shown in figure 4. In this work, true positives (TP), false positives (FP), false negatives (FN) and true negatives (TN) based on Intersection of union (IoU) are experimented and discussed in this section for analyzing the limitations from the top view perspective. In this research work, mAP is obtained for different combinations of dataset type as shown in section 3.

- True Positive (TP): A correct detection. Detection with IoU >= threshold
- False Positive (FP): A wrong detection. Detection with IoU < threshold
- False Negative (FN): A ground truth not detected
- The threshold for evaluation is set as 70 %

TP, FP and FN analysis for limitation in bicycle detection are shown in figures 2 and 3 from the results of the best model (8, see table 2) obtained for this work. The number of TP, FN and ground truth are discussed. The green bounding box represent ground truth, red bounding box represent the detected true positives and yellow bounding box represents false negatives. This analysis allows to determine the limitations such as view point variation, occlusion and detecting the exact number of bicycle present and for the evaluation of the mAP. When observed upto two bicycles there is no occlusion problem when multiple bicycles overlap (see figure 3). Whereas, for more than two bicycles, occlusions result in false negatives (see figure 2).

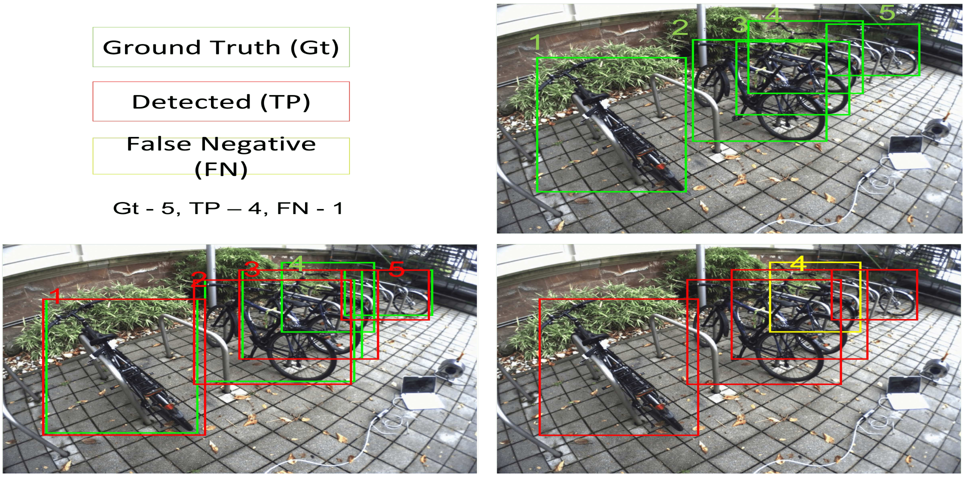


Fig. 2: TP, FP and FN analysis for limitation in bicycle detection. Here five ground truths present, but only four are detected. The ground truth which is not detected is a false negatives (FN) and it is shown in the last image in the second column with yellow bounding box.

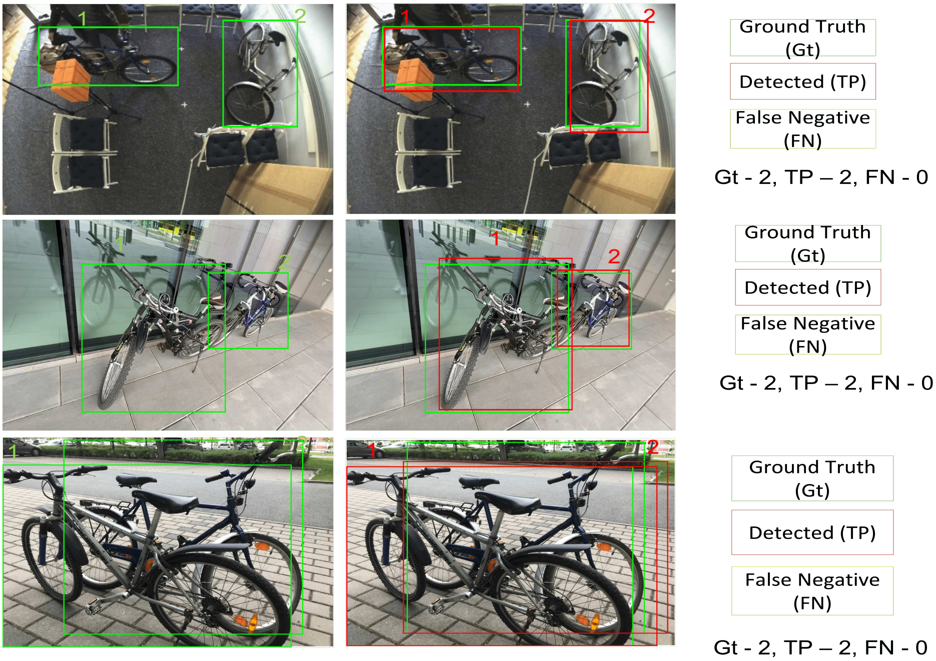


Fig. 3: TP, FP and FN analysis for limitation in bicycle detection.

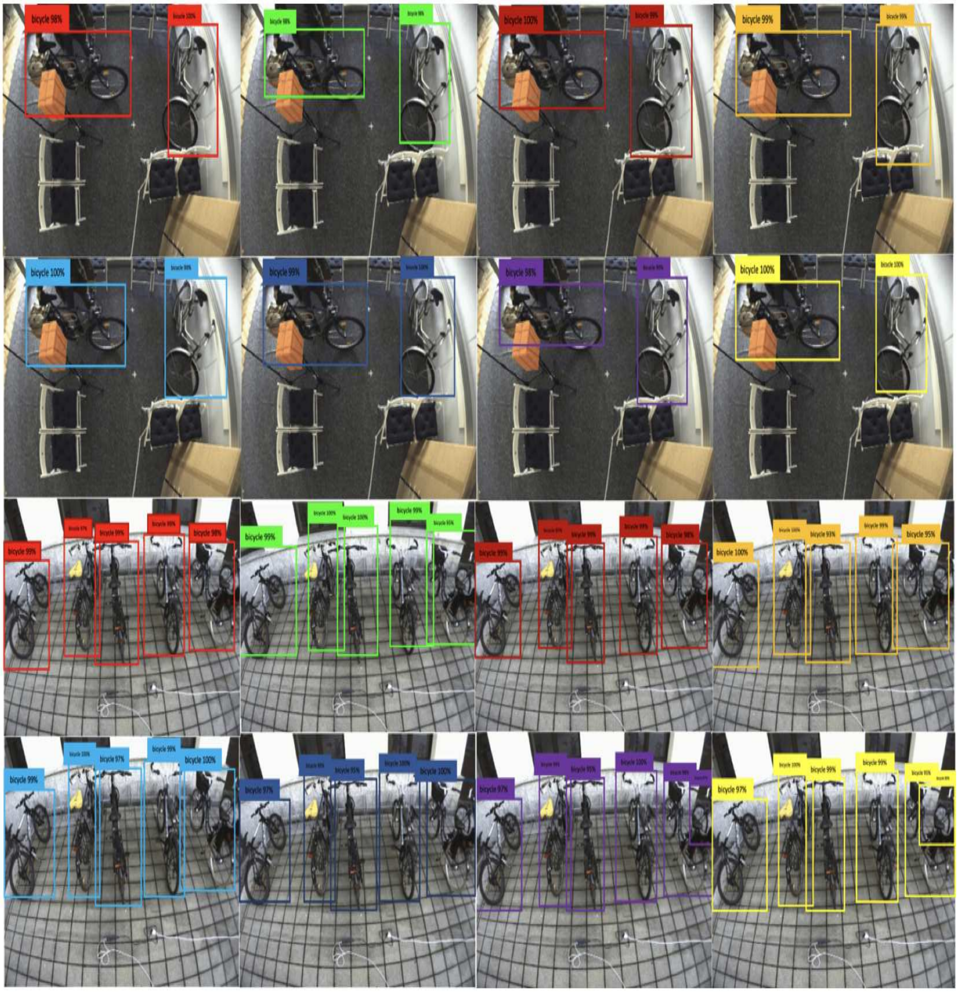


Fig. 4: Sample output from all eight models trained for combined dataset DS1, DS2, DS3 and DS4. The figure has 16 images. The eight models has unique colour for the predicted bounding box here. The first two rows are results from the Indoor dataset DS1 and DS3 and the last two rows are results from the outdoor dataset DS2 and DS4. The outputs of the models represented here are as follows with the first image in the first row as number 1 to the last image in the second row as 8 for the Indoor dataset, the same numbering is repeated for the last two rows of the image for Outdoor dataset. No.1 depicts FR-CNN Inception v2, No.2 depicts FR-CNN ResNet-50, No.3 FR-CNN ResNet-101, No.4 depicts SSD Inception v2, No.5 depicts YOLOv2 Darknet, No.6 depicts YOLOv3 Darknet, No7. depicts YOLOv2 PyTorch and No.8 depicts YOLOv3 PyTorch.

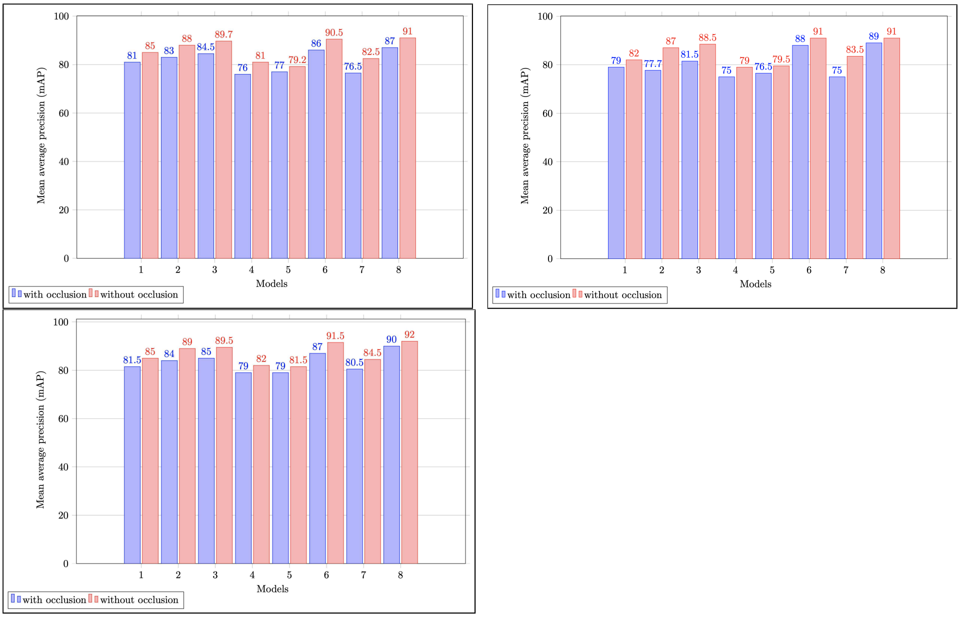


Fig. 5: The figure present here has three graphs which explain the Mean Average Precision obtained in y-axis and on x-axis all eight models shown in table 2. The first row shows results from dataset combination DS1 and DS3. The second row shows results from dataset combination DS2 and DS4. The third row shows results from combining all four dataset DS1,DS2, DS3 and DS4. The results are obtained based on with and without occlusions of bicycle.

From the fig 5, it is to be noted that, for this research work the number of bicycle present in a scene is limited to 10. Thus when experimenting with this dataset, good results were obtained. The dataset used in this research work was recorded from single view point but from different multiple camera deployed in a scene. Thus enabling to reduce problems occurring from view point variations. The mAP variations between with occlusion and without occlusion are less because the research work is focused more towards data driven approach, rather than model driven. This enables in achieving best results. It is clearly seen that model 8 out performs all other models.

7 Conclusion and Future Work

This work deals with bicycle detection from a top view perspective for surveillance application in bicycle parking area, inside public transportation's and near university area where there is more demand in everyday bicycle usage. Due to the scarcity of publicly available dataset, four customised datasets were prepared. They were combined into three categories and were used to benchmark the results on eight deep learning models. Evaluations are performed using mAP, TP, FP and FN analysis. The limitations were due to the single view perspective analysis of bicycle detection from the top view. One of the tedious work here was gathering the data and annotating multiple overlapped bicycles. In general overlapping of objects effect the performance of a deep learning model. Thus multiple overlapping of bicycle with bicycle in different view point may reduce the performance of detection. As the research work is well organized and data is gathered perfect, we have achieved better results.

During this research work, an occlusion problem occurred when one bicycle is completely hidden behind another bicycle from a top view perspective. This limitation can be encountered through quantifying the existing dataset by including synthetic datasets. Current work is under progress to overcome this limitation. This work focuses on creating a synthetic dataset using a game engine called Unreal Engine 4 (UE4), where the synthetic dataset represents scenes inside trains and trams used in Germany with humans and bicycle. This work focuses on bicycle detection from a multi-view perspective, where the real world and synthetic data fusion may provide better results. Furthermore, to reduce the effort and time involved in tedious process of manual labelling in supervised technique, the game engine generates ground truth for the synthetic dataset by incorporating required plugins and scripts.

Bibliography

[BM] Boris Sekachev, A. Z.; Manovich, N./, URL: https://github.com/openvinotoolkit/cvat.

[De21] Dequito, C.; Dichaves, I.; Juan, R.; Minaga, M.; Ilao, J.; Cordel, I. M.; Del Gallego, N.: Vision-based bicycle and motorcycle detection using a YOLO-based Network. In: Journal of Physics: Conference Series. Bd. 1922. 1, IOP Publishing, S. 012003, 2021.

[Fe10] Felzenszwalb, P.; Girshick, R.; Mcallester, D.; Ramanan, D.: Object Detection with Discriminatively Trained Part-Based Models. IEEE transactions on pattern analysis and machine intelligence 32/, S. 1627–45, Sep. 2010, URL: https://ieeexplore.ieee.org/document/5255236.

[GMC20] Garcia-Venegas, M.; Mercado-Ravell, D. A.; Carballo-Monsivais, C. A.: On the safety of vulnerable road users by cyclist orientation detection using Deep Learning. arXiv preprint arXiv:2004.11909/, 2020.

[HS06] Hinton, G.; Salakhutdinov, R.: Reducing the Dimensionality of Data with Neural Networks. Science (New York, N.Y.) 313/, S. 504–7, Aug. 2006, URL: https://www.researchgate.net/publication/6912170_Reducing_the_Dimensionality_of_Data_with_Neural_Networks.

[Le89] Lecun, Y.; Boser, B.; Denker, J.; Henderson, D.; Howard, R.; Hubbard, W.; Jackel, L.: Backpropagation Applied to Handwritten Zip Code Recognition. Neural Computation 1/, S. 541–551, Dez. 1989, URL: https://ieeexplore.ieee.org/document/6795724.

[Li16] Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.: SSD: Single Shot MultiBox Detector. In. Bd. 9905, S. 21–37, Okt. 2016, ISBN: 978-3-319-46447-3, URL: https://arxiv.org/abs/1512.02325.

[Li18] Liu, X.-Y.; Ding, Z.; Borst, S.; Walid, A.: Deep Reinforcement Learning for Intelligent Transportation Systems, Dez. 2018, URL: https://arxiv.org/abs/1812.00979.

[Op18] Ophoff, T.: Lightnet: Building Blocks to Recreate Darknet Networks in Pytorch./ , 2018.

[Re16] Ren, S.; He, K.; Girshick, R.; Sun, J.: Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. In. S. 1–10, Jan. 2016, URL: https://ieeexplore.ieee.org/document/7485869.

[RF18] Redmon, J.; Farhadi, A.: YOLOv3: An Incremental Improvement./, Apr. 2018.

[Sa18] Sang, J.; Wu, Z.; Guo, P.; Hu, H.; Xiang, H.; Zhang, Q.; Cai, B.: An improved YOLOv2 for vehicle detection. Sensors 18/, S. 4272, Dez. 2018, URL: https://www.researchgate.net/publication/329409946_An_improved_YOLOv2_for_vehicle_detection.

[Sh15] Shahraki, F. F.; Yazdanpanah, A. P.; Regentova, E. E.; Muthukumar, V.: Bicycle Detection Using HOG, HSC and MLBP. In (Bebis, G.; Boyle, R.; Parvin, B.; Koracin, D.; Pavlidis, I. T.; Feris, R. S.; McGraw, T.; Elendt, M.; Kopper, R.; Ragan, E. D.; Ye, Z.; Weber, G. H., Hrsg.): Advances in Visual Computing - 11th International Symposium, ISVC 2015, Las Vegas, NV, USA, December 14-16, 2015, Proceedings, Part II. Bd. 9475. Lecture Notes in Computer Science, Springer, S. 554–562, 2015, URL: https://doi.org/10.1007/978-3-319-27863-6%5C_51.

[VM19] Veres, M.; Moussa, M.: Deep Learning for Intelligent Transportation Systems: A Survey of Emerging Trends. IEEE Transactions on Intelligent Transportation Systems PP/, S. 1–17, Juli 2019, URL: https://ieeexplore.ieee.org/document/8771378.

[Zh18] Zhang, Z.; Wang, Y.; Chen, P.; Yu, G.: Intelligent Transportation Systems (ITS). In. S. 444–454, Jan. 2018, URL: https://www.researchgate.net/publication/326487592_Intelligent_Transportation_Systems_ITS.

[Zh19] Zhao, Z.-Q.; Zheng, P.; Xu, S.-T.; Wu, X.: Object Detection With Deep Learning: A Review. IEEE Transactions on Neural Networks and Learning Systems PP/, S. 1–21, Jan. 2019, URL: https://ieeexplore.ieee.org/document/8627998.

Multiple Sequence Alignment using Deep Reinforcement Learning

Roman Joeres[1]

Abstract: Multiple sequence alignment (MSA) is one of the primal problems in biology and bioinformatics. The question of how to align multiple sequences correctly is crucial for many other fields of research, e.g., gaining information about the evolutionary distance of two or more sequences and therefore about their corresponding species, finding protein targets for drugs, or finding a drug for a certain target protein.

Reinforcement learning (RL), and especially deep reinforcement learning (DRL), has become popular in recent years. To name just a few, DRL has shown major success in complex games such as Atari Games, Chess, and Go.

We model the problem of aligning multiple sequences as a Markov decision process (MDP) and examine the performance of different (D)RL algorithms compared to state-of-the-art tools.

Keywords: Bioinformatics; Multiple Sequence Alignment; Reinforcement Learning; Deep Reinforcement Learning

1 Introduction

MSA is one of the most important open fields in bioinformatics because it has applications in nearly every other field of bioinformatics research. Some of the most intuitive examples are phylogenetic trees that are used to describe the evolutionary distance between species. This can be measured relative to the number of similar or dissimilar patterns between sequences of the compared species [FD87]. Other applications are drug design and drug development where similar regions of sequences are matched and information on secondary structures and functionalities are inferred [Ko06; Li10].

Because this problem is NP-complete, the optimal alignment cannot be computed directly [WJ94]. Most classical approaches use a biological fine-tuned scoring function to score alignments computed using dynamic programming or fast Fourier transformation to reproduce the real evolutionary changes between sequences [Ed04; Ka02; THG94]. Aside from these, some new ideas use techniques from machine learning to find good or optimal alignments. RL provides techniques to solve multiplayer games or problems of sequential decision-making by learning and interacting in an MDP. The key idea in such settings is

[1] Saarland University, Saarland Informatics Campus (GradSchool), Campus E1 3, 66123 Saarbrücken, Germany
s8rojoer@stud.uni-saarland.de

to have an agent learning by trial-and-error in the MDP [SB18]. RL has been successfully applied to the traveling salesman problem [Ag19], solving the Rubik's Cube [KVW18], or optimizing a car manufacturing process [GGW20]. With methods from RL, DeepMind's AlphaFold won the CASP13 competition [Mo95], a comparison of different protein folding prediction programs that tries to find the correct fold of a protein structure [Se20].

In this paper, we give a comprehensive analysis of the performance of different RL algorithms in MSAs and offer a variety of interesting and promising ideas for future improvements. We use an MDP proposed by Mircea et al. [MBD14] who already received interesting results. At first, we will test multiple algorithms and optimization settings on DNA data from Mircea et al. Second, we test fewer algorithms on protein data from the BAliBASE dataset [TPP99].

2 Related Work

There are already approaches that apply RL to the problem of aligning multiple sequences. In 2014, Mircea et al. introduced RL to the field of MSAs [MBD14]. They focused on a tabular approach and proposed a first MDP to tackle the problem. Their work had promising results on some test data, and in a second paper with a slightly modified scoring function, the results were improved a bit [MBC16]. Jafari et al. introduced DQNs and an actor-critic algorithm in their work and could further improve the results of Mircea et al. in [JJR19]. But both papers suffer from the problem of a small selection of data to test their performance. In addition, both did not publish their code to reproduce the results and test it on larger benchmarks like BAliBASE [TPP99].

Other researchers also investigated the possibility to use RL techniques for MSAs using other models to represent sequence alignments. Edelkamp et al. [ET15] used a modified version of Monte Carlo tree search (MCTS) and achieved good results on some alignments from BAliBASE. Ramakrishnan applied DRL to the model of the Phylo game [Ka12]. But this model is too detailed to be applied to real-world problems with dozens of sequences and hundreds of molecules.

3 Multiple Sequence Alignment

The goal in MSA is to find mutations between sequences that model the evolutionary process the best. To test how well algorithms model this evolutionary process, benchmark datasets collect "true", hand-crafted solutions of alignments of multiple sequences. Before defining the different scorings in this field, we first define sequences and alignments formally.

Definition 3.1 (Sequence) *A sequence seq is a string over an alphabet Σ representing molecules M. For protein sequences, this alphabet contains 20 letters for the 20 amino acids (the molecules). For DNA sequences holds $\Sigma_{DNA} = \{A, C, G, T\}$, representing the four nucleic bases.*

Definition 3.2 (Alignment) *We denote an aligned sequence as* seq'*, a string over* $\Sigma' = \Sigma \cup \{-\}$*. The gaps,* $-$*, represent mutations such as insertions and deletions between sequences. Let SE be the set of all sequences of one alignment problem instance and SE′ be the set of all* seq'*. Then the alignment* $A \subsetneq SE'$ *contains possible aligned sequences and must fulfill three properties: (i)* $\forall\ seq_1, seq_2 \in A : |seq_1| = |seq_2|$*, (ii)* $\forall\ seq' \in A : seq' \setminus \{-\} = seq$*, and (iii)* $\nexists\ i \leq |A| : \forall\ seq' \in A : seq'_i = -$*.*

To align two or more sequences and to interpret the results of an alignment process, we have to score the resulting alignments to evaluate the quality of an alignment. Therefore, different scores have been developed and each of them has advantages and disadvantages as we discuss in the following.

Optimization Scores The sum-of-pairs-score (SP-score) is the most used score to align sequences. Used in a dynamic programming approach, one can compute the optimal alignment of k sequence of length at most n in $O(n^k)$ [NW70]. The SP-score for alignments $SP : SE' \rightarrow \mathbb{R}$ is defined using the function $sp : M \times M \rightarrow \mathbb{R}$ to score two molecules.

$$SP(A) = \sum_{i=1}^{N} \sum_{j=1}^{n-1} \sum_{k=i+1}^{n} sp\left(c_i^j, c_i^k\right) \quad \text{with} \quad sp\left(c_i^j, c_i^k\right) = \begin{cases} 0 & \text{if } c_i^j = c_i^k = - \\ 2 & \text{if } c_i^j = c_i^k \\ -1 & \text{if } c_i^j = - \text{ or } c_i^k = - \\ -2 & \text{if } c_i^j \neq c_i^k \end{cases}$$

measures the quality of aligned pairs of bases by assigning weights to types of pairs. We follow the definition of Mircea et al. [MBD14]. This kind of score is also used in state-of-the-art tools, namely CLUSTAL [THG94], MAFFT [Ka02], and MUSCLE [Ed04], which we later compare ourselfes to. Besides the SP-score, we use a relative score, called *column score* (C-score). This score $C : SE' \rightarrow [0, 1]$ is defined as

$$C(A) = \frac{\text{\# perfectly aligned columns in A}}{\text{\# columns in A}}.$$

Evaluation Scores Databases like the BAliBASE offer hand-crafted and hand-optimized solutions for alignment problems and from the supervised learning point of view, those solutions are the label for the problem instance. For the evaluation of an algorithm, one can compare the output of the algorithm with the labeled alignment $R \subsetneq SE'$ using the Q-score $Q : SE' \times SE' \rightarrow [0, 1]$ and TC-score $TC : SE' \times SE' \rightarrow [0, 1]$ [Ed04].

$$Q(A, R) = \frac{\text{\#correctly aligned pairs in A}}{\text{\#aligned pairs in } R}$$

$$TC(A, R) = \frac{\text{\#correctly aligned columns in A}}{\text{\#columns in } R}$$

3.1 Markov Decision Process

An MDP is defined as a tuple $(\mathcal{S}, \mathcal{A}, \mathcal{T}, \mathcal{R}, \mathcal{S}_0)$. The state-space $\mathcal{S}$ consists of all states and the action-space $\mathcal{A}$ represents all actions. The transition function $\mathcal{T} : \mathcal{S} \times \mathcal{A} \rightarrow \mathcal{S}$ describes to which states on can get. The reward function $\mathcal{R} : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \rightarrow \mathbb{R}$ maps states and actions to rewards. $\mathcal{S}_0$ is the initial state of the problem.

For us, $\mathcal{S}$ defines the possible orders of the sequences, including incomplete orders when not all sequences are aligned. These sequences are subsets of SE'. The applicable actions $\mathcal{A}$ remain in every state the same, i.e., every sequence can be aligned to the current alignment. The transition function $\mathcal{T}$ is, therefore, deterministic and maps a state and an action to the state representing the resulting order of sequences when aligning sequences (performing the actions) from the start state. Using this definition, a state can also be defined as the order of sequences (sequence of actions) that lead to a state. The rewards $\mathcal{R}$ are defined using a scoring function $score : \mathcal{S} \rightarrow \mathbb{R}$ used for optimization.

$$R(s, a_i, s') = \begin{cases} 0 & \text{if } n = 1 \\ -\infty & \text{if } a_i \in s \\ score(s') & \text{otherwise} \end{cases}$$

where $a_i \in s$ means that the i-th action was already performed earlier, i.e., the sequence i has already been aligned in s. The initial state $\mathcal{S}_0$ is the empty order, i.e., when no sequence is aligned. This model is visualized in Figure 1A.

4 Reinforcement Learning

RL is, besides supervised and unsupervised learning, the third big paradigm in machine learning. In contrast to supervised learning, RL does not depend on the existence of datasets. This learning technique is a trial-and-error approach that learns from the returns received from the environment, as depicted in Figure 1B. For each action the agent performs in the environment, it observes a reward that is used in reward functions to determine the quality of the selected action. The overall goal is to find a good policy, i.e., rules on how to behave in a state of the environment.

As in the papers of Mircea et al. [MBC16] and Jafari et al. [JJR19] we implement tabular agents (using the SARSA algorithm) as well as deep q-networks (DQNs) and apply both with three different reward structures, namely MC return, TD return, and λ-return. Jafari et al. used also an actor-critic approach with LSTM-cells. We will use the advantage actor-critic (A2C) consisting of an actor using policy gradients and a critic approximating the state-value function. To see the performance of policy gradients alone, we added the REINFORCE algorithm with and without a baseline to the portfolio of algorithms that are tested on MSAs.

We also implement an algorithm called "Upper confidence bounds applied to trees" (UCT) which is an improvement of normal MCTS. This algorithm also performs random rollouts

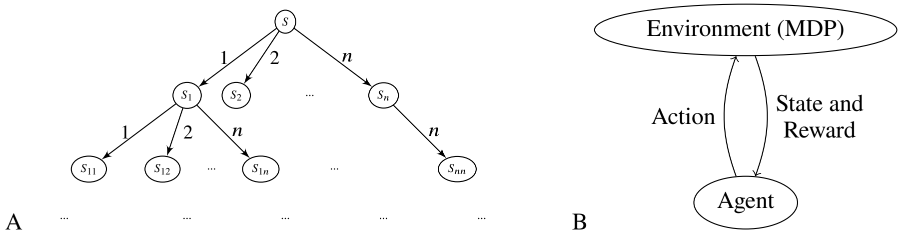


Fig. 1: A: Schematic overview of the model we use to describe multiple sequence alignments. B: Schematic view of a Markov decision process with the interactions between a learning agent and the environment.

in the state space to find a good action but improves its search using a tree policy to select a node in the known state space. The tree policy is given by

$$\max_a \left\{ v_a + \sqrt{\frac{\ln n}{n_a}} \right\} \tag{1}$$

where v_a is the average reward of action a, n is the total number of rollouts in the whole tree, and n_a the number of rollouts through this action. This formula is used to decide which child node to explore next by maximizing the score of the actions. If the child has not been explored yet, it expands the search tree in this node and performs a random rollout in one of the children [Co06].

5 Evaluation and Results

In the first step, we focus on the data that were already used by Mircea et al. [MBC16]. In Figure 2A, we can see exemplary results of the agents on the OxBench 429 data. This alignment contains 12 DNA sequences of length around 170 bases. For this task, our agents optimized the SP score and could outperform the classical reference tools as well as the results from Mircea et al. and Jafari et al.

One problem of SP-score optimization is the drop in performance in the alignments of the UCT-agent. The second part of the tree-policy (Equation 1) is most likely around 1, while the first part only depends on the reward (SP-score) of the alignments. Because the SP-score is unbounded, the first part dominates the formula. Here it can happen that the tree-policy always selects the same child because the SP-score of the first random rollout in the other children was too low. To solve this problem, we scaled the SP-score into the interval from 0 to 1. This is done using the center-star algorithm, which is proven to be a 2-approximation [Gu93]. The result is then used to estimate the upper bound, the lower bound is estimated as the highest score of 10 randomly arranged alignments. Using this trick, we could improve the SP-scores of UCT-alignments to a competitive level (see Table 1).

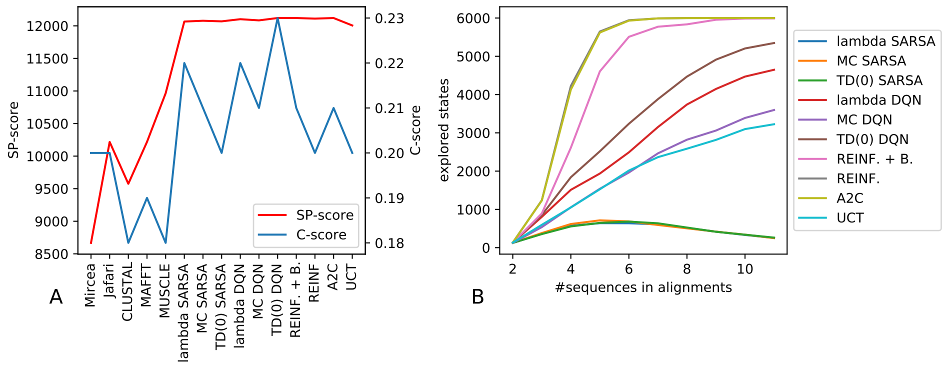


Fig. 2: Both plots show results from the OxBench 429 dataset. On the left, a comparison of the SP-scores and the C-scores of the 10 different agents and the reference algorithms. On the right, the number of explored states in the MDP is shown. The upper bound is given by the number of epochs (6000). The reference algorithms are not shown, because the numbers cannot be measured.

UCT-Mode	Hep.-C	P. Anubis	Ox469	Ox429	LGM	RLO	Ds. 1
Normal	18627	18848	639	11553	342	488	167
Adjusted	18627	18875	672	12006	348	488	167
Improvement	0.0%	0.1%	5.2%	4.0%	1.8%	0.0%	0.0%

Tab. 1: Improvement of adjusted UCT search with scaled rewards over normal UCT search with no scaling of the rewards.

	CLUSTAL	MAFFT	MUSCLE	SARSA	DQN	REINF.	A2C	UCT
Ox433	268	290	24	**353**	**353**	**353**	**353**	**353**
Ox641t2	659	723	442	**1053**	**1053**	**1053**	**1053**	**1053**
Ox34	355	377	-206	1215	**1223**	**1223**	**1223**	1208

Tab. 2: First protein analysis with data from the OxBench dataset.

In Figure 2B, we can see the number of unique explored states during the training. In contrast to tools like CLUSTAL, MAFFT, and MUSCLE, the presented agents perform many more alignments. From this, it follows that also the runtimes are much longer than for the classical tools. While CLUSTAL needs 0.06 seconds to find an alignment in OxBench 429, the SARSA agents need 20 minutes, and the policy agents up to 7 hours.

To extend this limited amount of data and to get an insight on how the tools might perform on protein sequence alignments, we choose three protein alignments from the OxBench dataset. OxBench focuses mainly on alignments between sequence families [Ra03]. The alignments consist of 3 and 6 sequences with lower sequence similarity but similar length compared to the DNA sequences. For those additional alignments, the results can be seen in table 2.

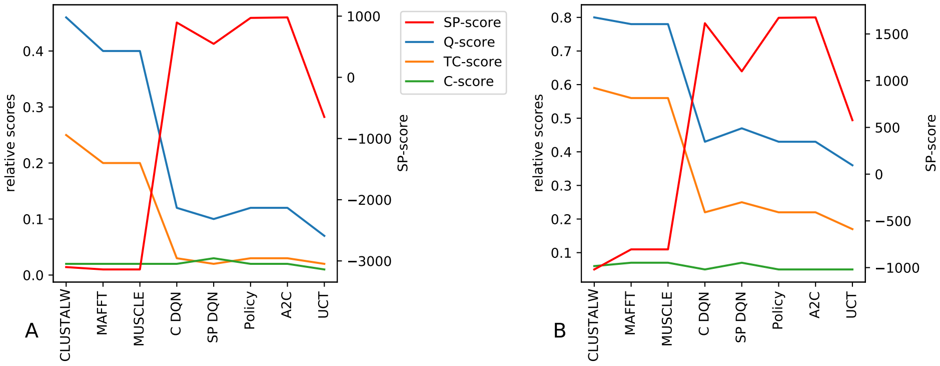


Fig. 3: Comparison of the performances of the algorithms with respect to different scoring functions. The values are averaged over the used alignments from the reference set 1.1 (left) and reference set 1.2 (right) of the BAliBASE dataset [TPP99].

In the second part of the analysis, we use alignments from the BAliBASE dataset [TPP99]. BAliBASE is made of reference subsets that contain sequences with similar properties. For each alignment in BAliBASE, there is a labeled alignment of the sequences. For our analysis, we use from the first reference set the alignments with fewer than or exactly ten sequences and an average length of the sequences of at most 500. These bounds are chosen because with longer sequences or more sequences the execution of RL algorithms takes too long. Those sequences can be split into two groups, namely the reference sets 1.1 and 1.2 containing alignments with lower and higher sequence similarities, respectively.

The comparison in this field is a bit different from the one before. Instead of analyzing an exemplary single sequence file, we average over the results within a reference subset. As the computation of these alignments takes much longer than aligning the few sequences above, we focus on DQN using the λ-return, the REINFORCE algorithm, the A2C algorithm, and the UCT agent. All of those agents will only optimize for the SP-score, except the DQN, it also optimizes for the C-score. We choose those agents because of their performance on the DNA data.

In Figure 3 the comparison of the average scores is denoted. Here, we additionally see the Q-scores and TC-scores of the alignments. The classical tools outperform the RL agents in terms of the Q-score and the TC-score due to the more fine-tuned scoring. On the other hand, the RL algorithms receive better SP-scores as they explicitly optimize for them. Again, this metric is evaluated using our simpler definition. We can also see that the RL tools narrow the gap to the state-of-the-art tools when the sequences to align are more similar as this is the case in the reference set 1.2 (Figure 3B).

6 Discussion

In Figure 2A, there is no big difference within the RL agents we implemented, because all of them work very similarly and the alignments are not complex. There are only a few DNA sequences, those are very similar and vary not much in length. The difference between our agents and the other RL agents is caused by the optimization score of the alignments. The agents of Mircea et al. and Jafari et al. optimized the C-score, not the SP-score, but the C-score is not a good measure for optimization. It automatically shrinks towards zero when there are more sequences or more different molecules in the type of sequence. The blue lines in Figure 2 show the optimization of the C-score and emphasize this problem. There are no big differences in the C-scores between the different agents and the scores are very low.

The results from the classical tools are optimized for their own SP-score, which is highly different from the simple one we use as described in section 3. In CLUSTALW, among others, this is done by sequence weighting based on the (dis-)similarity of a sequence to other sequences. The higher the dissimilarity or divergence of a sequence compared to others, the higher the sequence weight. These weights are used as multiplication factors for the sequences. There are several other improvements to this plain SP-scoring function presented above [THG94]. For the comparison, we computed our SP-scores of the alignments; therefore, those results are not very expressive. To get a more powerful result, one has to evaluate the alignments also using the SP-score function of the tools or know the labeled alignments as we do for the protein data.

When it comes to runtime, we see that the RL agents take much more time than the state-of-the-art tools caused by the increased number of alignments they have to compute. The reason for those additional alignments is rooted in the idea of RL and its trial-and-error character. To learn, an agent has to train for many episodes in its environment. Transferred to multiple sequence alignment, the agent has to perform many alignments to find out which are good ones and which are not that good.

The first results on the OxBench protein alignments are promising. Nevertheless, they are not very expressive, because the properties of the alignments are very similar to the ones from the dataset of Mircea et al. The pairwise sequence similarity is higher than in the BAliBASE dataset, the number of sequences is low, and the evaluation is done without labeled alignments. So, it is not very surprising that the RL tools perform better than the state-of-the-art tools. Furthermore, we can see that the adjusted UCT-algorithm still has some shortcomings. We see the scaling improves the performance (Table 1), but this does not always work (Table 2), as the estimates of the bounds might not be accurate enough.

The low performance has several reasons. If we inspect the alignments and in which order the sequences are aligned, we can get some additional insights in the alignment quality and how they were computed. The fine-tuned SP-score makes some implicit assumptions, such as one should align similar sequences first. From those assumptions, the alignments get a structure, i.e., the order of how the sequences are aligned. They produce complex

alignment-trees with the sequences in its leaves and the alignments of the corresponding leaves in the inner nodes. These structures are similar within the classical tools but cannot be found in the simpler, path-like alignments from the reinforcement agents. This might cause lower performance when it comes to the Q-score and TC-score.

An additional problem that is related to the way we defined the SP-score is the fragmentation of the sequences, i.e., there are many small gaps and many single nucleotides aligned in a longer segment of another sequence. This problem comes from the way we handle the gaps. In our definition, every gap gets the same penalty. Alternatively, extending an already existing gap is much more biologically meaningful and should therefore be punished less than opening a new gap as in convex gap costs [MBD14].

7 Conclusion

In this work, we applied multiple algorithms from the field of RL to the biological problem of multiple sequence alignments. We performed an extensive comparison with three state-of-the-art tools, namely CLUSTAL, MUSCLE, and MAFFT, and with other RL algorithms that were applied to multiple sequence alignments.

We tested ten different algorithms on several DNA sequence alignment problems and compared the results to other tools. We could easily reproduce results from other approaches that also used RL. Second, we applied the methods to alignments of protein sequences from the BAliBASE dataset.

We saw that RL performs well on sequences that have almost the same length. Unfortunately, for the DNA instances we used, no labeled solutions are available. So, we cannot compute Q-scores and TC-scores to see how well RL performs in terms of biological correctness. For sequences from the BAliBASE whose lengths are more different, the alignment quality sinks as multiple small gaps are interrupting the sequences in the final alignments. On most of the alignments, all state-of-the-art tools outperformed our agents.

But the major drawback of RL in multiple sequence alignments is the time that is needed to compute the alignments. Classical approaches solve the BAliBASE alignments within seconds or milliseconds; whereas the computations of our agents last for hours on the same sequence data. The reason for this is primarily based on the number of alignments a reinforcement agent computes, which is much higher than the number of alignments a classical alignment algorithm performs. For example, CLUSTAL performs mostly one progressive alignment that is then improved by a few iterative refinement steps, no matter how many sequences there are to align.

Although the results are not good, RL can help improve MSAs. Progressive alignments can be modeled as MDPs and RL can help distinguish good from bad alignment decisions. These new approaches should then be evaluated on expressive benchmarks such as BAliBASE with labeled alignments.

8 Future Work

To receive better Q-scores and TC-scores, one can use other variants of the SP-score for optimization. We use a SP-score with linear gap costs that lead to many single and small gaps in the alignment and multiple spread nucleotide islands. To prevent this, affine gap costs or the SP-Score implementations from state-of-the-art tools can be used.

One way to reduce the runtime and to increase the flexibility is to let the network generalize between alignments using additional data from the sequences and to modify the approach such that align-trees can be output. An idea could be to combine graphs and RL to directly compute align-trees. Therefore, each sequence is represented as a node in a fully connected graph. Then a graph neural network (GNN) can be used to choose an edge to contract by aligning the two nodes of the contracted edge. The resulting alignment structure can be more complex because alignment-trees can be produced. RL would be used to learn the optimal strategy that contracts edges.

9 Acknowledgement

This paper is based on my bachelor's thesis at the chair for modeling and simulation of Prof. Verena Wolf under the supervision of Timo P. Gros. I want to thank them for the opportunity to write my thesis on this topic and for their help during the development of the work.

Thanks also to Prof. Tobias Marschall at Heinrich-Heine University for being the co-supervising professor. His biological perspective and advice helped to interpret the results.

Bibliography

[Ag19] Agostinelli, F.; McAleer, S.; Shmakov, A.; Baldi, P.: Solving the Rubik's cube with deep reinforcement learning and search. Nature Machine Intelligence 1/8, S. 356–363, 2019.

[Co06] Coulom, R.: Efficient selectivity and backup operators in Monte-Carlo tree search. In: International conference on computers and games. Springer, S. 72–83, 2006.

[Ed04] Edgar, R. C.: MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic acids research 32/5, S. 1792–1797, 2004.

[ET15] Edelkamp, S.; Tang, Z.: Monte-carlo tree search for the multiple sequence alignment problem. In: Eighth Annual Symposium on Combinatorial Search. Citeseer, 2015.

[FD87] Feng, D.-F.; Doolittle, R. F.: Progressive sequence alignment as a prerequisitetto correct phylogenetic trees. Journal of molecular evolution 25/4, S. 351–360, 1987.

[GGW20] Gros, T. P.; Groß, J.; Wolf, V.: Real-Time Decision Making for a Car Manufacturing Process Using Deep Reinforcement Learning. In: 2020 Winter Simulation Conference (WSC). IEEE, S. 3032–3044, 2020.

[Gu93] Gusfield, D.: Efficient methods for multiple sequence alignment with guaranteed error bounds. Bulletin of mathematical biology 55/1, S. 141–154, 1993.

[JJR19] Jafari, R.; Javidi, M. M.; Rafsanjani, M. K.: Using deep reinforcement learning approach for solving the multiple sequence alignment problem. SN Applied Sciences 1/6, S. 592, 2019.

[Ka02] Katoh, K.; Misawa, K.; Kuma, K.-i.; Miyata, T.: MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic acids research 30/14, S. 3059–3066, 2002.

[Ka12] Kawrykow, A.; Roumanis, G.; Kam, A.; Kwak, D.; Leung, C.; Wu, C.; Zarour, E.; Sarmenta, L.; Blanchette, M.; Waldispühl, J. et al.: Phylo: a citizen science approach for improving multiple sequence alignment. PloS one 7/3, e31362, 2012.

[Ko06] Konagurthu, A. S.; Whisstock, J. C.; Stuckey, P. J.; Lesk, A. M.: MUSTANG: a multiple structural alignment algorithm. Proteins: Structure, Function, and Bioinformatics 64/3, S. 559–574, 2006.

[KVW18] Kool, W.; Van Hoof, H.; Welling, M.: Attention, learn to solve routing problems! arXiv preprint arXiv:1803.08475/, 2018.

[Li10] Li, Q.; Cheng, T.; Wang, Y.; Bryant, S. H.: PubChem as a public resource for drug discovery. Drug discovery today 15/23-24, S. 1052–1057, 2010.

[MBC16] Mircea, I.-G.; Bocicor, I.; Czibula, G.: A reinforcement learning based approach to multiple sequence alignment. In: International Workshop Soft Computing Applications. Springer, S. 54–70, 2016.

[MBD14] MIRCEA, I.-G.; BOCICOR, M.-I.; DÎINCU, A.: ON REINFORCEMENT LEARNING BASED MULTIPLE SEQUENCE ALIGNMENT. Studia Universitatis Babes-Bolyai, Informatica 59/2, 2014.

[Mo95] Moult, J.; Pedersen, J. T.; Judson, R.; Fidelis, K.: A large-scale experiment to assess protein structure prediction methods. Proteins: Structure, Function, and Bioinformatics 23/3, S. ii–iv, 1995.

[NW70] Needleman, S. B.; Wunsch, C. D.: A general method applicable to the search for similarities in the amino acid sequence of two proteins. Journal of molecular biology 48/3, S. 443–453, 1970.

[Ra03] Raghava, G.; Searle, S. M.; Audley, P. C.; Barber, J. D.; Barton, G. J.: OXBench: a benchmark for evaluation of protein multiple sequence alignment accuracy. BMC bioinformatics 4/1, S. 47, 2003.

[SB18] Sutton, R. S.; Barto, A. G.: Reinforcement Learning: An Introduction. The MIT Press, 2018.

[Se20] Senior, A. W.; Evans, R.; Jumper, J.; Kirkpatrick, J.; Sifre, L.; Green, T.; Qin, C.; Žídek, A.; Nelson, A. W.; Bridgland, A. et al.: Improved protein structure prediction using potentials from deep learning. Nature 577/7792, S. 706–710, 2020.

[THG94] Thompson, J. D.; Higgins, D. G.; Gibson, T. J.: CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic acids research 22/22, S. 4673–4680, 1994.

[TPP99] Thompson, J. D.; Plewniak, F.; Poch, O.: BAliBASE: a benchmark alignment database for the evaluation of multiple alignment programs. Bioinformatics (Oxford, England) 15/1, S. 87–88, 1999.

[WJ94] Wang, L.; Jiang, T.: On the complexity of multiple sequence alignment. Journal of computational biology 1/4, S. 337–348, 1994.

Appendix

Tool	Setting
CLUSTAL	Version 1.2.4 from December 16th, 2016 (Clustal Omega)
MUSCLE	Version 7.271 from July 3rd, 2020 (fast and progressive *FFT-NS-2* option [Ka02] p.4)
MAFFT	Version 3.8.31

Tab. 3: All of the tools are executed in default mode, so no search-influencing parameters were set.

General	
Network	Input x 256 x 128x x 64 x output
Activations	Mix of ReLU and Softmax
epochs	50*# seqs

SARSA and DQN		Policy Gradients	
α	0.1	α	0.01
γ	0.9	γ	0.99
λ	0.1	Baseline/critic	state-value network
ϵ-start	1	UCT	
ϵ-end	0	simulations	50*# seqs
ϵ time	10%	rollouts	1
TD-steps	TD(0)	C	1

Tab. 4: Hyperparameters used in the training of the agents.

Everything was executes on an Ubuntu Server, Version 18.04.4 LTS (Bionic Beaver) with 8 kernels and 4 GHz clocking.
The code for this paper and the underlying thesis is available on GitHub (https://github.com/Old-Shatterhand/MSADRL).

Anomaly Detection in Motion Timeseries using the Bosch XDK and Dynamic Time Warping

Julián Rico Mejía[1], Oscar Aguilar Aguila Isaías[2], Priyanka Paschapur[3]

Abstract: This paper presents the development of an anomaly detector for robotic movements using the dynamic time warping (DTW) algorithm and its implementation in Matlab. Data was collected by mounting the Bosch Cross-Domain Development Kit (XDK) sensor on a collaborative robot arm (Cobot), aiming at industrial applications in need for motion anomaly detection during repetitive tasks. The paper discusses practical issues like parameter tuning as well as algorithmic variants such as decoupling accelerometer and gyroscope data.

Keywords: Machine Learning, Dynamic Time Warping, Anomaly Detection, Bosch XDK, Collaborative Robot.

1 Introduction

Anomaly detection refers to the problem of finding patterns in data which do not conform to expected behaviour. It finds extensive use in a wide variety of applications such as fraud detection for credit cards, insurance or health care, intrusion detection for cyber-security, fault detection in safety-critical systems, and military surveillance for enemy activities [CBK09].

Anomalous data detection has been studied within statistics as early as the late 19th century. Nowadays, anomaly detection algorithms have been automated and systemized in many fields, such as computer science, especially in statistics applications, machine learning, data mining and information theory [BJ15].

The specific characteristics of an anomaly detection problem are determined by several factors. Important aspects of any anomaly detection technique are the nature of the input and output data, the type of anomaly, and the availability of labeled data. The input data is generally a collection of data instances and each data instance can be described by a set of attributes. The output data defines how the anomaly should be reported, by labels such as normal/anomalous or by scores, where each instance is assigned an anomaly score. The

[1] Hochschule Darmstadt, Fachbereich Elektrotechnik und Informationstechnik, Schöfferstraße 3, 64295, Darmstadt, julian.mejia@stud.h-da.de

[2] Hochschule Darmstadt, Fachbereich Elektrotechnik und Informationstechnik, oscar.isaias@stud.h-da.de. Stipendiary from CONACyT and DAAD

[3] Hochschule Darmstadt, Fachbereich Elektrotechnik und Informationstechnik, priyanka.paschapur@stud.h-da.de

nature of attributes can be of different types, such as binary, categorical, or continuous, and determines the applicability of different anomaly detection techniques [CBK09].

Movement recognition has been an active research area, with many practical applications in robotics, human-computer interaction, physical rehabilitation systems, and sign language. Nowadays, sensors with individual sensing capability are used in this field [CBK09]. Our research obtains the input data from the Bosch XDK sensor that allows for several sensing capabilities at once, especially those relevant for motion.

Different kinds of algorithms such as Support Vector Machines, replicator Neural Networks, and Density Estimation, have been used to implement Anomaly Detection [Ho14]. However, these methods compare similarly behaving data dimensions against each other, and classify sudden divergence as suspicious. In robot anomaly detection, many of the dimensions cannot be compared by common distance metrics as Euclidean distance. Dynamic Time Warping manages to compare timeseries that are not equal in size or that are similar but not fully synchronous by allowing one data point to be associated with more than one data point in the opposing timeseries.

The goal of this paper is to use DTW and machine learning on the data of robot movements and detect anomalies. Clustering of time series was used in the past by [ESL19] for data preprocessing in trajectory learning (motion) algorithms and [Ma19] who focused on anomaly detection of network traffic using DTW, applying this method on a labeled dataset for measuring the detection accuracy for different anomalies and providing the starting ground for our research. However it was further expanded to compare the analysis of dependent and independent DTW algorithms, as used in [Sh17], where the advantages and disavantages of dependent and independent DTW approaches were analyzed.

Henceforth, this work comprises a study analyzing the algorithm's performance applied to Anomaly Detection considering multiple factors such as different approaches to the DTW algorithm while using the Bosch XDK sensor as a data acquisition kit.

Section 2 of the paper presents the components of our system; section 3 describes the implementation of the algorithm; section 4 the performance criteria applied to our model; section 5 provides insight into the collaborative robot's movements, and section 6 details our analysis of the adaptability of DTW for this application.

2 Main components of the system

2.1 Bosch Cross-Domain Development Kit (XDK)

The Bosch XDK is, according to Bosch [Cr20], the "Swiss army knife" for IoT, and can be used for rapid prototyping in different areas of automation. It is an almost ready sensor which requires minimum programming experience to use [Bo20]. This paper utilizes two of its sensors for data collection: the accelerometer and the gyroscope, to analyze

movement. This data is comprised of accelerometer data in directions x, y, z and gyroscope's data in directions x, y, z and finally, their respective timestamp.

2.2 Dynamic Time Warping

Finding anomalies based on a collection of samples without hard criteria for the distinction between normal and anomalous movements requires a comparison rule. Similarities between these samples can be assessed in a consistent manner. As the recordings from an accelerometer and a gyroscope are presented to the identification algorithm as a vector containing the sampled data throughout a single execution of the movement to be identified, the first method that comes to mind for determining the similarity between sequences is using the Euclidean distance between them.

However, for the given application, the Euclidean distance will fail to account for differences between two timeseries, which may not have the exact same length (thus having different amount of points to compare in a pairwise manner) or may contain similar information occurring at different moments. The first issue will render the Euclidean distance mathematically impossible to implement without ignoring chunks of information and the latter will cause a high distance output for sequences that may be very similar but have delays with respect to one another. These two issues are represented in Fig. 1, which depicts two timeseries with different lengths but similar behavior. Data contained within the yellow circle (A) would have to be ignored to implement the algorithm and data in the green circles (B and C) is expected to be associated together, but it would not be the case [Co17].

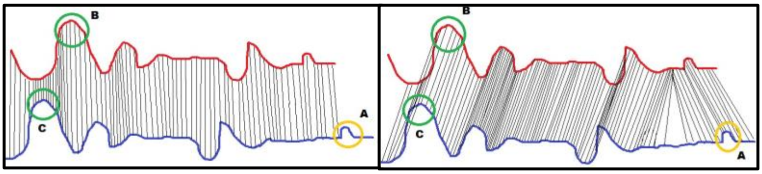


Fig. 1: Euclidean (left) vs DTW (right) Matching

The Remaining Useful Life (RUL) similarity models are used to predict when a given machine or equipment will fail to take preventive actions beforehand. Different approaches to this exist but the Pairwise Similarity Model computes the distance between different timeseries, which is fitting for our goal [RU20]. There are several approaches to achieving this but for the sake of this application the Dynamic Time Warping (DTW) was selected as its implementation is not challenging and it solves the comparison issues that arise from Euclidean distance measures.

DTW manages to compare timeseries that are not equal in size by allowing one data point to be associated with more than one data point in the opposing timeseries. Moreover, each measurement will be associated with a corresponding one from the opposing timeseries based on its Euclidean based similarity to measurements in the vicinity. Thus, two points

that are very similar but not taken at matching times will be linked together which accounts for differences in data capturing of the movements to be processed for anomaly detection.

As opposed with Euclidean distance matching, it can now be seen from the previous graph that all data points are used to determine the similarity of the blue curve with respect to the red one and that data within the green circles is linked together, as expected.

2.3 Classification of Distance Measures

Once a consistent method for assessing the similarity between two timeseries has been defined, the processing of this information must be established. As the system is intended to be implementable on virtually any movement (even non-robotic), a statistics-based approach seems viable.

After DTW is performed to the incoming measurements, information is reduced from a multidimensional representation into a one-dimensional distance measurement. Given that there are enough measurements, for a given application it can be assumed that the distance of each new measurement to a selected template will follow a normal distribution. From fitting the training data to a normal distribution, the first two parameters of the anomaly detector appear: the mean and the variance [Ma20].

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{1}$$

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \mu)^2 \tag{2}$$

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\cdot\left(\frac{x-\mu}{\sigma}\right)^2} \tag{3}$$

Equation 3 is commonly known as probability density function. This will provide the likelihood of the occurrence of a given input, in our case, a distance to the template. By establishing a threshold of likelihood for the distances, the third parameter for the anomaly detector arises.

The main idea is to test the likelihood of every distance obtained to a template and if this result fails to meet the threshold then the measurement will be considered as anomalous [Ma20]. It is worth noting that this must only occur for measurements whose distance lays in the right tail of the probability density function. If this distinction is not considered, then measurements that are very similar to the template would also be considered as anomalies because of the unlikelihood of this scenario, cf. Fig. 2.

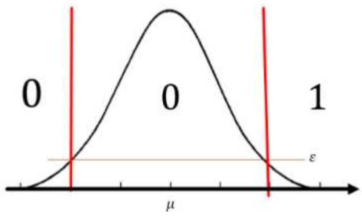


Fig. 2: Detection criteria along the probability density curve

The output of the anomaly detector algorithm is then represented in equation (4), where ε represents this so-called threshold.

$$y(x) = \begin{cases} 0 : \text{if } f(x) \geq \epsilon \\ 1 : \text{if } f(x) < \varepsilon \text{ and } f(x) > \mu \end{cases} \quad (4)$$

The mean and the variance are a result of fitting the training data to a Gaussian distribution. As no feedback is given to the algorithm at the end of this process this is an unsupervised methodology. A Gaussian distribution is assumed a-priori based on the assumption that there will be enough data for such a distribution to fit to the data and the resulting model is not compared to any other information.

However, this model will not be unique, especially if the number of samples used to fit the model decreases. Different selections of training data will result in different models or, in other words, different means and variances. This adds to the issue of setting a coherent value for the anomaly threshold, as the performance of the algorithm will completely depend on the three parameters (mean, variance, and threshold) to be optimally tuned for the specific application they are to be used in.

Although there is no simple way to predict the behavior of the data, there are means to assess whether a given sample is anomalous or not. This can be used to label the training data not only to assess the performance of the system, but also to feed this information back to the system so that the threshold can be set to optimize the performance of the anomaly detector given a probability density function for the data. As labeled data is fed back to the system so that its performance can be tuned (as opposed with the probability density function), the determination of the threshold is performed in a supervised manner.

The reason for choosing this hybrid approach instead of a purely supervised one is that the amount of positives (anomalies) is expected to be minimal with respect to the amount of negatives within all the data collected. The nature of anomalies is not easy to predict as anything that causes a significant deviation from the template will result in an anomaly. Because of this, a typical supervised approach, like a neural network, would face trouble identifying the "anomalous" category.

3 Matlab Implementation of the Algorithm

Data collected from the sensor is presented to the algorithm in Matlab as a collection of csv-files that contain the information from all 7 data sources (3 dimensions from the gyroscope, 3 from the accelerometer, and the timestamp). This information must be extracted into Matlab structures so that it can be used in the algorithm.

Once the data is available in a workspace structure, it must be divided into a training set and an evaluation set. A template is selected from the training set to be the standard against which all measurements are going to be compared. All the training data points are then compared to the template using DTW to obtain similarity values based on distance to the template. For this Matlab provides the command *dtw(X,Y)* which will output the distance between X and Y. The function will also be useful when classifying test data and new incoming data.

For fitting the model the Matlab command *fitdist(X,DISTNAME)* is leveraged. This function will return a variable containing information on the probability distribution of X when fitted to the type of distribution DISTNAME, including the mean and variances, which are needed to build the anomaly detector. A normalization of the data (subtracting the mean and division by the standard deviation of the dataset) supports fitting to a Gaussian model and makes data comparable when it comes from different sources and happens to be of different scale or range.

The search for the last parameter, the threshold, is done by searching for the value that maximizes performance of the model based on mean and variance obtained from *fdist*. To do this, the likelihood of a given value according to the obtained model must be calculated. This can be obtained directly from equation 3 or by using the expression *Y = normpdf(X, MU, SIGMA)*. This is the same expression that will allow for the implementation of the anomaly detector when evaluating test data or new incoming data after the '*dtw*' function has been used on it.

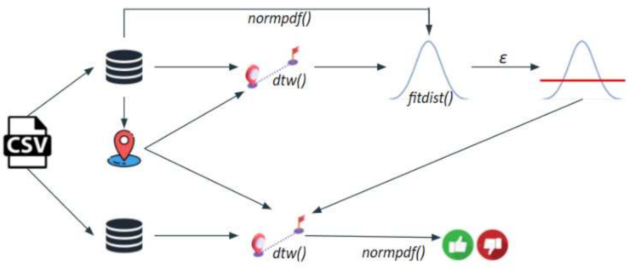


Fig. 3: Graphical overview of the algorithm and the main Matlab functions involved in it

4 Performance Criteria

Model evaluation is an integral part of the model development process. Evaluating model performance with the data used for training is not acceptable because it can easily generate overoptimistic and overfitted models. To avoid overfitting, a test set or collection of takes (not seen by the model) is used to evaluate model performance.

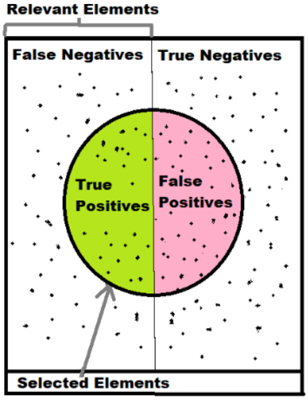


Fig. 4: Possible Outcomes of Anomaly Detection

In order to test the performance of the model, precision, recall and F1 were taken into account. Within the context of this paper, these concepts are defined as follows using the possible outcomes of anomaly detection seen in Fig. 4:
Precision refers to the number of correct classifications among all identified anomalies.

$$\text{Precision}=\frac{\text{True Positive}}{\text{True Positive}+\text{False Positive}} \tag{1}$$

Recall illustrates the percentage of all anomalies that the model found.

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive}+\text{False Negative}} * 100\% \tag{2}$$

F1 represents the harmonic mean of precision and recall and is typically used when both precision and recall matter, which is typically the case [Mo20].

$$\text{F1} = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}} \tag{3}$$

5 Movements of the Collaborative Robot (Cobot)

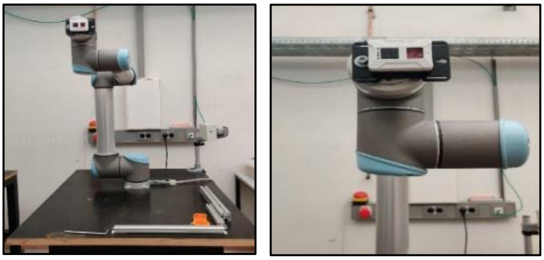


Fig. 5: UR Cobot (left) and XDK Sensor Placement (right)

5.1 Sensor placement, Robot Trajectory and Test Cases

The Bosch XDK Sensor was attached to the Cobot Tool Center Point (TCP) with the orientation as shown in Fig. 5. The robot followed a trajectory consisting of cartesian paths in 3 different planes, covering the workspace evenly to capture motion data from both gyroscope and accelerometer. 88 normal datasets and 16 anomalous datasets were obtained. Since the industrial robots are highly consistent in movement, in most cases, no anomalies would normally be available. Hence, to obtain anomalies some disturbances were introduced by human-robot interactions such as blocking the robot's path, bringing it to (safe) emergency stops, restarting the trajectory at random points by remote control, changing the sensor's orientation, pausing the robot's trajectory, and hitting (safely) the robot. This corresponds with industrial safety practices of Cobots which typically have human interactions.

5.2 Testing regime and observed results

Changed Parameters	Recall	Precision	F1
Training pool size	0.9091	1	0.9524
Gyroscope data only	0.7	1	0.8235
Accelerometer data only	0.6	1	0.75
Without Normalization	1	0.137	0.241
Using filter	0.6	1	0.75
Excluding timestamp	0.6	1	0.75
Increase anomalies ratio	0.6	1	0.75

Tab. 1: Overview of test results

The following test regime was carried out leading to the results in Tab. 1:

1. All data included: This means that all 88 datasets as well as 16 anomalies were evaluated under a single run of the algorithm. After this test, the datasets were

divided into a training pool (to train the algorithm) and a test pool (to test the trained algorithm).

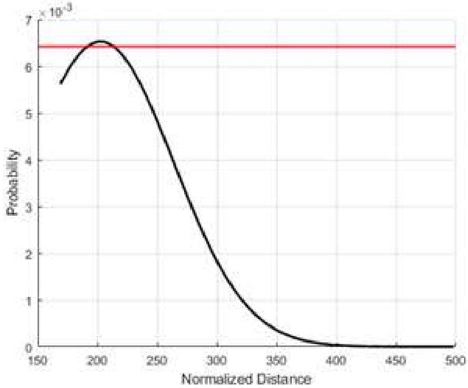


Fig. 6: Probability density graph based on normalized distances of experimental data

2. Testing only accelerometer data: This means that the first 3 dimensions, as well as their timestamp (dimension 7 of any dataset), were used.

3. Testing only gyroscope data: This means, the remaining dimensions, as well as the timestamp of the dataset were used. This improved model performance.

4. Changing training pool size: The effects of different training pool sizes were assessed by adding 6 datasets to the training pool in contrast to other parts of the regime. The best F1 is obtained by increasing the training pool size.

5. Excluding filter or normalization: During data pre-processing, a median filter was used to eliminate extreme outliers and data normalization was used to make data fit into the same normal range. The F1 did not change by turning off the filter and the worst F1 was obtained if the data is not pre-processed (normalized).

6. Increase anomalies ratio: Adding more anomalies in the training pool resulted in identifying more anomalies than before, as the algorithm becomes more robust to recognizing anomalies recorded during this research for the tests.

7. Shuffling Datasets: A pseudo-randomizer function which shuffles the data sets used for training and testing and is then fed to the algorithm. Shuffling the datasets leads to a better F1 score and is useful in identifying more number of anomalies. Among the 105 datasets obtained from Bosch XDK Sensor, training was run on 40 datasets (training dataset), and the model was tested against the other 65 datasets (called the validation dataset or testing set). The goal of cross-validation was to test the model's ability to predict anomaly in new data that was not used in estimating it, in order to flag problems like overfitting or selection bias.

The best and the worst F1 are 0.9524 and 0.241 respectively and this shows that the algorithm is heavily affected by the training pool size and data pre-processing. Accelerometer data seems to be noisier than the gyroscope's and hence provides poorer results by only using it as compared to only analysing gyroscope results. However, increasing the anomalies ratio within the training pool is effective in identifying all the anomalies, if those are labelled correctly as anomalies. Therefore, labelling is important, hence the mentioned combination of supervised and unsupervised processing is helpful.

6 Adaptability of DTW applied to the developed algorithm

6.1 Dependent and Independent DTW

As the data coming from the sensor is constituted of 7 different dimensions, it is of particular interest to look into how the DTW algorithm is taking this information into account and how it affects the performance of this application. The built-in function of Matlab takes a data series in, regardless of its number of dimensions, and proceeds to calculate the distance between both timeseries in what is described by [Sh17] as a "dependent" approach to the DTW. However, in this same work, it is described how not necessarily all dimensions of a given application are directly dependent with each other and how this affects the result of the anomaly detection.

The next step of this study was then directed into precisely this: implementing different methodologies for the DTW calculation and evaluating how the result is influenced by this. Results for a completely dependent approach were already obtained from the previous phase and two further variations were proposed:

- A completely independent calculation in which each dimension is compared only with the corresponding timeseries in the template and the individual results of all dimensions are then summed together to obtain a single value.
- A clustered independent calculation where data coming from the accelerometer is treated as a 3D timeseries and then compared in a dependent manner with the corresponding timeseries of the template. The result from this calculation is summed up with the result of the gyroscope´s data which is treated in the same way and with the result of the timestamp which is purely one-dimensional.

For this phase, the same implementation was used, except that the calculation of the distances through DTW was expanded to include the new approaches.

For the evaluation of all the approaches the size of the training and testing data sets was varied from 20 to 75 with step sizes of 5 to find an optimum size for each case. Tab 7.1 shows the performance of the algorithm while varying the training pool size for each of the proposed approaches of the DTW:

Pool Size	Dependent F1 (Test)	Dependent F1 (Training)	Clustered F1 (Test)	Clustered F1 (Training)	Independent F1 (Test)	Independent F1 (Training)
20	1	0.8	1	0.7368	1	0.7368
25	1	0.8	1	0.8571	1	0.8
30	1	0.8	1	0.8571	1	0.8571
35	1	0.9565	1	0.8571	1	0.9091
40	1	0.9565	1	0.8571	1	0.9565
45	1	0.9565	1	0.9565	1	0.9565
50	1	1	1	0.9565	1	0.9565
55	1	1	1	0.9565	1	0.9565
60	1	1	1	0.9565	1	1
65	1	1	1	0.9565	1	1
70	1	1	1	0.9565	1	1
75	1	1	1	0.9565	1	1

Tab. 2: Performance for different DTW approaches in training and test environments

The performance achieved in both training and in testing for all three approaches is outstanding. The difference lies in the training performance of the different approaches. We can notice that the dependent approach reaches an ideal performance with a training pool size of 50, while the independent does so with 60 and the clustered one never reaches the ideal training performance, although it still obtains an F1 of more than 95% with 45 samples. As the accelerometer and the gyroscope are attached together and their measurements are the result of the same movements it is not surprising that this application tends to behave better under a dependent approach. This would be the recommended implementation for a real-life scenario due to its simple implementation and desired results, while mentioning that any of the three approaches would yield an excellent performance.

Conclusions

Anomaly detection based on the automatic determination of the parameters of the probability density function (including a threshold) of the distance of a certain timeseries to a given "ideal" sequence proved to be a very suitable way of detecting deviances in the movement of a robotic arm. The achieved system performance exceeded our initial expectations; upon increasing the number of anomalies presented to the algorithm during training the final performance generally tends to increase.

DTW provided a versatile and handy concept that is useful to compare timeseries that may differ in length and may lack synchrony between them without losing information and retaining a real measure of similarity, without a high computational complexity. Different approaches for its implementation, such as clustering input dimensions, can provide advantages on particular applications. However, for our Cobot scenario, either of the

approaches showed a sufficiently good performance.

Acknowledgements

We would like to thank Prof. Dr.-Ing. Karl Kleinmann who provided guidance throughout this research and Mr. Timm Streul, Ms. Rashmi Rajanna who supported this work during its initial phases.

Bibliography

[BJ15] Blomquist, H.; Möller,J.: Anomaly Detection with Machine Learning. Quality Assurance of Statistical data in the Aid community, Uppsala Universiteit. 07/2015.

[Bo20] Bosch, YouTube Channel, https://www.youtube.com/watch?v=FAlyjBO0-7g&ab_channel = Bosch Connected Devices and Solutions, accessed: 05/2020.

[CBK09] Chandola, Varun; Banerjee, Arindam; Kumar, Vipin: 2009. Anomaly detection: A survey. ACM Comput. Surv. 41, 3, Article 15 (July 2009), 58 pages.

[Cr20] Cross Domain Development Kit XDK, https://www.Bosch-connectivity.Com/products/cross-domain/cross-domain-developement-kit/, accessed: 05/ 2020

[Co17] Costa B.G. et al. (2017) Fault Classification on Transmission Lines Using KNN-DTW. In: Gervasi O. et al. (eds) Computational Science and Its Applications – ICCSA 2017.-ICCSA 2017. Lecture Notes in Computer Science, vol 10404. pp.174-187 Springer, Cham.

[ESL19] Eiband, T.; Saveriano, M.; Lee, D.: Intuitive Programming of Conditional Tasks by Demonstration of Multiple Solutions, (IEEE): IEEE Robotics and Automation Letters, vol. 4, no. 4, pp. 4483-4490, Oct. 2019

[Ho14] Hornung, R. et al.: Model-free robot anomaly detection, (IEEE): 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 3676-3683, 2014

[Ma20] Machine Learning, https://www.coursera.org/learn/machine-learning, accessed: 03/2020.

[Ma19] Mahmoud, D. et al.: Anomaly Detection Using Dynamic Time Warping, (IEEE): 2019 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC), pp. 193-198, 2019

[Mo20] Model Evaluation, https://www.saedsayad.com/model_evaluation.htm, accessed: 20/06/2020.

[RU20] RUL Estimation Using RUL Estimator Models, https://de.mathworks.com/help/predmaint/ug/rul-estimation-using-rul-estimator-models.html, accessed: 08/2020.

[Sh17] Shokoohi-Yekta; et al.: Generalizing DTW to the multi-dimensional case requires an adaptive approach. Data Mining and Knowledge Discovery 31, pp. 1–31 (2017).

Informatik und Medizin

Identifying a Trial Population for Clinical Studies on Diabetes Drug Testing with Neural Networks

Tim Löhr[1]

Abstract: This project aims to model an end-to-end workflow of implementing different Artificial Intelligence (AI) tools for a clinical environment. A possible use case, such as the selection process of patients for a novel treatment, will be conducted by estimating the hospitalization time with a Neural Network on an Electronic Health Record (EHR) of diabetes. Then, Explainable AI (XAI) methods are computed for models trained with a Random Forest to evaluate the predictions. The diabetes readmission EHR dataset from the University of California, Irvine (UCI) Diabetes is used for this project. The trial population is selected by predicting the expected days for a person being hospitalized. An arbitrary boundary is set for choosing whether or not a patient shall be included into the trial. If so, a clear explanation of how the prediction is calculated and additional possible risk factors will be given in order to make the workflow explainable. This project shows that given a proper explanatory approach, AI can be a useful tool for the modern clinical environment. The workflow finally reveals that AI can be a beneficial support tool for doctors in the patient selection process.

Keywords: Machine Learning in the Industry 4.0; Clinical EDA; Data Analysis; AI in Medicine; Neural Networks; Diabetes

1 Introduction

Due to a new law in Germany, enforced on the first of January 2021 by the German Health Minister [Bu19], electronic patient files will be nationwide standardized in Germany. This opportunity can be seized to increase the impact of AI on the healthcare market to improve the health of Germany's population and facilitate doctor's work. Electronic Health Records offer a variety of different insights and application fields [SA19] [Am20], such as:

- Discover novel disease treatments.
- Improve patients diagnosis.
- Improve personalized healthcare.

In the year 2020, there are seven million cases of diabetes in Germany as stated by the Robert Koch Institut (RKI) [In20]. Even for well known diseases like diabetes, there is continuous research and novel drugs are invented on a frequent basis. Unfortunately, not every person is suitable for obtaining a treatment with a novel drug.

[1] Friedrich Alexander Universität Erlangen-Nürnberg, 91052 Erlangen, Deutschland timloehr@icloud.com

The project's background is an artificially created scenario. The assumption is, that the database collected through the new law offers an easily accessible API to gather data from. Electronic Health Records are under high data privacy conditions. They can only be used if the data is completely anonymized and patients can not be traced back. It is further assumed that those issues have already been taken care of and therefore all limitations and regularity steps have been successfully established.

After preparing the EHR for this privacy requirements, the data can be utilized to gain insights and evaluations out of it. Using AI applications in hospitals is a highly classified realization, because wrong decisions can cause a dangerous outcome towards patients. For this reason, implementations of *AI in Medicine* require a very detailed explanatory analysis of the predicted outcome. Keeping this background in mind, the project is an implementation of an end-to-end workflow of how AI is used in the real clinical environment.

The first step is to figure out which patients are most likely suitable for receiving novel diabetes treatments. This can be achieved by a patient selection process, as stated by Dr. Toddenroth et. al. from the FAU [FK13] and a more recent paper from Szu Yeu Hu et. al. [SYH19]. Then an uncertainty estimation of the prediction is computed with the Aequitas toolkit from the University of Chicago. This toolkit is used to explain the predictions with Explainable Artificial Intelligence (XAI) methods. This process is supposed to enable an understanding of how certain an estimation is with respect to different dataset imbalance problems, e.g. demography. The research question for this project can therefore be concluded as:

Can AI be safely applied in the real clinical environment if it provides enough explainability for its predictions?

2 Methods

The dataset is originated from the UCI Machine Learning Repository [Re08] collected from 1999 to 2008 with over 100000 entries to train on. Various features such as demographics, diabetes conditions or medications are provided. In total, there are 55 features from which 36 are included into the modeling, because the other 19 features were not useable due to missing and NaN values. An important detail is the distribution of ethnicity, gender and age. This disparity is used in section *Results* for evaluating the uncertainty of the model. Figure 1 and 2 show, that the median age of diabetes patients is 70 to 80 years. Figure 1 shows that the gender can be ignored, because it is almost equally distributed among all patients. The ethnicity disparity reveals the biggest concern, because among different ethnicities there is a huge distribution imbalance. Asian, African and Hispanic people are very underrepresented in comparison to Caucasian people. This plays an important role for the later upcoming predictive analysis and explanatory analysis. Keeping these distributions

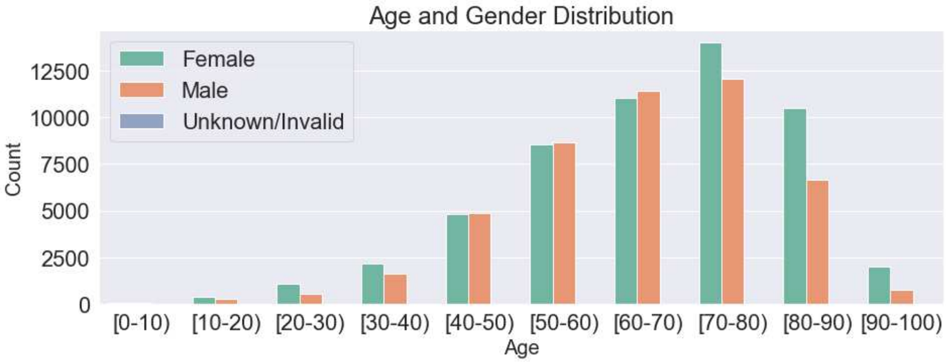


Fig. 1: The bivariate diagram shows the age disparity among all patients. The dataset is further split into the gender bins.

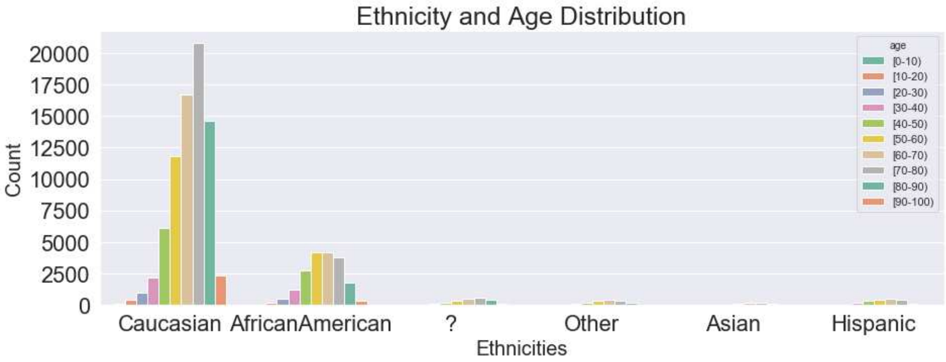


Fig. 2: The bivariate diagram shows the ethnicity disparity among all patients. The dataset is further split into the gender bins.

in mind, the methods used for this project will be conducted in the following respective order:

- Model a Neural Network and Random Forest.
- Explain the models with uncertainty estimation and metrics.
- Explain the predictions with SHAP and LIME.

SHapley Additive exPlanations (SHAP) is a python package for approximating the Shapely values. Those values can be used for various explanatory approaches, but for this project it is used for plots like in Figure 6. Local Interpretable Model-Agnostic Explanations (LIME) is a python package used for the LIME XAI approach.

Preprocessing is most of the time the first step before the training of an AI model. Various unnecessary features were dropped from the dataset and categorical features were transformed into a one-hot encoded representation. A supervised Neural Network always has a response variable, called a label. The *time in hospitalization* feature is used as the label, because it reveals the severity of each patient's diabetes disease. Usually, the longer people need to be hospitalized after receiving a treatment, the worse the condition of this person is. Predicting a low hospitalization time with a novel treatment can indicate compatibility with the patient and therefore he or she is a good match to be selected.

Neural Networks are able to achieve better results than standard Machine Learning approaches like the Random Forest. This can be proven due to the universal approximation theorem. Neural Networks are less explainable than most Machine Learning techniques, because they are mostly seen as a *blackbox*. Nevertheless, it was not feasible to compute the SHAP or LIME metrics for the Neural Network, that is why the Random Forest was used.

The Neural Network for this project is modeled with a Keras sequential function, consisting of Dense, Dropout, Dense Variational and a Distribution Lambda layers to estimate the hospitalization time with a certain probability. The Distribution Lambda layer from Tensorflow incorporates an ensemble of models which reports the variability of the prediction. It can be viewed as taking the mean of various regressions and measures then the qualitative difference between all outputs. Those layers are usually used in the clinical environment to reduce the risk of:

- Aleatoric Uncertainty: statisical uncertainty (*known unknowns*).
- Epistemic Uncertainty systematic uncertainty (*unknown unknowns*).

Aleatoric uncertainty will be evaluated with respect to the ethnicity distribution of the data. The neural networks prediction will be measured based on different evaluation metrics:

- Precision, Recall and F1 Score.

- Brier Score.
- AUC Score.

The loss function for the Neural Network is the negativ log likelihood function [DZ18]:

$$\text{Loss} = -log(y) \text{ for each prediction } y$$

The loss generally increases when the regression is unclear about its prediction and shrinks with increasing certainty. At this point, the model is able to predict how long a person is going to stay in the hospital with respect to the demographics and other features.

In conclusion to guarantee full transparency of the patient selection process, Shapely values and LIME try to explain which features led to which output. In other words, which features have a positive and a negative impact on the hospitalization time. The key difference between these two XAI methods is that Shapely values want to explain *how* a feature contributes to the overall prediction, while leaving out other combinations of features. LIME wants to explain what the most important features were in general for the prediction.

3 Results

After having the workflow in mind, the first step is to train the Neural Network and evaluate its results. The model trained for 50 epochs and reached a loss of 2.97. Figure 3 shows the course of events for different evaluation metrics. For example, the *Brier Score* itself is best when it is as low as possible. The other scores are best when they are as close to 100

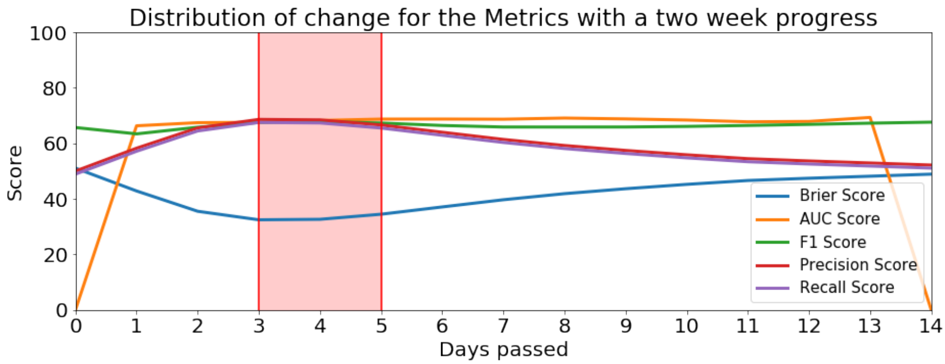


Fig. 3: The graphic shows the temporal change of the metrics with respect to a different time constraints on the boundary. The best values for the metrics can be found between the third and fifth day. Setting the boundary somewhere in the red highlighted area produces the best and most certain estimation.

percent as possible. The x-Axis represents the predicted time of hospitalizations in days. All metrics have an optimal score for the boundary at day four, as can be seen in Figure 3. After selecting the most promising discrimination boundary at day four, the uncertainty estimation with respect to ethnicity can be computed.

Figures 4 and 5 highlight that the true positive rate for Caucasian people is higher compared to the other ethnicities. The false-negative rate of other ethnicities is higher compared to the reference Caucasian people.

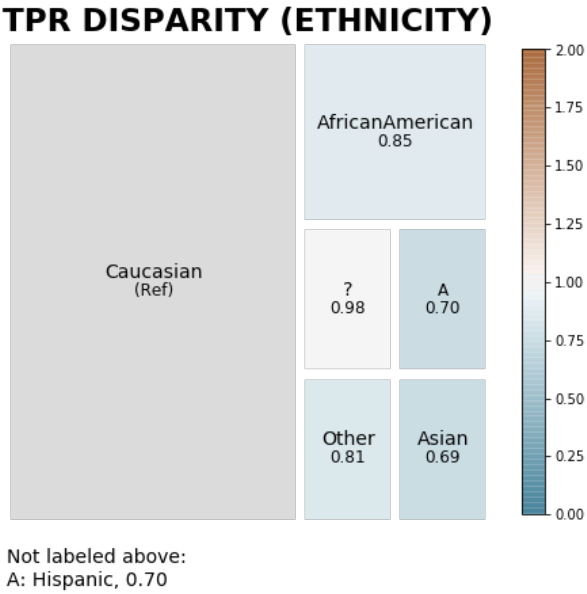


Fig. 4: True positive rate among ethnicities with Caucasians as reference group with the discrimination boundary at day four.

The next step is the investigation of how each prediction is computed and what the most important features for those predictions are. SHAP and LIME are used for this purpose, as mentioned in the Section Methods. Unfortunately, it was not possible to compute these two XAI methods with the Neural Network model, because the computational time due to many parameters was not feasible within the time constraint. To still generate valueable insights, a Random Forest was used to train on the same train- and testdata like the Neural Network. The Random Forest has fewer parameters and therefore the computation is much faster. The scores for the prediction of the hospitalization time were slightly worse.

The second interpretational approach are the Shapely values, which can be seen in Figure 6. This Figure lists 20 features with both their positive and negative impact on the prediction of the hospitalization time. Negative (blue) values on the x-Axis have an decreasing impact

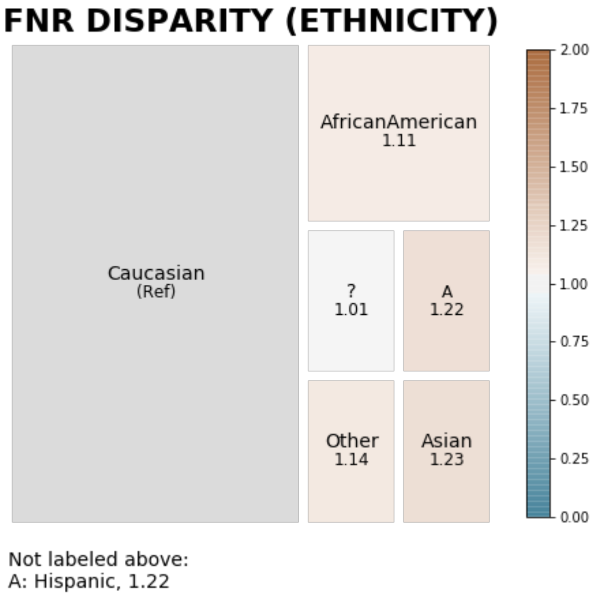


Fig. 5: False negative rate among ethnicities with Caucasians as reference group with the discrimination boundary at day four.

on the hospitalization time and positive (red) values on the x-Axis will increase the time of hospitalizations.

4 Discussion

Training the model alone is not sufficient enough for the selection of a patient into the trial population. There needs to be a specific boundary indicating after which time of hospitalization the diabetes is so serious, that this person can be taken into consideration for the trial. If this boundary is not set, the selection process can not be accomplished. Figure 3 not only shows the course of events, but it also highlights that different boundaries for the x-Axis produce various score curves. The optima is between the third and firth day.

The uncertainty estimation can be evaluated for different demographics and boundaries against each other. Figures 4 and 5 indicate, that if a patient is being selected and he or she is Caucasian, the probability of a positive or negative choice is more accurate compared to other ethnicities. Looking back on Figure 2, this important prediction disparity can be explained by the huge imbalance of ethnicities in the training data.

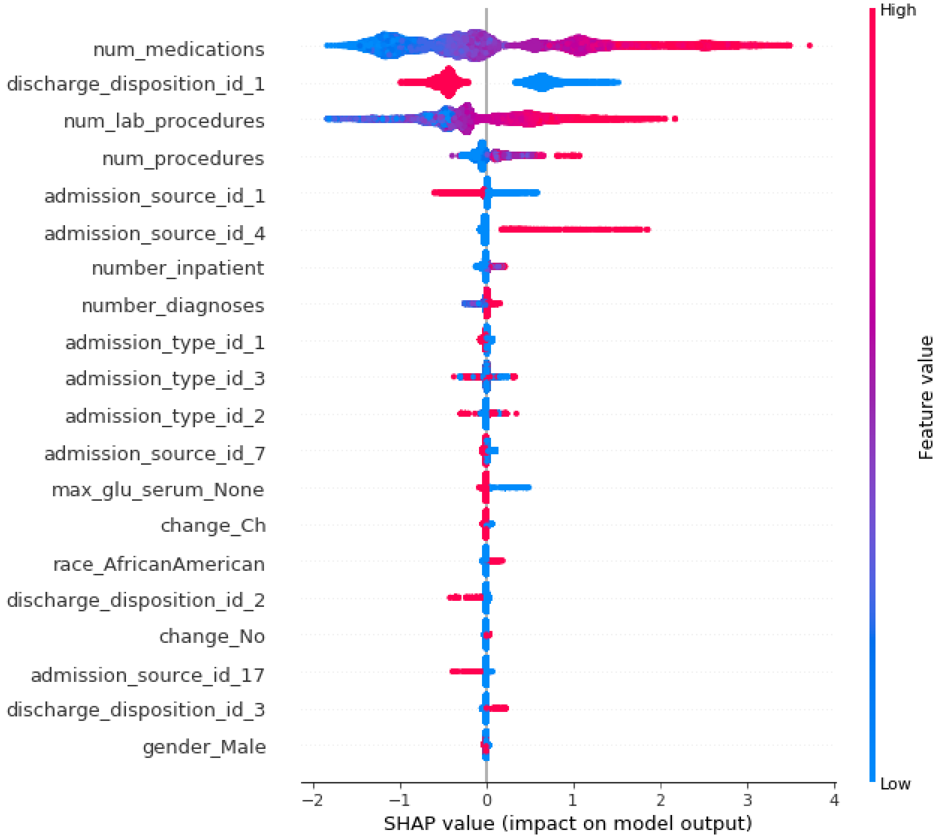


Fig. 6: Important features computed by SHAP.

Shapely values and LIME values need to be interpreted in a different way than the metrics. Figure 7 illustrates the single components which led to the prediction of 6.42 days. The meaning of these categorical features are provided by the following bullet points:

- discharge_disposition_id_1: Discharged to home.
- admission_source_id_5: Transfer from a Skilled Nursing Facility (SNF).
- admission_source_id_4: Transfer from a hospital.
- admission_source_id_10: Transfer from critial access hospital.

This patients EHR data has the features *admission_type_id_1* and discharge_disposition_id_6. This means that **not** having the features admission_source_id_5 and 6 actually has a **positive** (fewer days in hospital) impact on the hospitalization time. In contrary, when a patient has the feature *discharge_disposition_id_1*, he or she usually stays longer in the hospital. Also e.g. if the patient is taking the medication *Migitol*, he or she generally will stay up to one day (0.87) longer in hospital than patients who do not take this drug. Explanations like this can be conducted for all features to further investigate new insights into which effects medications and treatments can have on the patients health. This can be highly informative knowledge when making the choice to select patients into a clinical trial.

The distribution of *num_medications* can be seen in Figure 8. This Figure shows a right skewed normal distribution. Based on this distribution, it can be seen that a normally distributed numerical feature can have both a negative and a positive impact on the prediction of the most important features computed by SHAP, as shown in Figure 6. The more medications a person takes, the worse the condition and the longer the hospital stay for this person is, and vice versa. Whereas the feature *admission_source_id_4* always has a positive impact on the hospitalization time (higher hospitalization time).

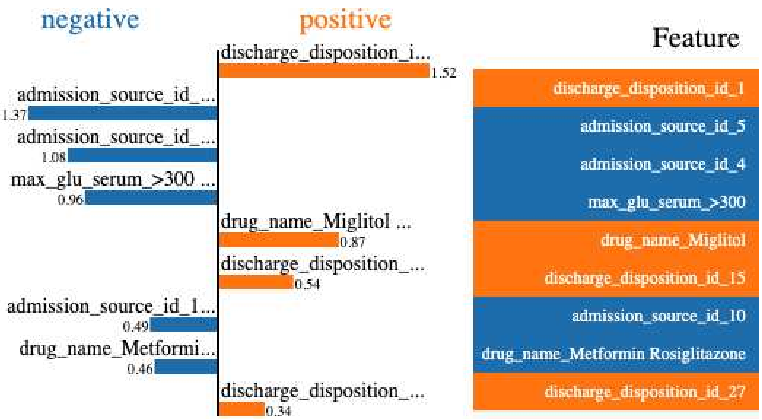


Fig. 7: Important features computed by LIME.

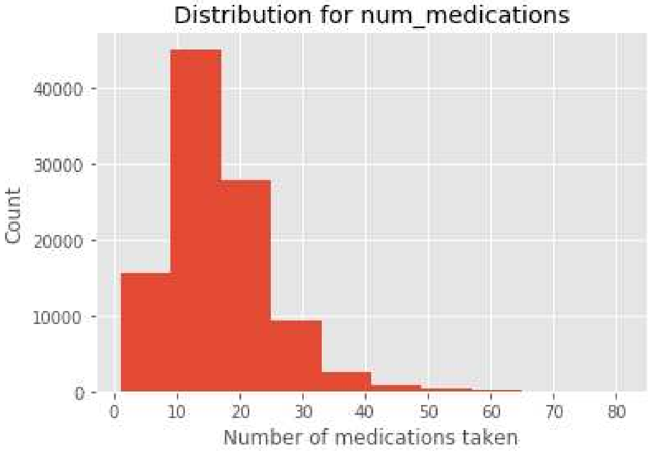


Fig. 8: Normal right skewed distribution of the feature *num_medications*.

After taking all evaluations and explanatory steps into consideration: Can AI safely be applied in the Industry 4.0 and especially hospitals? Keeping the results in mind, it appears that AI shows many valueable traits.

Pros:

- Support for doctor's decision.
- Patient preselection saves a lot of time.
- New insights into effects of medications or other features on patients.

Cons:

- Unstructured EHR data is very hard to handle.
- Not sure if every EHR dataset can be used to gain information out of it.
- The minimum necessary needed data amount for highly explainable predictions is unclear.

These techniques can be a good decision-support technology for doctors. If there are for example 2000 patients stored in the hospital's database, this preselection process can shrink the number of possible suitable patients already down in advance. Doctors will make the final decision based on their experience, but they can end up saving a lot of time by looking at already e.g. 20 preselected patients instead of investigating 2000 patient files. So even if the predicted population is not going to be selected, it has no dangerous impact for patients

in general. Doctors should be taught to develop more trust into a well trained and good performing AI systems, because they could benefit from it.

5 Conclusion

The conclusion of this entire workflow leads to three final questions to be answered.

Can it be applied without additional doctor's approval?

At this day and age not. What if only one person dies because of a false decision of the AI's prediction? This is a very sensitive topic, that's why the law and regulatory clinical instances like the FDA in the United States or the Arzneimittelbehörde in Germany must be involved when implementing such technologies in the clinical environment.

Can it be applied as a support technology?

If a doctor validates the results, it can be used to support the doctors decision for their clinical trials and save them up a lot of time.

Does a combination of metrics improve the overall explainability and how many metrics are enough for a sufficient explainable approach?

Using at first the F1-Score for evaluating the model's performance and then computing the uncertainty estimation with respect to critical features shows reasonable model decision making. Furthermore it can be seen, that many different approaches reveal different insights into the prediction. A good combination of metrics can enhance the overall performance, but there is no gold standard of metrics that explain 100 percent of the models behavior.

6 Future Work

For future research further improvements can also be made, such as:

- Try different XAI methods or explain more features in different ways.
- Try different Neural Network architectures with GPU training for higher accuracies.
- Model also the Epistemic Uncertainty with different Neural Network Layers: For example: Tensorflow offers the DenseVariational layer for a more robust Network prediction, as explained in Section Methods (epistemic uncertainty).
- Apply SHAP and LIME to a Neural Network.

Bibliography

[Am20] Amisha, Paras Malik, Monika Pathania Vyas Kumar Rathaur: Overview of artificial intelligence in medicine. Report, 2020.

[Bu19] Bundesgesundheitsministerium: Spahn: „Nur wenn wir die Chancen der Digitalisierung nutzen, koennen wir die Patientenversorgung besser machen. Technical report, 2019.

[DZ18] Donglai Zhu, Hengshuai Yao, Bei Jiang Peng Yu: Negative Log Likelihood Ratio Loss for Deep Neural Network Classification. Technical report, 2018.

[FK13] Felix Köpcke, Dorota Lubgan, Rainer Fietkau Axel Scholler Carla Nau Michael Stürzl Roland Croner Hans-Ulrich Prokosch Dennis Toddenroth: Evaluating predictive modeling algorithms to assess patient eligibility for clinical trials from routine data. Report, 2013.

[In20] Institute, Robert Koch: Diabetes Surveillance. Technical report, 2020.

[Re08] Repository, UCI Machine Learning: Diabetes 130 US hospitals for years 1999-2008. Technical report, 1999-2008.

[SA19] Shahabeddin Abhari, Sharareh R. Niakan Kalhori, Mehdi Ebrahimi Hajar Hasannejadasl Ali Garavand: Artificial Intelligence Applications in Type 2 Diabetes Mellitus Care: Focus on Machine Learning Methods. Report, 2019.

[SYH19] Szu Yeu Hu, Enrico Santus, Alexander W. Forsyth Devvrat Malhotraid Josh Haimson Neal A. Chatterjee-Daniel Kramer Regina Barzilay James Tulsky Charlotta Lindvall: Can machine learning improve patient selection for cardiac resynchronization therapy? Report, 2019.

Explainable Diagnosis of COVID-19 from Chest X-ray Images via CNNs

Emre Arkan[1], Jan Malte Beckert[2]

Abstract: This work demonstrates how Convolutional Neural Networks (CNNs) can be used to identify signs of *COVID-19* from Chest X-rays (CXRs) and discusses the challenges of deep learning with small datasets. In order to validate the model's performance, two novel explanation methods *LIME* and *Grad-CAM* are explored. Additionally, they serve to further increase users' confidence in specific classifications. Since the explanation results revealed model biases, additional preprocessing mechanisms were explored: A U-Net-based lung segmenter is introduced to the preprocessing pipeline, which masks all non-lung parts of the CXRs images. Subsequently, the segmentation and non-segmentation results were evaluated with regard to both their performance metrics and interpreted explanation results.

Keywords: COVID-19; Chest X-ray; CNN; Grad-CAM; LIME; Explainable AI

1 Introduction

When Chest X-ray (CXR) images are already been taken for the diagnosis of other diseases or injuries, an image-based classifier can be used with marginal costs to unveil previously undetected COVID-19 cases. Additionally, CXR images are non-invasive [1]. This work showcases a binary classifier that distinguishes between *COVID-19* and *NO FINDING*, i.e., healthy using a CNN (NASNetLarge) pre-trained on the ImageNet database [2].

Existing CXR image-based classifiers do not disclose any reasoning for why a particular decision was made. Ribeiro *et al.* [3] states that explainers help raise users' confidence both in the overall model and in a particular classification. This is especially important in medical settings where confidence requirements are high and insufficiently met by common metrics. Therefore, an explainer is used to validate the training results. Two approaches which facilitate explanation capabilities are explored: *LIME*, a black box explainer and *Grad-CAM*, a white box explainer utilizing the model's structure, respectively.

First, related work is discussed in Sect. 2. A description of the datasets that are used in this work is given in Sect. 3. Fundamental methods used in this work for data preprocessing, image segmentation, classification, and explainability are discussed in Sect. 4. The results are portrayed in Sect. 5 and are discussed in Sect. 6. Finally, the paper closes with a summary and an outlook on future work in Sect. 7.

[1] Universität Leipzig, Institut für Informatik, Germany, emre.arkan@studserv.uni-leipzig.de
[2] Universität Leipzig, Institut für Informatik, Germany, beckert@studserv.uni-leipzig.de

2 Related Work

Sethy *et al.* [4] demonstrate the detection of COVID-19 using CXR images with multiple CNNs as feature extractors and Support Vector Machines as classifiers. The authors have reached 95.38% of accuracy utilizing a pre-trained ResNet50.

As will be mentioned in Sect. 3, the selected datasets are of significantly different size. The challenges of deep learning in the context of COVID-19 with scarce data availability are discussed by Tartaglione *et al.* [5]. Using image segmentation to remove possible bias sources is proposed and the significance of adequate training data balancing is emphasized. Furthermore, the difficulties that emerge when training large models on scarce data such as overfitting are discussed: Apart from showing good metrics throughout validation, the models specialize in features attributed to the overrepresented class.

Ribeiro *et al.* [3] propose *LIME*, a universal explainer for classifiers. *LIME* trains a separate, interpretable model approximating a given prediction. The limitations of using pixel-regions to explain model predictions for image classifiers are emphasized: Supposing darker lungs on CXR images indicate COVID-19 infections, this behavior can not be explained using a heatmap of certain regions of the patient's lung. *Grad-CAM* was presented by Selvaraju *et al.* [6]. It is a white box explainer exploiting the activations of neurons which is reliant on the structure of the model. Since it is applied on CNNs exclusively, its applications are less universal compared to *LIME*, but promise more accurate explanation results.

3 Description of the Datasets

For the task at hand, CXR images for both healthy (*NO FINDING*) and *COVID-19* diagnosed patients are needed. CXR images can be distinguished by the position of the entry point of the roentgen rays: For CXR images in *posterior-anterior* (PA) view, the rays enter the thorax from behind and, vice versa, for *anterior-posterior* (AP) CXR images. *Lateral* CXR images enter the body from the side. PA and AP CXR images are the most prominent and can also be found predominantly in CXR image datasets. CXR in PA-view outperform AP-view images in areas such as clarity. Hence, they are used in favor of AP-view images although they commonly require the patient to be in better condition. [7]

Cohen *et al.* [8] provides a manually populated dataset containing CXR images of COVID-19 patients. The dataset initially contained 123 images annotated with sex, age, finding, survival, date, and location data. The dataset was updated and substantially expanded as described in [1]. The authors have collected 584 CXR images related to COVID-19 (196 of which are in PA-view) from all over the world on GitHub in a respective repository. The growth of the dataset, however, has stagnated around October 2020.

Wang *et al.* [9] assembled a dataset containing 112,120 CXR images from 30,805 patients with various diseases. 39,302 of the 67,310 PA CXR images show no findings. The images

are derived from hospitals' picture archiving and communication systems and extracted from radiological reports using Natural Language Processing techniques. The datasets that existed before the COVID-19 pandemic are significantly larger than the COVID-19 dataset. Hence, the training data is prone to be heavily unbalanced. However, a random subset of the dataset was created, which contains ca. 4000 of the 39,302 PA CXR images, as the whole dataset is not required for the trainings in the scope of this work.

4 Methods

This section presents the essential methods used in this work. The details of the preprocessing in conjunction with the U-Net, which is used to segment the CXR images prior to the classification, are delineated. Moreover, the CNN used for image classification is described together with the details of the training. Finally, both *LIME* and *Grad-CAM* are introduced, which provide a method for explaining predictions made by the model.

4.1 Preprocessing

Initially, common data preparation techniques are applied to the datasets, which are described in Sect. 4.1.1. These steps are incorporated into the preprocessing pipeline of all experiments. Sect. 4.1.2 presents a supplemental approach: A U-Net is used to mask the lung parts of the CXR images.

4.1.1 Data Preparation

First, the datasets, as described in Sect. 3, are loaded, divided into *Training*, *Validation*, and *Test* datasets. Next, the *ImageDataGenerator* [10] is used to standardize the images and to create additional variations of the images by rotating the images within a range of 10 degrees. Except for rotation, which is applied to *Training* data exclusively, the preprocessing steps are applied to all subsets. The model requires the images to have $n * n$ dimensions, however, most images in the datasets are of different aspect ratios. Thus, the images are padded by adding black bars along the longer sides of the image in conjunction with centering. Since the datasets, i.e., the classes are heavily unbalanced, a countermeasure must be taken. One technique is to use class weights derived from the relative underrepresentation of each class [11]. The class-weights serve as a multiplier that penalizes the network greater when it misclassifies an image from an underrepresented class. On the contrary, Japkowicz [12] proposes down-sizing of the over-represented class. Subsets of equal size are created by decreasing the largest dataset to the size of the smallest one.

4.1.2 Segmentation

The U-Net is a neural network based on the CNN architecture. In contrast to the other popular CNN architectures, it is an autoencoder extended by cross-connections with the aim of applying semantic segmentation. Ronneberger *et al.* [13] have used this architecture to prevent the loss of information. In addition, this architecture reduces the size of the input image several times and then enlarges the image again. The first step is mainly to filter the type of the object from the location of the object in an image. In the second step, incrementally downsized images from the first step are extrapolated to high-resolution images. Fig. 1 illustrates the structure of a U-Net.

In this paper an existing U-Net-based lung segmenter [14] is utilized to mask the lungs from the CXR images. The segmenter was trained on CXR images and masks from the Montgomery County CXR Set [15] and from the SCR database [16]. Ginneken *et al.* [16] provide 247 hand-generated masks for CXR images from the JSRT (Japanese Society of Thoracic Radiology) database [17], the Montgomery County CXR Set contains 138 masks. Both datasets contain CXR images in PA-view only.

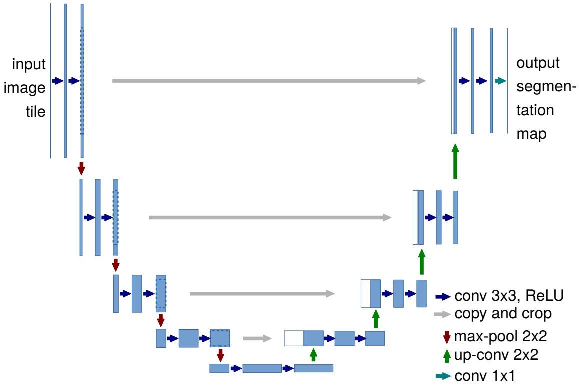


Fig. 1: Structure of the U-Net adapted from [13]

This U-Net consists of 4 max-pooling operations and 4 up-convolutions, respectively. Each max-pooling operation halves the resolution and doubles the number of filters for the subsequent convolutional layers. Both the input and the output have a size of $256 * 256 * 1$. Sect. 4.1.2 shows validation scores for *Lung Segmentation (2D)*.

The aim of the segmentation is to reduce complexity as the model should only consider the lungs for the classification.

	Montgomery County CXR	**JSRT/SCR**
IoU	0.956	0.971
Dice	0.972	0.985

Tab. 1: Intersection over union and dice scores for both Montgomery County CXR and JSRT/SCR datasets evaluated using 5-fold cross validation [14].

4.2 Training

The initial approach was to apply transfer learning on various models. The best-performing model was selected for further improvements. Considering that patients that require AP-view CXR images are typically in worse condition compared to PA-view CXR images (see Sect. 3), a bias source is presented: The model might learn to associate AP-images with *COVID-19* and PA-view with *NO FINDING*, respectively. Hence, only PA-view CXR images were selected for the training.

The CNN architecture used in this paper is based on the NASNetLarge, which Zoph *et al.* [18] have introduced. The pre-trained model on ImageNet [2] has a top-1 and top-5 accuracies of 0.825 and 0.960 [19], respectively. The feature extractor of NASNetLarge is used as a base for the model and a custom classifier is appended to be trained for a binary classification between *COVID-19* and *NO FINDING*.

The custom classifier begins with a global max-pooling layer, which is followed by a dropout layer disabling 30 % of the neurons randomly during the training. A dense layer containing 256 nodes succeeds the dropout layer. A regularizer function is applied to the output of the layer, i.e., an L2 regularizer. The last layer is preceded by a leaky ReLU and has a softmax activation. Due to the structure of the NASNetLarge, the input size of the images changes the architecture of the network [18]. Therefore, the input size of the images for transfer learning must be the same as the initial training. Thus, the default input size of $331 * 331 * 3$ was used, which leads to a total of $1,032,962$ parameters in the new classifier.

Each model was initially trained with frozen weights in the feature extractor. This first step was followed by the fine-tuning, where the weights of the feature extractor, i.e., NASNetLarge were also included in the training.

All trainings were monitored for early stopping in order to avoid overfitting. Both during the main training and the fine-tuning area under the precision-recall curve (AUC-PR) was observed and the training was stopped if this metric did not improve in the last 10 or 3 epochs (*patience*), respectively. These patience values were chosen experimentally according to preliminary training results. Afterwards, the weights of the best epoch were restored.

4.3 Explanation

In order to elevate users' confidence in the model, explainers are employed. Two approaches to explanation emerge: Explaining the general mechanisms of the model, which lead to classifications or explaining the classification of a specific image. Both techniques presented in this section generate explanations for a given image.

4.3.1 LIME

In contrast to other explainers, *LIME* does not analyze the structure and weights of a model to explain its predictions. Instead, *LIME* trains a separate model locally on a specific prediction. Hence, *LIME* remains indifferent to the nature of the model and can be used to explain predictions not only for neural networks but any classifier independent of its structure. Furthermore, *LIME* can be used more widely as both image and text classifiers can be analyzed. *LIME*'s approach is to create a linear approximation of the model, which can then be used to explain the prediction. First, the input state is fed into the model. Next, input nodes are modified by a small non-zero random value and the changes in the output class are noted. Such modifications are named *samples* that are used to construct the linear approximation. In the context of image classifiers, this means that some pixel values are modified. By tuning the pixel values, their effect on the output class can be observed, and therefore, it can be evaluated whether pixel-regions are increasing the model's confidence in a prediction, are mostly irrelevant to the prediction, or decreasing the model's confidence.[3] *LIME* produces different explanations depending on the number of samples selected. Various sample sizes were evaluated.

Fig. 2 illustrates the sampling mechanism: The blue/pink regions represent the black-box model where different colors indicate different classes and the bold red cross marks the prediction that is being explained. The remaining objects embody the modified input values/instances which are used to build the linear approximation of the model represented by the dotted line. The sizes of the remaining objects indicate their weights.

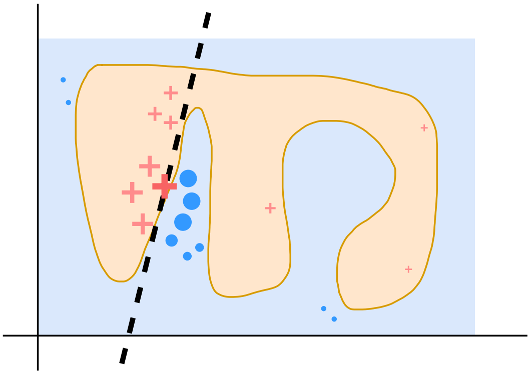


Fig. 2: Example illustrating the *LIME* sampling mechanism adapted from [3]

4.3.2 Grad-CAM

Gradient-weighted Class Activation Mapping (*Grad-CAM*) is a white box explainer that is applied to CNN models. It uses the activations of neurons to create an explanation and is, therefore, reliant on the structure of the model. *Grad-CAM* allows for the creation of visualizations that are class-discriminative, meaning they highlight regions that support the classification of the identified class only instead of highlighting all regions that had an impact on the decision. This is achieved by creating a weighted combination of feature maps for a given class. Layers closer to the top layer are given higher weights. In contrast to Class Activation Mapping (CAM) [20], *Grad-CAM* can explain the predictions of CNN-based models without modifying their structure and can be applied to models with fully connected

layers such as the model employed in this work. Since *Grad-CAM* is not training a separate model, it is orders of magnitude faster to compute explanations compared to *LIME*. [6]

5 Results

Sect. 5 shows the evaluation scores of chosen metrics for each of the trained models. The test dataset consists of a total of 40 CXR images when no class weighting is applied. *Original* means that neither the images for the training nor the images for the validation and test have been segmented. *Segmented*, on the other hand, represents the model that has only seen segmented images. Each model was evaluated twice, i.e., before and after fine-tuning.

Fig. 3 illustrates the training history for both models, where the accuracy and the AUC-PR are presented. Furthermore, a comparison of *LIME* and *Grad-CAM* for both models is visualized in Fig. 4. Even though the metrics for each model are comparable, the results with *Grad-CAM* showed poor performance initially. Therefore, another model with segmentation was trained, which showed similar performance but required more epochs to converge.

With respect to different sample sizes for *LIME*, it could be observed that the explanations converge to stable results for high numbers of samples. However, the computational cost increases respectively. Sample numbers of the order of 1000 have shown to provide stable explanation results.

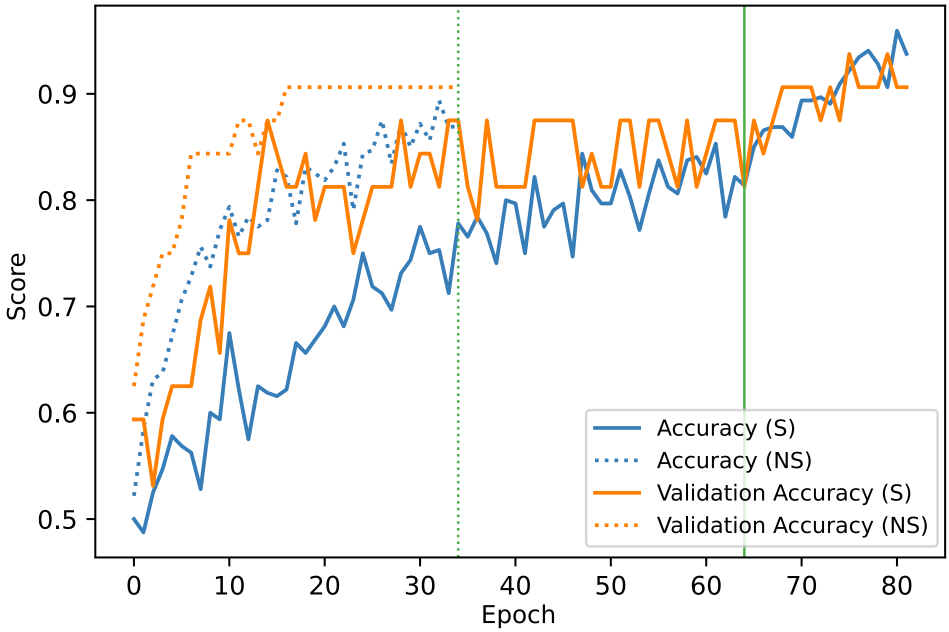
6 Discussion

Initial trainings showed that the class weight-based approach is clearly inferior to the approach where the sample sizes for each class are equal. As can be seen in Sect. 5, the class weight-based approach performed very poorly in recognizing COVID-19 cases (*Specificity*).

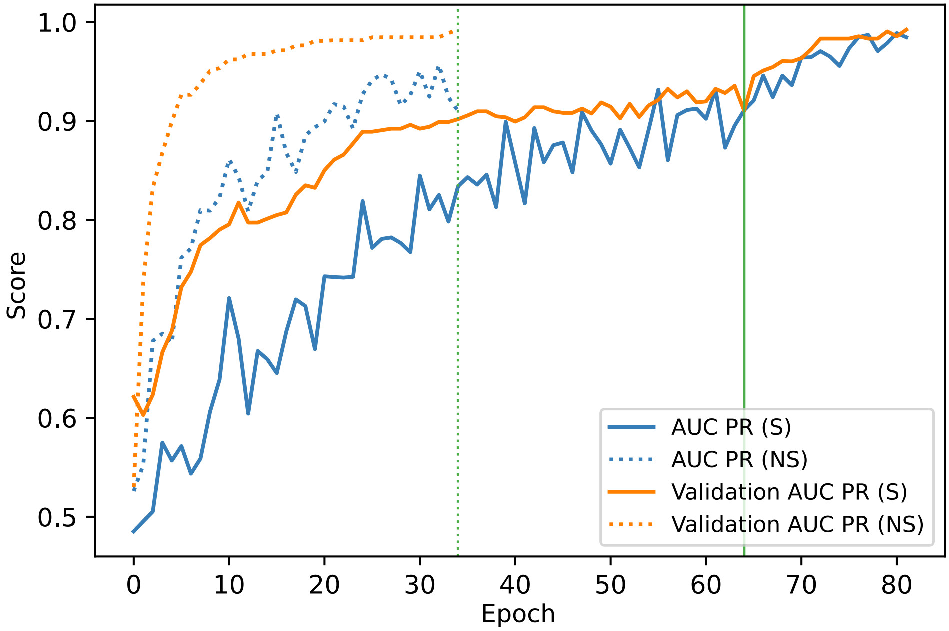
Model	**Metric**						
	Accuracy	Precision	Sensitivity	Specificity	F_1-Score	AUC-PR	AUC-ROC
Original	0.825	0.762	0.888	0.773	0.821	0.875	0.918
Original (FT⋆)	0.850	0.800	0.888	**0.818**	0.842	0.925	0.932
Original CW†	**0.883**	**0.873**	**1.000**	0.417	**0.932**	**0.957**	0.903
Segmented	0.825	0.762	0.888	0.773	0.821	0.878	0.896
Segmented (FT⋆)	0.850	0.773	0.944	0.773	0.850	0.942	**0.943**
Segmented CW†	0.850	0.842	**1.000**	0.250	0.914	0.941	0.860

⋆After fine-tuning. ; † Class Weight.

Tab. 2: Test results of trained models. Models trained with the class weight-based approach have the same values before and after fine-tuning.



a) Training and validation accuracy of both models.



b) Training and validation AUC-PR for both models.

Fig. 3: Green lines mark the beginning of the fine-tuning. Straight lines are for the model trained with segmented (S) CXR images. Dashed lines are for the model trained with CXR images that are not segmented (NS).

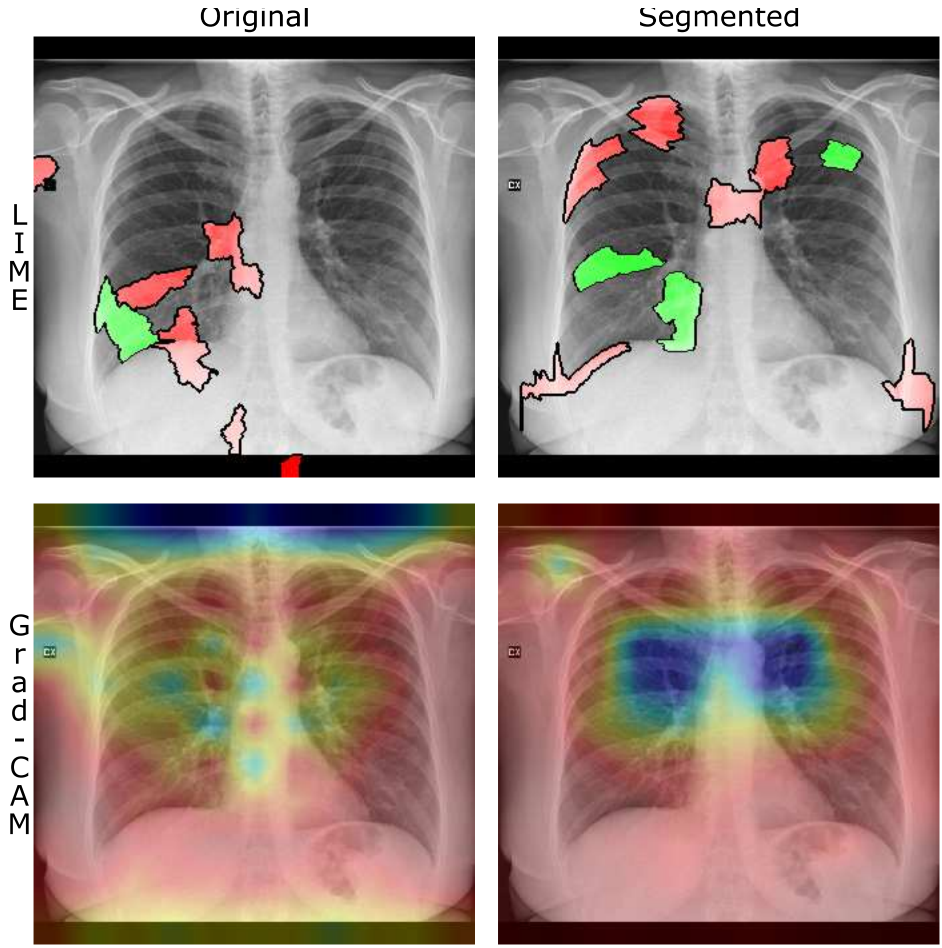


Fig. 4: *LIME* and *Grad-CAM* Explanations for the same image. This image was labeled with *COVID-19* and was also classified accordingly from both models. For *LIME*: Green markings indicate areas that support the decision, i.e., *COVID-19*. Red markings, on the other hand, are against the decision of the model. For *Grad-CAM*: The blue regions in the heatmap indicate more activation in the CNN.

After achieving 85 % accuracy with the NASNetLarge (see Sect. 5), interpretability of the decisions of the model was essential. Two techniques were tested with the model: *LIME* and *Grad-CAM* (see Sect. 4.3.1 and 4.3.2). As can be seen in Fig. 4, both techniques have different explanation results, though some parts overlap. However, *Grad-CAM*'s explanations showed that the model mainly focused on irrelevant areas of the images, e.g. markings or areas outside the lungs (see Fig. 4).

The model was trained with segmented CXR images and was capable of achieving very similar results (see Sect. 5). However, as can be seen in Fig. 3, the model requires almost

twice the number of iterations to converge with segmented images. This phenomenon could be due to the artifacts in the original images, such as markings indicating CXR view (see Fig. 4), which lead the model to detect patterns in the dataset that have no correlation with a COVID-19 diagnosis.

Furthermore, the experiment to compare class weights for unbalanced training with sample size balancing was repeated with segmented CXR images. The model performed even worse in recognizing COVID-19 labeled images (see *Specificity* in Sect. 5).

7 Conclusion and Future Work

The trained models showed good performance but left room for improvement. In regard to interpretability, the segmentation has improved the quality of explanations for both *LIME* and *Grad-CAM*. However, due to the construction of *LIME*, the changes made to the input images are random and tend to wonder outside the lung area. Furthermore, *Grad-CAM* is much faster than *LIME* as it basically extracts the information from a single pass of the input image through the model, which makes it more suitable for interactive medical applications.

Improvements to the classification quality can be achieved, e.g. by training a model from scratch with one or more large CXR datasets with multiple classes. Then the trained model can be used as a feature extractor to train a new model only to classify between *COVID-19* and *NO-FINDING*.

Moreover, the performance of the explainers should be validated by domain experts with regard to their capability to identify indicators of COVID-19 within the lungs. Cooperations in this regard are already in progress with a radiologist.

References

[1] J. P. Cohen, P. Morrison, L. Dao, K. Roth, T. Q. Duong, and M. Ghassemi, "Covid-19 image data collection: Prospective predictions are the future," *arXiv 2006.11988*, 2020. [Online]. Available: `https://github.com/ieee8023/covid-chestxray-dataset`.

[2] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *2009 {IEEE} Conference on Computer Vision and Pattern Recognition*, IEEE, 2009. DOI: `10.1109/cvpr.2009.5206848`.

[3] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should i trust you?: Explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA: Association for Computing Machinery, 2016, pp. 1135–1144. DOI: `10.1145/2939672.2939778`.

[4] P. K. Sethy, S. K. Behera, P. K. Ratha, and P. Biswas, "Detection of coronavirus disease (COVID-19) based on deep features and support vector machine," *International Journal of Mathematical, Engineering and Management Sciences*, vol. 5, no. 4, pp. 643–651, Aug. 2020. DOI: 10.33889/ijmems.2020.5.4.052.

[5] E. Tartaglione, C. A. Barbano, C. Berzovini, M. Calandri, and M. Grangetto, "Unveiling COVID-19 from Chest X-ray with deep learning: a hurdles race with small data," 2020. arXiv: 2004.05405.

[6] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," *International Journal of Computer Vision*, vol. 128, no. 2, pp. 336–359, 2020. DOI: 10.1007/s11263-019-01228-7. arXiv: 1610.02391.

[7] A. Er and A. Murphy, *Chest (PA view)*. [Online]. Available: https://radiopaedia.org/articles/chest-pa-view-1 (visited on 05/02/2021).

[8] J. P. Cohen, P. Morrison, and L. Dao, "COVID-19 Image Data Collection," 2020. arXiv: 2003.11597. [Online]. Available: http://arxiv.org/abs/2003.11597.

[9] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "ChestX-ray: Hospital-Scale Chest X-ray Database and Benchmarks on Weakly Supervised Classification and Localization of Common Thorax Diseases," *Advances in Computer Vision and Pattern Recognition*, pp. 369–392, 2019. DOI: 10.1007/978-3-030-13969-8_18.

[10] *ImageDataGenerator (TensorFlow Core v2.4.1)*. [Online]. Available: https://www.tensorflow.org/api%7B%5C_%7Ddocs/python/tf/keras/preprocessing/image/ImageDataGenerator (visited on 05/06/2021).

[11] X. Guo, Y. Yin, C. Dong, G. Yang, and G. Zhou, "On the class imbalance problem," *Proceedings of the 4th International Conference on Natural Computation, ICNC 2008*, vol. 4, pp. 192–201, 2008. DOI: 10.1109/ICNC.2008.871.

[12] N. Japkowicz, "The Class Imbalance Problem: Significance and Strategies," *Proceedings of the 2000 International Conference on Artificial Intelligence*, pp. 111–117, 2000. DOI: 10.1.1.35.1693.

[13] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9351, pp. 234–241, 2015. DOI: 10.1007/978-3-319-24574-4_28. arXiv: 1505.04597.

[14] *Lung Segmentation (2D)*. [Online]. Available: https://github.com/imlab-uiip/lung-segmentation-2d (visited on 04/29/2021).

[15] S. Jaeger, S. Candemir, S. Antani, Y.-X. J. Wáng, P.-X. Lu, and G. Thoma, "Two public chest X-ray datasets for computer-aided screening of pulmonary diseases.," *Quantitative imaging in medicine and surgery*, vol. 4, no. 6, pp. 475–477, 2014. DOI: 10.3978/j.issn.2223-4292.2014.11.20.

[16] B. van Ginneken, M. B. Stegmann, and M. Loog, “Segmentation of anatomical structures in chest radiographs using supervised methods: A comparative study on a public database,” *Medical Image Analysis*, vol. 10, no. 1, pp. 19–40, 2006. DOI: 10.1016/j.media.2005.02.002.

[17] J. Shiraishi, S. Katsuragawa, J. Ikezoe, T. Matsumoto, T. Kobayashi, K.-i. Komatsu, M. Matsui, H. Fujita, Y. Kodera, and K. Doi, “Development of a digital image database for chest radiographs with and without a lung nodule,” *American Journal of Roentgenology*, vol. 174, no. 1, pp. 71–74, Jan. 2000. DOI: 10.2214/ajr.174.1.1740071.

[18] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, “Learning Transferable Architectures for Scalable Image Recognition,” *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 8697–8710, 2018. DOI: 10.1109/CVPR.2018.00907. arXiv: 1707.07012.

[19] *Keras Applications*. [Online]. Available: https://keras.io/api/applications/ (visited on 05/02/2021).

[20] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, “Learning deep features for discriminative localization,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 2921–2929. DOI: 10.1109/CVPR.2016.319.

Building a GAN for Replicating Epithelial Impedance Spectra for ML-based Pattern Recognition

Lena Jurkschat, Benjamin Schindler

Abstract: Impedance spectroscopy is a common method in the field of biotechnology to measure electrical conductivity of special cell lines (i.e. ephitelial). Based on the measured impedance spectra, machine learning (ML) techniques including random forests and feedforward networks are increasingly used to determine physiological properties of the underlying cell tissue and to detect a wide range of diseases. However, training ML models for this purpose typically requires large amounts of data and real cell tissue measurements are costly to obtain due to their experimental setup. This paper introduces a Generative Adversarial Network (GAN) which meets the high demand for training data by replicating impedance spectra from a given data set. As a proof of concept, we show that GANs are capable of generating spectra that have a similar shape to the original ones and could therefore be used to overcome a lack of training data.

Keywords: GAN, impedance spectroscopy, neural networks, epithelia

1 Introduction

Artificial Neural Networks (ANNs) cover a variety of tasks, such as classification, regression and translation of texts. A special kind of ANNs are so called *Generative Adversarial Networks* (*GAN*), which are composed of a *Generator* and a *Discriminator* subnetwork. As generative models, GANs learn distributions within a data set and, if trained successfully, are able to generate new samples from them. GANs are very well known for their capacity of generating fake pictures of persons [Goo+14]. Trained on large sets of high-dimensional data (e.g. pixels of an input image), a GAN is able to generate fake pictures that a human cannot distinguish from real ones.

In the field of biotechnology, impedance spectroscopy is widely used in the investigation of epithelial tissues to determine their electrical conductivity and the associated flow of certain ions using so called Ussing Chambers [LSH04]. Furthermore, measured impedance spectra reveal information about physiological properties of the underlying cell tissue (e.g. membrane capacitance and subepithelial resistance) and can be used for various medical applications, such as detecting breast cancer [Rah+20] and muscular damage [Mły+19]. In the last decade, various cell cultures, e.g. HT29/B6 and IPEC-J2, were studied using impedance spectroscopy, in which the complex valued impedance is determined as a function of the frequency of an alternating current [Gün+12].

It has been shown that machine learning algorithms are able to analyse impedance spectra and approximate the sought physiological properties through patterns in the data [SBG13].

For this purpose, impedance spectra have been modeled to overcome the lack of expensive measured data, which are needed in a great amount for the training process [Sch18]. Thus, high accuracies in the determination could be achieved, but new synthesis methods are needed that allow a more realistic modeling of the cell tissues based on real measurements. Here we introduce a new approach to replicate impedance spectra using GANs in order to enlarge the data quantity. In this paper, synthetic data is used for this purpose, but the principle procedure could also be easily applied to real impedance measurements in the future.

This paper describes the implemented GAN and presents some of the generated spectra. Although the dimension of the data is comparatively low (84 values per sample against pictures with thousands of pixels), it was possible to create a GAN that accomplishes that challenge. We present an architectural small network that is able to create not distinguishable artificial impedance spectra from four randomly chosen values of latent space.

Section 2 introduces the data and in section 3, the functionality of the GAN is explained. The generated impedance spectra is presented and evaluated in section 4. A conclusion of the approach is given in section 5.

2 Data

For this work, a data set of modelled impedance measurements on the epithelial cell line HT29/B6 under physiological control conditions was adopted from [Sch20]. HT29/B6 is a well-studied carcinoma cell culture derived from the human colon. The synthesised data is based on experimentally estimated value ranges and additional error modelling which are both obtained from [Sch18]. The data set has the advantage that it has already been used to predict epithelial properties using machine learning and is extensively characterised (cf. [Sch20]).

The data set includes $150,000$ spectra, each consisting of 42 measurements taken at different frequencies ($1.3Hz - 16350Hz$). As complex values, impedances comprise real and imaginary parts. Impedance spectra are therefore often displayed in Nyquist representation, in which the imaginary part is plotted against the real part (Fig. 2). Looking at real and imaginary parts as separate features, this results in 84 values as input for the neural network.

3 Generative Adversarial Networks

Generative Adversarial Networks (GAN) are a special kind of Artificial Neural Networks [Goo+14]. Generally it is build out of two separate models, that are trained against each other.

There is a generator model G that takes random noise as input and is supposed to produce realistic but synthetic data as output, in our case the generated impedance spectra. The

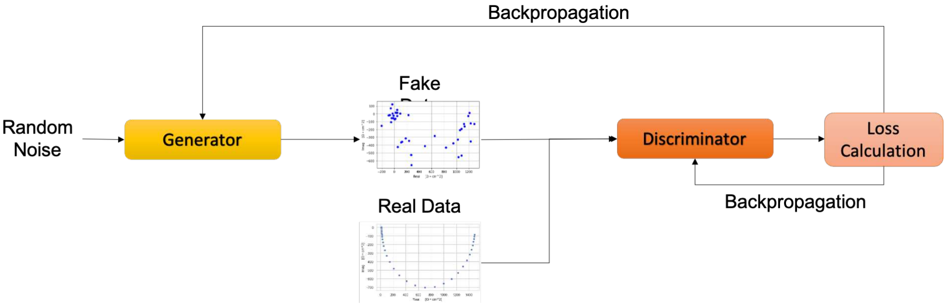


Fig. 1: Overview of GAN functionality, here for impedance spectra

discriminator model D then distinguishes whether the given data belongs to the original data or is fake data from the generator. The error of that model (i.e. incorrect identified fake and real data) is then propagated to itself and the generator to train its recognition of fake data and the production of synthetic data respectively. Here, the Binary Cross Entropy Loss (BCE Loss) was used as the error function, calculated on the basis of the correctly and incorrectly categorized data.
An overview of the GAN training for the special case of impedance spectra can be found in figure 1.

The model was build using the *PyTorch*[1] neural network framework. The generator model in this paper uses four randomly generated numbers (i.e. Noise) approximately in the range of the impedances as an input for the generator to generate the impedance spectra.

Compared to many state-of-the-art ANNs, its architecture is relatively small. It was build as a fully connected feedforward neural network, using a combination of Linear and LeakyReLU (Leaky Rectified Linear Unit) layers [NH10] as a base model for both, generator and discriminator. Additionally the discriminator uses the sigmoid function at the ouput layer for classification into real and fake. For testing different architectures, an ELU (Exponential Linear Unit) has been added to the generator of the base model. The differences of both models are discussed in section 4.
However, ultimately 5 Layers for the generator were used and 7 for the discriminator including input and output layer, so its training time was short although no GPU was used. The Generator uses $[4, 8, 16, 42, 84]$ neurons belonging to the layers and the Discriminator $[84, 64, 32, 16, 8, 4, 1]$ respectively. The complete model was trained with all given data samples and for 30 epochs with a batch size of 32, optimized using the AdaGrad optimizer [DHS11] with an initial learning rate of 0.001.

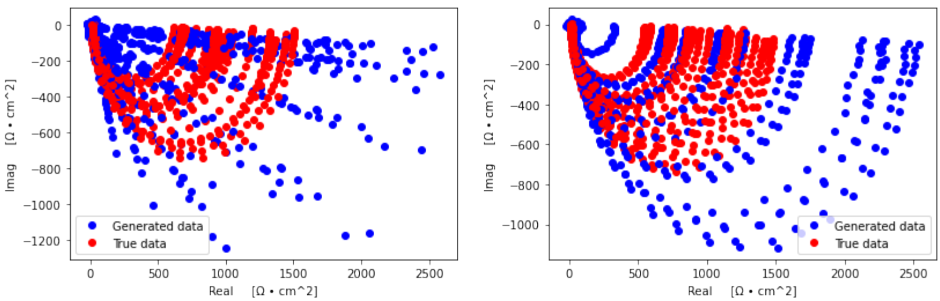
[1] `https://pytorch.org/docs/stable/nn.html`

4 Results

Within the process, two major GANs have been trained. Firstly, one GAN without an ELU Unit within the Generator and secondly, one where the ELU Unit has been added to the second-to-last layer due to the already successful training of other GANs [Agg+19]. For the target of generating new impedance spectra, the generator functionality has to be proofed. Therefore, the Results of the generator models have been compared graphically to the training data (Fig. 2). Twenty generated impedance spectra and another twenty randomly selected spectra from the original data set were chosen to be plotted.

While testing various units, the introduction of ELUs instead of solely using ReLUs was the key to train the right contexts from the data. This improvement is shown in Figure 2. Utilizing only ReLU units led to a generator model that learns a more linear structure than the typical semi-circle shape it is supposed to generate (Fig. 2(a)). Using an ELU Unit additionally, it turns out that the generated data has the supposed semi-circle shape but scales wider in its value range (Fig. 2(b)).

To validate the training process, the GAN losses were also recorded. Within the process, the losses of the generator and discriminator should adapt to each other as the generator improves and discriminator has a harder task to differentiate true from fake data. At the end of the process, the discriminator is not able to distinguish fake from real, so the discriminator gets ~ 50% right. As seen in figure 3, the Generator starts with a higher loss, but adapts to the discriminator, as the generation of fake data gets better. Obviously, the discriminator behaves vice versa. Note that it's typical for a GAN that the loss is rising in the beginning due to the adversarial training.



(a) First Approach using a LeakyReLU Unit (b) Second Approach including an ELU Unit

Fig. 2: Comparison of original (red) and replicated (blue) impedance spectra generated with a GAN using only LeakyReLU layers (a) vs. additional ELU layers (b).

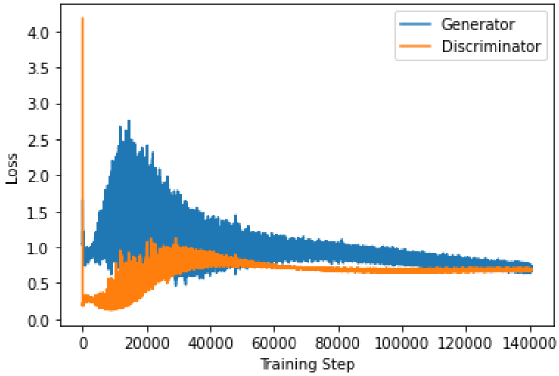


Fig. 3: Loss of the ELU Discriminator and Generator Model during the training process.

5 Conclusion

It has been shown that Generative Adversarial Networks are capable of replicating impedance spectra from modeled data. Even though the generated data scales wider then the true one (fig. 2(b)), it creates the right shape for various new data. Further data sets have to be investigated to validate, if this model is able to replicate spectra for other cell lines and cell conditions with more complex impedance curves.
Therefore, the method presented here can already be used to enlarge impedance data sets. However, the procedure developed in this paper must be seen as a first step. For future work, the won spectra have to be analyzed with proper quantitative metrics to determine if they are reliable compared to measured spectra. Additionally, the impedance curves must be mapped with the corresponding cell model parameters, in order to use the won data for regression tasks and pattern recognition with biological application such as in [SBG13].

Acknowledgement

We thank Thomas Schmid from the Machine Learning Group of Leipzig University for his supervision in the field of epithelial impedance spectroscopy, for providing the data set and for his beneficial advice during the implementation of the network.

References

[LSH04] Hongyu Li, David N Sheppard, and Martin J Hug. “Transepithelial electrical measurements with the Ussing chamber”. In: *Journal of Cystic Fibrosis* 3 (2004), pp. 123–126.

[NH10] Vinod Nair and Geoffrey E. Hinton. “Rectified Linear Units Improve Restricted Boltzmann Machines”. In: *Proceedings of the 27th International Conference on International Conference on Machine Learning*. ICML’10. Haifa, Israel: Omnipress, 2010, pp. 807–814. ISBN: 9781605589077.

[DHS11] John Duchi, Elad Hazan, and Yoram Singer. “Adaptive Subgradient Methods for Online Learning and Stochastic Optimization”. In: *Journal of Machine Learning Research* 12.61 (2011), pp. 2121–2159. URL: http://jmlr.org/papers/v12/duchi11a.html.

[Gün+12] Dorothee Günzel et al. “From TER to trans- and paracellular resistance: lessons from impedance spectroscopy”. In: *Annals of the New York Academy of Sciences* 1257 (June 2012), pp. 142–151. ISSN: 0077-8923. DOI: 10.1111/j.1749-6632.2012.06540.x. URL: https://doi.org/10.1111/j.1749-6632.2012.06540.x.

[SBG13] Thomas Schmid, Martin Bogdan, and Dorothee Günzel. “Discerning Apical and Basolateral Properties of HT-29/B6 and IPEC-J2 Cell Layers by Impedance Spectroscopy, Mathematical Modeling and Machine Learning”. In: *PLOS ONE* 8.7 (July 2013), pp. 1–12. DOI: 10.1371/journal.pone.0062913. URL: https://doi.org/10.1371/journal.pone.0062913.

[Goo+14] Ian Goodfellow et al. “Generative Adversarial Nets”. In: *Advances in Neural Information Processing Systems*. Ed. by Z. Ghahramani et al. Vol. 27. Curran Associates, Inc., 2014. URL: https://proceedings.neurips.cc/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf.

[Sch18] Thomas Schmid. “Automatisierte Analyse vonImpedanzspektren mittelskonstruktivistischen maschinellen Lernens”. Dissertation. Universität Leipzig, 2018. URL: https://nbn-resolving.org/urn:nbn:de:bsz:15-qucosa2-318388.

[Agg+19] Karan Aggarwal et al. “Regression with Conditional GAN”. In: *arXiv preprint arXiv:1905.12868* (May 2019).

[Mły+19] Marcel Młyńczak et al. “Impedance Spectroscopy Method to Detect Pelvic Floor Muscle Damage—A Feasibility Study”. In: *World Congress on Medical Physics and Biomedical Engineering 2018*. Singapore: Springer Singapore, 2019, pp. 875–878. ISBN: 978-981-10-9038-7.

[Rah+20] Sam Matiur Rahman et al. “Ensemble-Based Machine Learning Algorithms for Classifying Breast Tissue Based on Electrical Impedance Spectroscopy”. In: *Advances in Artificial Intelligence, Software and Systems Engineering*. Cham: Springer International Publishing, 2020, pp. 260–266. ISBN: 978-3-030-20454-9.

[Sch20] Benjamin Schindler. “Maschinelles Lernen zur paarweisen Analyse epithelialer Impedanzspektren”. German. MA thesis. Universität Leipzig, 2020. URL: https://nbn-resolving.org/urn:nbn:de:bsz:15-qucosa2-742028.

Informatik und Gesellschaft

Designing an ethical technology project with the help of data feminism

Experiences of the implementation of the data feminism concept in the students' project 'Questioning Street Names Leipzig'

Magdalena Bui [1], Lea-Kathrin Gleißner [2], Fey Kühn [3], Amelie Nenninger [4]

Abstract: Algorithms and new technologies help people in several life situations, but society pays a high price for their advantages. Several scandals occurred recently, showing that algorithms are neither neutral nor fair – quite the contrary: They discriminate people as humans do. One approach to create less biased data science projects is the "Data Feminism" method, presented by Catherine D'Ignazio and Lauren F. Klein in their book of the same title. This paper evaluates how feasible the method can be implemented in student projects based on the experiences four Leipzig students made by trying to implement the method into their project 'Questioning Street Names Leipzig'. The paper focusses on three main concepts: subjective viewpoints and context, crediting all forms of labour, and building and linking communities through public tagging events, thus opening the academic question for some citizen science help. The project utilizes open data and open data sources such as Wikidata and OpenStreetMap. The authors of "Data Feminism" want to encourage students, as well as academic professionals, to think about their bias in their data and to use the data feminism approach to reduce the impact of them and create more ethical computer science projects.

Keywords: data feminism, Citizen Science, OpenStreetMap, Wikidata, open data, report of experiences

1 Introduction

In December 2020, health care workers of the Stanford University Medical Center protested against the hospital's COVID-19 vaccine distribution plan made by an algorithm. Only seven out of over 1.300 frontline workers were chosen to receive one of 5.000 vaccinations [GH20]. In June 2020, an algorithm graded the A-levels in Great Britain to prevent grade inflation due to the pandemic. However, students from lower socio-economic backgrounds had their grades drastically downgraded in comparison to their

[1] Universität Leipzig, Fakultät für Mathematik und Informatik, Institut für Informatik, Augustusplatz 10, 04109 Leipzig, mb13vaxo@studserv.uni-leipzig.de
[2] Universität Leipzig, Fakultät für Mathematik und Informatik, Institut für Informatik, Augustusplatz 10, 04109 Leipzig, lea-kathrin.gleissner@studserv.uni-leipzig.de
[3] Universität Leipzig, Fakultät für Mathematik und Informatik, Institut für Informatik, Augustusplatz 10, 04109 Leipzig, kuehn@studserv.uni-leipzig.de
[4] Universität Leipzig, Fakultät für Mathematik und Informatik, Institut für Informatik, Augustusplatz 10, 04109 Leipzig, an40vyki@studserv.uni-leipzig.de

teachers' predictions. The results of pupils enrolled in private schools on the other hand were upgraded [Ma20]. In September 2020, Ph.D. student Colin Madland tweeted about how the face of his black colleague was not recognized in Zoom while using virtual backgrounds. While posting pictures of proof, Twitter's mobile version automatically cropped the picture, putting Madlands face in the centre while completely ignoring his colleague's face. Twitter denied finding evidence of racial bias in their software while testing but recognized the importance of further research [Di20]. These are several examples of how personal and social biases influence the work of computer science. Communities like the "Forum InformatikerInnen für Frieden und gesellschaftliche Verantwortung" are interested in critical discussions on these topics [FIfF]. Additionally, mathematicians Hannah Fry and Cathy O'Neil tackle this question with their books "Hello World" and "Weapons of Math Destruction". The film "Coded Bias", produced by Shalini Kantayya, also thematises these biases. Another publication from 2021 which deals with computer science systems and their social effects is "Your computer is on fire" by Thomas S. Mullaney, Benjamin Peters, Mar Hicks and Kavita Philip. Another new publication on how to solve the aforementioned problems caused by biased data is the book "Data Feminism" by Catherine D'Ignazio and Lauren Klein.

This paper reports on the experience of four German students who implemented the data feminism concept in a student project during the winter term 2020/2021. Data feminism is a new approach, published in 2020. According to the research of the authors of this paper, its implementation into student projects has been rarely discussed so far. The authors of this paper want to share the values and limits of the concept regarding their personal learning curve and their project 'Questioning Street Names Leipzig'. The project focusses on representation biases in street names concentrating on people included in the FLINTA*-acronym.The acronym is German and stands for for women, lesbians, intersex people, non-binary people, transgender people and agender people. Intersexuality is a medical term to describe people who were not born neatly fitting into the biological categories of 'male' and 'female' [HR17]. Non-binary people do not feel like fitting into the categories 'men' or 'women' [Gl] and agender people identify as having no gender [Fe21]. Furthermore, transgender people in general do not identify with the gender they were assigned at birth [Gl], in contrast to cisgender people, who do identify with their gender assigned at birth [Gl]. Cisgender men are not included in the FLINTA* acronym, because in contrast to FLINTA* they benefit from different privileges, because of their gender. Privileges grant advantages, benefits, and respect due to the sole reason of belonging to a specific social identity group, e.g. different religions, race, gender identity, sexuality, and class [Ho21].

According to the Merriam-Webster Dictionary bias is "a personal and sometimes unreasoned judgment [just like a] prejudice" [MW] The Project Implicit by Harvard University describes implicit bias as an unconscious factor in human decision making, shaped by different circumstances like cultural upbringing [Pr11]. It is hardly accessible for the human conscious awareness and can differ from one's explicit attitude, the behaviour one concisely thinks about, and acts upon.

First, the concept of data feminism is introduced and following this, the project ‘Questioning Street Names Leipzig’ is briefly described. The paper focusses on three main points: knowing one’s personal viewpoint and context, appreciating people and their labour, as well as networking and linking communities. It closes with a discussion of its own method and the data feminism approach and attempts to answer the question: “How and to what extend were the principles of data feminism implementable for the student project ‘Questioning Street Names Leipzig’?”

2 The method: data feminism

Data feminism, in the words of one of its authors, Catherine D’Ignazio, is "data science with an intersectional feminist lens. It takes all inequality into account at every stage of the data processing pipeline" [Co20]. It frames data as something inherently influenced by human bias. Therefore, the incidents described above are "results coming out of these algorithms with the flawed data we are feeding in" [Co20]. Klein and D’Ignazio have devised seven principles of data feminism.[5]

According to the authors, these principles should serve as tools for data scientists who want to stay mindful of the bias in their data [DK20]. The first principle, "examine power" [DK20], and the fourth, "rethink binaries and hierarchies" [DK20], ask the data scientist to examine the power structure around them and recognize biases in their surroundings [DK20]. The principles "elevate emotion and embodiment" [DK20] and "consider context" [DK20] teach that data is not cold and objective but informed by their context and as capable of evoking emotion which should not be shunned, but examined and valued [DK20]. "Embrace pluralism" [DK20] and "make labor visible" [DK20] focus on appreciating the value and work of everyone who led to the success of a project, from the scientists who laid the groundwork over local communities that assisted in gathering the data to the people who build computers and hardware. Through all these principles, data scientists can "challenge power" [DK20] and produce work that challenges the status quo and questions existing hierarchies, even in their own field [DK20]. Data feminism is a guideline for data science projects. It aims to credit everyone involved in a project equally, avoid reinforcing discrimination, make the field a more accessible and welcoming place for people from all backgrounds [DK20].

3 The project ‘Questioning Street Names Leipzig’

The first semester of the Digital Humanities programme at the University of Leipzig includes the module ‘Introduction to the Digital Humanities’. Students get an insight into

[5] An appropriate graphic can be found at:
http://datafeminism.io/wp-content/uploads/2020/12/DF_POSTER_ENGLISH.pdf

various disciplines and take their first steps in the digital processing of data. During the programme, students get together in groups and carry out a three-month project which they describe and reflect on in a project report.

Our project ‘Questioning Street Names’ comprises of two parts: The statistical analysis of bias and the development of an alternative, utopian map. From December 2020 until recently, we have been meeting weekly to move our project forward. During the week, we worked asynchronously on subgoals we discussed in plenary and formulated a new agenda for the next meeting. Completing a project exclusively at a distance and without the possibility of a single physical meeting brings many new challenges. We have been networking with people, projects, associations, and communities working on issues of equality, social recognition, and digital solutions. With the help of the open-source project EqualStreetNames.Brussels[6], we created a map of Leipzig that makes the gender bias in the Leipzig streetscape visible. It can be found at EqualStreetNames.Leipzig[7] website. For this purpose, we organised tagging days where we invited and guided people with different levels of experience in the digital field to link Geodata and Wikidata. During the tagging days, we contacted the OpenStreetMap community. One of us has acquired knowledge in HTML and CSS to enable the construction of our website. The website aims to present a utopian map, visualising a reverse score of the ratio in gender representation, to dignify people who are currently not visible in the cityscape and published it on the ‘Questioning Street Names Leipzig’ website[8].

4 Designing a data feminism project – key concepts

During the lecture series “Changing Course” at Brown University, D’Ignazio and Klein gave a talk on their book. They ended the talk with a slide called “What Students and Scholars Can Do” [Br20]. We want to use this slide as a basis, as it directly addresses students. While our project is explicitly feminist in its content, we believe that the principles can be implemented in all kinds of projects to create fair and less biased computer science systems.

4.1 Strong roots: Knowing about one’s own viewpoint and the context

Five principles were presented as suitable for projects conducted by students on the lecture slide. The principles “consider context” and “rethink binaries and hierarchies” were left out, but we have decided to include them in our work. In the beginning of our project, we contextualized ourselves, especially regarding the advantages and obstacles we

[6] https://equalstreetnames.brussels/
[7] https://leipzig.equalstreetnames.eu/de/index.html#10.05/51.3433/12.3895
[8] https://imlabormitlea-code.github.io/FLINTA-MAP-Leipzig/index.html

experience. We educated ourselves on the topics of feminism and algorithmic biases to form a stable basis for our work.

Our goal was to avoid looking at our data purely from only a technological point of view that disregards the sociocultural context of the data and our work. By doing so, we could avoid the so-called projection error. This bias was described by Caroline Criado-Perez in "Invisible Women" as the error that occurs when people design systems that fit them personally but fail to fit other parts of the population [Pe21]. We are planning on creating an explanatory text for the website EqualStreetNames.Leipzig, and we have shared our contextualisation on our own website to be transparent.

The principle "consider context" [DK20] goes hand in hand with "examine power" [DK20] in that they both focus on background knowledge of the system we live in. Our work concentrates on the existing unfairness regarding the representation of genders in the streetscape of Leipzig. Therefore, our interpretation of the principle fits well with the interpretation of D'Ignazio and Klein, which emphasises the interrogation of existing unfairness in the world [Br20].

Furthermore, we linked the two discussed principles with "elevate emotion and embodiment" during our project process. On their slide, Klein and d'Ignazio interpreted this principle as an encouragement to "[e]xperiment with creative forms of data presentation and communication" [Br20]. We experimented with colours and their feminist associations in our website design and plan to redesign the website of EqualStreetNames.Leipzig, putting more emphasis on the genders that are not visible in Leipzig's streetscape and improving the visualisation of transgender people. "Emotion and embodiment" [DK20] also refers to the idea that neither work nor knowledge can be neutral or objective [DK20]. As a group of FLINTA* who are directly affected by the bias we examined, our work was informed by our emotional attachment and reaction to the project matter itself, and we kept this in mind.

The previous examinations outline the first steps in the process of designing a Data Feminist project: identifying imbalances of power, educating oneself on social biases and recognizing the context of the work. From this basis, we can move on to the second step and to the principles "rethink binaries and hierarchies", "challenge power" and "make labor visible".

4.2 Steady growth: Appreciating people and their labour

The principle "examine power" strongly interacts with the principle "rethink binaries and hierarchies". As a project focused on gender biases, we committed to "challeng[ing] the gender binary" [Br20] in favour of challenging "other systems of counting and classification that perpetuate oppression" [Br20]. Hence, we examined the representation of FLINTA*, not just on the representation of cisgender women.

Data feminism asks scholars and students to “credit [their] data sources and research support staff” [Br20]. It taught us to credit every contribution to the project and we acknowledge the people who aided us at the end of this paper. Especially the work of women in the field of computer and data science is historically invisible, and how the labour of workers in the global south or from social lower classes is frequently forgotten when it comes to crediting data science work [DK20]. Anno Kirchner argues in their seminar work on memorial plaques in Jena that the current pandemic shows which work, done by whom, is valued by society [Ki20]. Furthermore, Criado-Perez examines data on invisible work in her book in a more detailed way [Pe19].

D’Ignazio and Klein implement the seventh principle “make labor visible” [DK20] with extensive and transparent citation politics, which features detailed explanations on whom they cited and which concepts they focus on [DK20]. Although the method was too time consuming for this project, we hope that more future works will include such transparency and academic accountability. The repository of ‘Questioning Street Names Leipzig’ will be published once the data preparation has reached a more complete state. The unpublished, detailed university report can be made available on request.

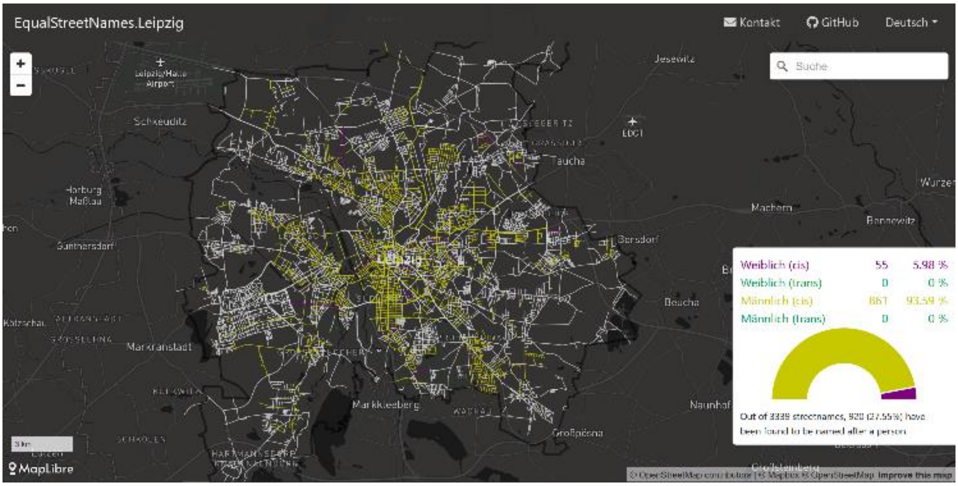


Fig. 1 Screenshot of the EqualStreetNames.Leipzig website (01.05.2021) [Eq21]

Our interpretation of “make labor visible” surpasses the topic of citation politics. It takes the phrase literally. The reproduction of EqualStreetNames.Leipzig shows via a statistic whose work and accomplishments are being honoured with a street named after them as shown in Fig. 1. We could identify two gender gaps. First, there is a gap between the representation of cis men (nearly 94% of street names) and cis women (nearly 6% of street names) [Eq21]. Taking a closer look at the non-visualized parts of the statistic, one can see that 0% of streets have been named after non-binary or transgender people [St21]. We

want to emphasise this lack of representation to call attention to these forgotten contributions and histories. This is our more extensive demand on the principle “Rethink Binaries and Hierarchies” [DK20].

During our project process, we have received plenty of positive feedback on our work from fellow students, academics in computer science and humanities fields and from people who work outside of academia altogether. Despite the obstacles posed to us by the pandemic, we could create a project from which others benefit. Through this, we identified the value student projects can have for academic conversations. We had the advantage of being able to experiment with concepts and ideas and contribute to Digital Humanities conversations. Our background in the humanities proved more helpful when it came to examining reasons for the uneven representation of genders, but we could not have completed our analysis and visualisation without our computer science skills. We learned to value the disciplines equally and believe that the Digital Humanities can be a bridge between the humanities and computer science. Sarah A. Lang also shows that there is a perceived hierarchy between those disciplines, which also relates to gender topics [La20]. Working on our project showed us that student projects are capable of challenging those hierarchies in academia.

With the second part of our project, we served the principle “challenge power” by creating the base for an alternative map and “collect[ing] counter data” [Br20]. Once completed, this map aims to simulate how a streetscape could look like if the work of FLINTA* was honoured and appreciated like the accomplishments of cisgender men are today. It will also introduce the historical figures we chose, thus making their labour visible.

The next paragraph will show that we “introduce[d] new communities to data and digital tools” [Br20]. Since our group cannot represent all viewpoints, we reached out to several FLINTA* groups via messenger app. We were interested in their input and perspective for the alternative map. This brings us to the principle “embrace pluralism”.

4.3 Bearing fruit: Networking and linking communities

“Embrace pluralism” [Br20] invites us to “[i]nclude more people in data projects – impacted communities especially [and c]enter the work of minoritized people and follow their leadership” [Br20]. Multiple points of view were included by asking for proposals for the alternative map. As a group of minoritized people, we recognized each other as equally competent and elected to decide against appointing a leadership role. Instead, we learned from each other and grew to respect and appreciate our different perspectives.

Even though the proposals from other FLINTA* for the alternative map gave us pertinent insights and research material, our tagging days were more important regarding knowledge transfer. The algorithm of the EqualStreetNames projects works with the ‘sex or gender’ tag from Wikidata [Se21]. To link the streets matching the gender of the person they are named after, Wikidata must be linked to the corresponding street on

OpenStreetMap. For this purpose OpenStreetMap provides the 'name:etymology:wikidata' tag.

Leipzig has around 3000 streets and squares [St18]. The meaning and history of these names are listed in the "Verzeichnis Leipziger Straßennamen mit Erläuterungen" (list of Leipzig street names with explanations) [St18]. The city of Leipzig also provides an XCSL file on their website in which they organize every street by district. We used this sheet to track our progress. Tagging all the streets is a laborious task and in a team of four, we would not have been able to finish the whole cityscape alone. In order to accomplish our goal of connecting communities, sharing knowledge and creating the map, we organized the tagging days.

During the two events, 17 volunteers from different communities helped us tag the streets. We shared our appeal to join the effort via different channels, including messenger apps, Digital Humanities mailing lists and the student black board of the University of Halle. From there, our appeal was spread further into the OpenStreetMap and Wikidata communities. We were able to connect people from various communities, skill levels and backgrounds during these events. People who had never worked with OpenStreetMap before were able to learn technological skills, and people who did not know about gender bias gained awareness. As shown below, those events were successes, both for our project and for data feminism, as we were able to "introduce new communities to data and digital tools" [Br20], and we will include community building events in future projects. Still, there are a few shortcomings of these events to be considered. Firstly, the organisation of such an event is a significant workload for which students and scholars might not have time. We were, for example, unable to generate sufficient metadata to evaluate the events. Secondly, work done by human beings – especially by beginners – is prone to errors, and quality control can be difficult. Despite these shortcomings, we value citizen science to get student projects out of the seminar room as a very powerful method.

Georg Verweyen, an active member of the OpenStreetMap community, assisted us with quality assurance. He also created graphics showing the improved ratio of tagged streets in Leipzig during our tagging events, as shown in Fig. 2. The graph shows that with increasing time of the day, more and more streets were tagged. We started both of our tagging events at ten in the morning. The biggest incline in percentage occurred around lunchtime (one pm till two pm). On the first day, we managed to tag around twelve percent of OpenStreetMap ways. On the second event day, sixteen percent were tagged. Many of the participants returned from the last tagging day, and we were able to streamline our explanation process and start much sooner. Around ten percent of ways were tagged in the time between our events. A few participants of our first tagging day expressed that they wanted to continue after the event. At the end of the second event, we managed to tag all streets in Leipzig who were named after a person that has a Wikidata entry. Only the ones who do not have a Wikidata entry yet are not tagged.

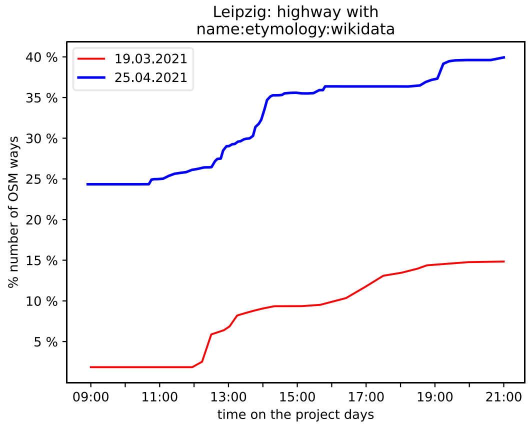


Fig. 2 Monitoring of the improvement of tagged streets in Leipzig during our project days [Ve21]

5 Method discussion

This paper discusses the experience of four German first semester students with the implementation of data feminism. Therefore, the thoughts and opinions discussed in this paper are difficult to reproduce. Adhering to the “elevate emotion and embodiment” and “consider context” principles of data feminism, we never claimed neutrality in our project, knowing that this is not common in computer science. Before we thought about the implementation, we discussed the value of data feminism and aimed to follow the principles. After finishing parts of the project, we grew more and more aware that the concept itself and its implementation is worth further examination and interesting for our academic colleagues. Due to this late insight, we did not produce metadata or worked on an evaluation concept for data feminism before starting to work on ‘Questioning Street Names Leipzig’. For this reason and the fact that our students’ group was very small, this paper is limited in its reproducibility.

We suggest collecting metadata both during community events and during the process of the project. During the tagging events, we want to collect data to better measure the impact and success of the events. For this, we recommend building a questionnaire that participants can fill out before the actual event where data on already existing knowledge and motivation is collected. We believe that questions that generate either binomial data or ordinal data would be sensible. After the events, we would like to collect data on how long participants worked, what they learned and gather extensive feedback on the perceived quality of the event. More important, in our opinion, is the evaluation of the process of the student working on the data feminist project. This could be done with a questionnaire that can be repeated before, during and after the project. Before the project starts, we suggest questions that determine how much the student knows about data

feminism, how important they think it is, how it is relevant for their work and which principles they deem most important. At the same time, we want to track where the student sees problems and difficulties and which aspects of data feminism they do not deem significant for their work as of now. For those questions, we also suggest the collection of ordinal data via sliding scale. During and after project, the questions of importance, relevancy and significance can be repeated to track how the student body's opinion changes during the project.

The data feminism concept was developed in an American intersectional feminist context and needs to be adapted for German and European needs. We considered the special history Germany and Leipzig as an Eastern German city have when we researched the development of the street names in the Leipzig cityscapes. We believe the principles themselves to work for multiple countries and societies, but we agree with Mithu Sanyal: people should not entirely avoid reading American theories on these social issues, but they should not exclusively read American theories [Sc21]. German publications dealing with feminism and social issues in computer science are e.g. "Netzpolitik – Eine feministische Einführung" by Francesca Schmidt or "Wenn KI dann feministisch – Impulse aus Wissenschaft und Aktivismus" by netzforma* e.V.. Accountability should not be built from a single idea [Br17] and other methods and concepts like "Explainable AI" can be discussed[9] and merged with data feminism issues to create individual project solutions for less biased computer science projects.

6 Conclusion

"Data science is the art and science of turning data into insights" [MS20]. The project 'Questioning Street Names Leipzig' turns street name data into insights about gender bias by making them visible on EqualStreetNames.Leipzig and gives an alternative with its own website 'Questioning Street Names'. It is a data science project and as such it must be critical about the impact it can have. Since data science and its possible biases gained more attention over the last years, we consider the implementation of methods that deal with these biases of great importance. The scandals and events shown in the introduction are just a few examples. Data feminism and its seven principles are one way to rethink (data) science and build more ethical digital projects. This paper discussed the question: "How and to what extend were the principles of data feminism implementable for the student project 'Questioning Street Names Leipzig'?"

Implementing data feminism is a laborious and time-consuming task. D'Ignazio and Klein do not require every data scientist and student in particular to completely fulfil all demands of their concept, as it can easily by overwhelming. Therefore, everyone must decide how to deal with limited time, financial resources and pressure to publish projects [Br20]. Acknowledging that, we want to encourage academics, data scientists and especially

[9] Francesca Schmidt and Prof. Dr. Frieder Stolzenburg discussed this topic in the context of the lecture series "We need to talk about AI" from the "KI & Wir" network [Ri].

students to implement and adapt the principles to their own needs. The authors provide additional information in the form of a reading group, the recordings are available on YouTube [Ca]. Moreover, they gave talks at many universities and institutions like the Brown Institute [Br20].

Our project was valued for its impact even though it is technologically simple. As students, we have more freedom to experiment with different methods and to create original projects. The citizen science approach enabled us to include viewpoints from outside of a university context, thus enabling us to challenge traditional academic barriers. Especially our Digital Humanities standpoint served us well to link social discourses and technological issues. The skills we learned from the data feminism concept were mostly new to us and thereby challenging. Considering one's own viewpoint and bigger contexts, making labour visible through citation politics and acknowledgements, doing work addressing social problems and including a community of people in a project – all of those are methods that we will incorporate in future projects. Returning to the scandals and the discrimination caused by algorithms that we mentioned in the introduction, we think data feminism can help students in their education and in building a consciousness for the effects their data science projects can have. Therefore, we hope the concept will be integrated into university teachings.

Acknowledgements

We want to thank those who developed ethical, (data) feminist and critical computer science concepts and the people who introduced us to them. These people were authors, content creators, podcasters, activists, and friends, who generated and shared educational work with us – partly for free. The team of the 'Informatica Feminale', above all Silvia Bauer and Nadja Geisler, introduced parts of our team to the field of feminist and critical computer science.

We especially want to thank Catherine D'Ignazio and Lauren Klein. We thank Jonathan Beliën from EqualStreetNames Belgium and the volunteers who generated linked data during our tagging events. We have benefitted from free access to tutorials on HTML and CSS. Thanks to Thomas Köntges for teaching the class 'Introduction to Digital Humanities' and his support and Anno Kirchner for making their work available to us. We are grateful for the interest in our project and the platforms offered to us to spread our thoughts and project work by the Frauenkultur Leipzig e.V., the OpenStreetMap Germany community, the Saxorum blog and the team of #empowerDH. We appreciate #empowerDH and Jana Keck for their input. Many thanks go to Georg Verweyen for providing graphics and assisting with quality control.

References

[Br17] Brown, A. M.: Emergent strategy. Shaping Change, Changing Worlds, AK Press, 2017.

[Br20] Brown Institute: ‘Changing Course’: Lectures on Data Representations and Equitable Approaches to Computing - Data Feminism Catherine D’Ignazio, MIT and Lauren Klein, Emory University, https://brown.columbia.edu/changing-course/, last access: 01.05.2021

[Ca] Catherine D’Ignazio, https://www.youtube.com/user/kanarinka, last access: 01.05.2021

[Co20] Corbyn, Z.: Catherine D’Ignazio: 'Data is never a raw, truthful input – and it is never neutral’ https://www.theguardian.com/technology/2020/mar/21/catherine-dignazio-data-is-never-a-raw-truthful-input-and-it-is-never-neutral, last access: 01.05.2021

[Di20] Dickey, M. R.: Twitter and Zoom's algorithmic bias issues, https://techcrunch.com/2020/09/21/twitter-and-zoom-algorithmic-bias-issues/, last access: 01.05.2021

[DDK] D’Ignazio, C., Klein, L. F. Diaz, M.: Data Feminism infographics, http://datafeminism.io/blog/book/data-feminism-infographic/, last access: 01.05.2021

[DK20] D’Ignazio, C., & Klein, L. F.: Data feminism, Mit Press, Cambride, 2020.

[Eq21] EqualStreetNames.Leipzig, https://leipzig.equalstreetnames.eu/en/index.html#10.28/51.3433/12.3895, last access: 04.05.2021

[Fe21] Ferguson, S.: What Does It Mean to Be Agender?, https://www.healthline.com/health/agender, last access: 04.05.2021

[FIfF] Forum InformatikerInnen für Frieden und gesellschaftliche Verantwortung e.V.: Wir über uns https://www.fiff.de/about, last access: 04.05.2021

[Fr19] Hannah Fry: Hello World, How to be Human in the Age of the Machine, W. W. Norton & Company, New York City, 2018

[GH20] Guo, E., & Hao, K.: This is the Stanford vaccine algorithm that left out frontline doctors, https://www.technologyreview.com/2020/12/21/1015303/stanford-vaccine-algorithm/, last access: 28.04.2021

[Gl] Glossary of terms, https://www.stonewall.org.uk/help-advice/faqs-and-glossary/glossary-terms, last access: 04.05.2021

[Ho21] Hofmann, M. A., Privilege and Intersectionality, https://guides.rider.edu/privilege, last access: 03.05.2021

[HR17] Human Rights Watch: I Want to Be Like Nature Made Me, https://www.hrw.org/report/2017/07/25/i-want-be-nature-made-me/medically-unnecessary-surgeries-intersex-children-us#:~:text=The%20genital%20surgeries%20done%20on,limit%20options%20for%20future%20surgery, last access: 04.05.2021

[La20] Lang, S. A.: The Computational Humanities and Toxic Masculinity - A Long Reflection: https://latex-ninja.com/2020/04/19/the-computational-humanities-and-toxic-masculinity-a-long-reflection/, last access: 29.02.2021

[Ki20] Kirchner, A.: Jenas Gedenktafeln aus (queer-)feministischer Perspektive - Missverhältnisse und Potentiale einer regionalen Erinnerungskultur, Universität Jena, Jena, 2020

[Ma20] Mahdawi, A.: It's not just A-levels – algorithms have a nightmarish new power over our lives, https://www.theguardian.com/commentisfree/2020/aug/19/its-not-just-a-levels-algorithms-have-a-nightmarish-new-power-over-our-lives, last access: 01.05.2021

[MS20] Magenheim J. & Schulte C.: Data Science Education. In: Tatnall A. (eds) Encyclopedia of Education and Information Technologies, Springer, Cham, https://doi.org/10.1007/978-3-030-10576-1_253

[MW] Merriam-Webster: Bias, https://www.merriam-webster.com/dictionary/bias, last access: 04.05.2021

[Pe21] Perez, C. C.: Invisible women, Exposing data bias in a world designed for men. Chatto & Windus, London, 2019.

[Pr11] Project Implicit: Frequently asked questions, https://implicit.harvard.edu/implicit/faqs.html#faq13, last access: 03.05.2021

[Ri21] Ringvorlesung Aufzeichnung „Explainable AI: Eine mögliche Lösung des Diskriminierungsproblems?", https://ki-convention.com/ringvorlesung-aufzeichnung-explainable-ai-eine-moegliche-loesung-des-diskriminierungsproblems/, last access 03.05.2021

[Sc21] Scheller, J.: «Identitätspolitik bringt uns dazu, in Mustern zu sprechen, die wir eigentlich sprengen wollen»: Mithu Sanyal befreit ein Reizthema von Tabus, https://www.nzz.ch/feuilleton/identitaetspolitik-mithu-sanyal-befreit-die-debatte-von-tabus-ld.1605577, last access: 03.05.2021

[Se21] sex or gender (P21), https://www.wikidata.org/wiki/Property:P21, last access: 03.05.3021

[St18] Stadt Leipzig Amt für Statistik und Wahlen: Verzeichnis Leipziger Straßennamen mit Erläuterungen, https://static.leipzig.de/fileadmin/mediendatenbank/leipzig-de/Stadt/02.1_Dez1_Allgemeine_Verwaltung/12_Statistik_und_Wahlen/Raumbezug/Leipziger_Strassennamen.pdf, last access: 03.05.2021

[St21] Statistics.json, https://github.com/EqualStreetNames/equalstreetnames-leipzig/blob/master/data/statistics.json, last access: 03.05.2021

[Ve21] Verweyen, G.

Die BYTE Challenge – ein digitaler Technik-Wettbewerb

Stefan Hildebrand[1], Carolin Neumann[2], Luisa Wolf , Annika Tauchnitz , Robin Krüger

Abstract: Die BYTE Challenge ist ein digitaler Wettbewerb für Schüler*innen aus ganz Deutschland, der Informatik, Informationstechnische Grundlagen sowie deren gesellschaftliche Bedeutung vermittelt und auf die Reduzierung bestehender Ungerechtigkeiten im MINT-Bereich hinwirken soll. Dazu soll die Teilnahme in jeder Hinsicht niedrigschwellig möglich sein; die Teilnahme ist kostenlos.

Keywords: Informatik für alle; Digitale Lehre; Informatische Bildung; MINT; Andere Fachrichtungen; Unterrichtsaktivitäten

1 Motivation und Zielsetzung

Der Bedarf an Informatischer Bildung wird unterschiedlich begründet, vom Training von Problemlösefähigkeiten als Teil der Allgemeinbildung bis zur besseren Nutzung digitaler Werkzeuge. [SR18] Die Vermittlung von Informationstechnischen Grundlagen (ITG) sowie Informatik erfolgt in Deutschland sehr unterschiedlich zwischen den Bundesländern, Schulformen und Klassenstufen, zumindest ITG ist an den meisten Schulen vorgesehen. [Wi21][SH21]

Es bestehen im MINT-Bereich dauerhaft Disparitäten in der Bildung und Auswahl von Studienfächern. In der Informatik beträgt der Mädchenanteil unter den Studienanfänger*innen nicht einmal 20% (WS 2011/12 sowie WS2018/19) [St19, S. 240] Ebenso bestehen deutliche soziale und erhebliche zuwanderungsbezogene Disparitäten. [St19, S. 271] [St19, S. 304 f.] Weiterhin herrscht ein Mangel an Informatiklehrer*innen [Mü17], obwohl es einen Fachkräfteengpass bei Berufen im Bereich der Informatik gibt, sowohl mit, als auch ohne Hochschulabschluss. [19b, S. 18 f.]

Daraus ergibt sich die Frage, wie Informatische Bildung und auch Interesse an MINT-Fächern für mehr Schüler*innen zugänglich gemacht und damit mehr Bildungsgerechtigkeit geschaffen werden kann: Für Schüler*innen, an deren Schulen zu wenig Informatik-Lehrkräfte zur Verfügung stehen oder Informatik-Unterricht nur in geringem Umfang vorgesehen ist, für Mädchen sowie für Kinder mit sozialen und zuwanderungsbezogenen Nachteilen.

[1] Gesellschaft für Informatik e.V., Hochschulgruppe Berlin-Brandenburg, Anna-Louisa-Karsch-Straße 2, 10178 Berlin, Deutschland stefan.hildebrand@byte-challenge.de
[2] Gesellschaft für Informatik e.V., Hochschulgruppe Berlin-Brandenburg, Anna-Louisa-Karsch-Straße 2, 10178 Berlin, Deutschland info@byte-challenge.de

Als eine Möglichkeit dafür wird im vorliegenden Papier die BYTE Challenge vorgestellt. Das ist ein Wettbewerb für Schüler*innen der Sekundarstufe I (ab 2022 auch Sek. II), der Informatische Grundkenntnisse vermitteln und die Teilnehmenden dafür begeistern soll.

Die BYTE Challenge findet zum ersten Mal vom 18.03.2021 bis zum 18.06.2021 statt. Die bisherigen Erfahrungen und Auswertungen werden daher als Zwischenstand dargestellt.

2 Konzept

Das Projekt ist vollständig ehrenamtlich aufgebaut, soll aber möglichst viele Schüler*innen deutschlandweit erreichen. Daher ist es naheliegend, Angebote auf einem digitalen Konzept aufzubauen. Beispielsweise bieten Education-Start-Ups zum staatlichen Bildungssystem alternative Wege zur Wissensvermittlung und zeichnen sich durch eine hohe Anwendungsorientierung, den Fokus auf die Vermittlung technologischer Zukunftsfähigkeiten sowie niedrige finanzielle Einstiegsbarrieren aus [19a].

In der Pandemie ist auch die schulische Kommunikation zwischen Lehrer*innen und Schüler*innen außerhalb des Präsenzunterrichts besonders in den Vordergrund gerückt, in digitaler Form gleichberechtigt neben der konventionellen analogen Form [20, S. 10 f.]. Dazu werden vielerorts mindestens ergänzend digitale Kommunikationsmittel eingesetzt [BCW21, S. 3][We21] Die digitale Kommunikation erweist sich als tauglich für die Lehre, von der Augenheilkunde [Mo21a] bis zur Mathematik [We21]. Das digitale Home Schooling zeigt sich beispielsweise im Mathematikunterricht als besonders fruchtbar, beim aktiven Lernen, z.B. mithilfe interaktiver Lernsoftware wie GeoGebra. „Zusammenfassend ermöglicht durch die Pandemie beschleunigte Etablierung digitaler Lehre nicht nur die digitale Transformation bestehender Lehrformate, sondern auch die Entwicklung innovativer, inhärent digitaler Formate.“ [Mo21b]

Daher wird die BYTE Challenge als komplett digitaler Wettbewerb angeboten. Dieses Format bietet einerseits die Möglichkeit, einen dritten Vermittlungskanal neben Erziehungsberechtigten und Lehrer*innen zu nutzen. Als extrinsische Motivation können dabei Preise vergeben werden, sodass eine Entkopplung vom Schulnotensystem möglich ist. [Ko16]

Die BYTE Challenge hat eine Dauer von drei Monaten, in denen Kurse und Aufgaben aufgeteilt in drei zeitlich parallele Kategorien angeboten werden: Coding, C+ und Online-Seminare. Diese lassen sich auch in die von Seegerer; Romeike [SR18] verwendeten Kategorien G1 bis G4 einteilen, s. Tab. 1.

Darüber hinaus bieten wir eine Reihe von Orientierungsangeboten an, um die Möglichkeiten nach der Schule und für freiwillige Engagements kennenzulernen, über das Gebiet der Informatik hinaus.

Kategorie	Beschreibung	BYTE-Kurse
G1 – Denkweisen	z.B. Computational Thinking, algorithmisches, kreatives, problemlösendes Denken	Coding-Kurse
G2 – Fluency	tieferes Verständnis verwendeter Technologien; Studierende/ Schüler*innen sollen befähigt werden, Informatiksysteme effizient und gewinnbringend zur Lösung von Problemen einzusetzen	
G3 – Wissenschaft	Aufzeigen von zentralen Ideen und Schlüsselkonzepten der Wissenschaft Informatik und Vermittlung eines breiten Bildes der Disziplin und grundlegender Konzepte	C+ Kurse
G4 – Gesellschaft	Verständnis für den Einfluss und die Auswirkungen von Informatik und Informatiksystemen auf die Gesellschaft und das persönliche zukünftige Leben	Online-Seminare

Tab. 1: Zuordnung Kategorien – BYTE-Kurse

2.1 Teilnahmevoraussetzungen

Für die Förderung von mehr Bildungsgerechtigkeit wird ein inklusiver Ansatz verfolgt, d.h. die Angebote stehen allen Schüler*innen offen, müssen dabei aber besonders attraktiv und hilfreich für diejenigen sein, die nicht bereits umfassende Förderung und Vorkenntnisse im Informatik-Bereich erhalten haben.

Viele bereits existierende Wettbewerbe und außerschulische Angebote hingegen thematisieren nur ein spezielles Gebiet oder erwarten: Grundkenntnisse von Beginn an, eine betreuende Lehrkraft, spezielle Hardware oder eine Anfahrt an einen bestimmten Ort.

Dies ist jedoch nicht vereinbar mit dem Ziel der Förderung benachteiligter Kinder mit wenigen Vorkenntnissen. Die BYTE Challenge achtet daher auf möglichst niedrige Teilnahmehürden. Es werden keine fachspezifischen Vorkenntnisse vorausgesetzt. Zur Teilnahme werden ausschließlich Web-Anwendungen verwendet, die in allen gängigen Browsern, den gängigen Betriebssystemen und auf einer Vielzahl von Endgeräten (Tablet, PC, Laptop) ohne besondere Hardwareanforderungen lauffähig sind. So besteht auch keine Ortsbindung.

Die Teilnahme kann selbstständig oder zusammen mit einer betreuenden Lehrkraft erfolgen.

2.2 Preise

Durch die Teilnahme an den Kursen der BYTE Challenge können die Schüler*innen Punkte in unserer eigenen digitalen Währung ‘Byte’ sammeln, aufgeteilt in 8 ‘Bits’ pro ‘Byte’. Dabei treten die Schüler*innen nicht gegeneinander an, sondern erhalten unabhängig voneinander Punkte nur für ihre eigenen Ergebnisse. Anschließend können sie selbstbestimmt ihre Preise auswählen, die im Anschluss nach Hause geliefert werden. Dazu wird eine Art Online-Shop eingerichtet, in der ausschließlich mit “Byte”-Gutscheinen bezahlt werden kann. Bei

der Gestaltung des dort erhältlichen BYTE-Sortiments werden die Teilnehmenden mit einbezogen und auf ihre Wünsche explizit eingegangen. Dieses Konzept ermöglicht, dass nur wirklich erwünschte Preise versendet werden. Diese Produkte (Preise) im BYTE-Sortiment sind nirgendwo sonst erhältlich und können auch nicht gegen Geld erworben werden.

2.3 Infrastruktur

Die Inhalte und Prüfungen des Wettbewerbs werden in einem Lernmanagementsystem (LMS) zur Verfügung gestellt. Die Programmieraufgaben können über die Scratch-Webseite (www.scratch.mit.edu) oder eine Scratch-App bearbeitet werden. Für die Live-Seminare kommen zusätzlich Live-Streams auf YouTube, Wooclap und “Frag Jetzt” zum Einsatz. Darüber hinaus gibt es ein Chatforum für die Teilnehmenden, die begleitenden Lehrkräfte und die ehrenamtlich Unterstützenden, um Fragen schnell und direkt beantworten zu können – hier sind die Teilnehmenden vollständig pseudonymisiert. Die Registrierung zur Teilnahme, die datenschutzrechtliche Einwilligung der Erziehungsberechtigten, der Upload der Scratch-Spiele, der Online-Shop für die Preise sowie Kontaktformulare für Fragen und Anmerkungen sind Bestandteile der Homepage.

2.4 Coding-Kurse

Die Coding-Kurse der BYTE-Challenge sollen die Programmierung in Scratch von null an vermitteln. Sie sind aufgeteilt in drei Phasen. In Phase 1 erhalten die Teilnehmenden eine Einführung in die Programmierumgebung und häufig verwendete Scratch-Blöcke, die in kurzen Fragmenten miteinander verbunden und ausprobiert werden. In Phase 2 wird nach Anleitung ein kleines Spiel aufgebaut. In Anlehnung an das Konzept der Anchored Instruction (Gallenbacher - Anchored Instruction) entsteht von den ersten Blöcken in Phase 1 bis zum Ende von Phase 2 schrittweise und ohne Kontextwechsel das vorgegebene Spiel, das selbst eine (sehr kleine) Geschichte erzählt. Zu Beginn stellen wir einen Charakter vor, der im Spiel über die Tastatur gesteuert wird und die*den Teilnehmer*in repräsentieren soll. Um eine implizite Diskriminierung durch den Avatar zu vermeiden, haben wir im Icon-Paket Figuren unterschiedlicher Hautfarbe, gelesener Geschlechter sowie Fortbewegungsarten (laufend, im Rollstuhl) berücksichtigt, in den Tutorials verwenden wir mehrere verschiedene dieser Icons. Die Auswahl überlassen wir jedoch den Teilnehmenden selbst. In der Geschichte soll die Hauptfigur über einen fiktiven Marktplatz laufen, Hindernissen aus dem Weg gehen und einen Kuchen zu einem Ziel bringen. Um auf Umweltverschmutzung aufmerksam zu machen, wird herumliegender Müll eingesammelt. Im Laufe der Programmierung des Spiels entlang der Geschichte werden grundlegende Konzepte wie Schleifen und Abfragen eingebunden und erklärt. Algorithmen spielen jedoch keine große Rolle und wir legen keinen Wert auf Datenstrukturen. Dank der einfachen Bedienung von Scratch rückt das Spiel in den Vordergrund, sodass die Programmierung eher im Hintergrund erlernt wird. In der abschließenden Phase 3 sollen die Teilnehmenden ihr erlerntes Wissen anwenden

und kreativ eigene Ideen einbringen, indem sie ein eigenes Spiel programmieren. Dabei können sie aus dem Ergebnis von Phase 2 aufbauen oder etwas komplett eigenständiges programmieren

Phase 1 und die erste Hälfte von Phase 2 werden in Form von Video-Tutorials sowie alternativ in Form von bebilderten Texten vermittelt. Die zweite Hälfte von Phase 2 wird nur noch anhand von Text-Bild-Tutorials vermittelt, Phase 3 enthält lediglich die Aufgabenstellung, aber kein neues Material mehr. Die Lernerfolgskontrolle und Bewertung erfolgt für Phase 1 und die erste Hälfte von Phase 2 mithilfe von Quizzes im Lernmanagementsystem, die fertigen Spiele aus Phase 2 und Phase 3 hingegen werden von uns anhand einheitlicher Bewertungsbögen von Hand korrigiert.

2.5 Online-Seminare

Die Online-Seminare behandeln Themen an der Schnittstelle von Informatik und Gesellschaft. Dazu gehören einerseits digitale und soziale Kompetenzen, die erforderlich sind, um sicher und erfolgreich mit digitalen Medien umzugehen. Hierzu behandeln wir Themen wie digitale Souveränität und Cybermobbing. Weiterhin werden Anwendungsbereiche von Technik präsentiert, z.B. das Potenzial von Informatik in der Landwirtschaft oder im Umweltschutz.

Darüber hinaus bieten wir Online-Seminare zur Orientierung an, wo wir Studienmöglichkeiten vorstellen und Programme, um sich ehrenamtlich oder freiwillig zu engagieren. Dabei steht stets das Interesse an Technik im Vordergrund.

Insgesamt werden acht Kurse mit folgenden Themen angeboten: Mindset, Digitale Souveränität, Orientierungsstudium, Ausbildung und Studium, Freiwilligendienste, Digitale Zivilcourage, Ehrenamt sowie Umwelt und Technik

Die Online-Seminare erfolgen als Livestream auf YouTube, werden aber auch über das LMS veröffentlicht. Während des Livestreams können die Teilnehmenden Fragen via „Frag Jetzt“ stellen, die entweder live durch die Referent*innen oder im Anschluss von uns schriftlich beantwortet und veröffentlicht werden. Um auch gehörlosen Teilnehmenden die Inhalte zugänglich zu machen, werden die Videos untertitelt. Während des Live-Streams werden außerdem mittels Wooclap-Umfragen und Quizze durchgeführt und die Ergebnisse im Live-Stream eingeblendet. So wird gleich von Beginn an ein Bezug zu den Teilnehmenden hergestellt. Eine Woche vor dem Live-Termin wird im LMS ein Vorbereitungsquiz veröffentlicht, für die Teilnahme gibt es ein “Bit” als Belohnung. Nach dem Seminar findet zu jedem Thema ein kurzes Überprüfungsquiz als Lernerfolgskontrolle mit stufenweiser Bewertung statt.

Damit soll die Interaktion als signifikanter Faktor für die Effektivität einer Lehrmethode [Mo21b] verstärkt werden.

2.6 C+ Kurse

C+ Kurse sind Technikvertiefungskurse, die einen Einblick in gesellschaftlich interessante Themengebiete bieten sowie die Vielfalt der Informatik und ein Grundverständnis für bestimmte neue Technologien vermitteln.

Technologien bekommen durch die Vermittlung in den Medien eine größere Bedeutung. Beispielsweise durch die "Digitalstrategie der Bundesregierung wird das Thema Künstliche Intelligenz (KI) auch für die Schule zunehmend relevant." [SLR19] Auch für die berufliche Entwicklung der Teilnehmenden sind die Themen der C+ Kurse relevant, denn „Die Bedeutung von Informatik nimmt nicht nur in immer mehr Bereichen unseres täglichen Lebens zu, sondern auch in immer mehr Ausbildungsrichtungen.“ [SR18]

Themen wie KI, Virtuelle Realität, Computerbestandteile, Internet der Dinge oder digitale Währungen finden im Lehrplan aber bislang keinen Platz, weswegen wir sie in den C+ Kursen behandeln.

Die Kurse basieren auf Interviews mit Fachleuten auf den jeweiligen Gebieten oder sind aufgrund unserer Recherchen als Erläuterungen aufgebaut. Die Bewertung ist bewusst „leicht“ gehalten, die Inhalte werden mithilfe von TikTok-Videos und Quizzes aufgelockert, um zur Teilnahme zu motivieren.

2.7 Finanzierung & Ehrenamt

Um Disparitäten aufgrund der sozioökonomischen Situation der Erziehungsberechtigten/ Eltern abbauen zu können, darf die Teilnahme keine Gebühren von den Schüler*innen oder deren Eltern voraussetzen. Daher wird die BYTE Challenge ausschließlich von ehrenamtlich Engagierten getragen, die sich in der Hochschulgruppe Berlin-Brandenburg der Gesellschaft für Informatik e.V. organisieren. Materielle Unterstützung erhält die BYTE Challenge in Form von Sponsoring und Spenden von Unternehmen, teils finanziell, teils als Sachspenden wie Serverhosting, das fertig gehostete Lernmanagementsystem, Merchandise-Artikel, Räumlichkeiten und Video-Technik.

2.8 Social Media & Öffentlichkeitsarbeit

Über die Öffentlichkeitsarbeit soll die BYTE Challenge potentiell interessierten Schüler*innen bekannt gemacht werden. Dazu erfolgte der Versand von E-Mails an alle auffindbaren Adressen von Oberschulen in Deutschland, innerhalb der GI, sowie an einige Elternverbände.

Um eine möglichst große Bandbreite an Schüler*innen zu erreichen, erfolgt zudem eine breit angelegte Öffentlichkeitsarbeit Social Media-Kanälen, insbesondere auf TikTok und

Instagram. Auf Instagram werden regelmäßig optisch ansprechende Posts mit inhaltlichem Mehrwert veröffentlicht, thematisch liegt der Fokus auf Technik, Berufswahl und Tipps zu digitalen Kompetenzen. Auf TikTok finden sich Informationsvideos zu aktuellen Themen, die sich bei der BYTE Challenge abspielen. Weiterhin achten wir darauf, unseren vor allem jungen Followern einen vielfältigen und reflektierten Blickwinkel auf gesellschaftlich viel diskutierte Themen zu bieten. Über Facebook sollen vorwiegend Eltern und Lehrer*innen angesprochen werden.

2.9 Kooperationspartner & Partnerprojekte

Wir arbeiten eng mit verschiedenen Initiativen zusammen. Wir integrieren in spezifischen Kursen Hinweise auf weiterführende, gemeinnützige und kostenlose Angebote. Wir verweisen beispielsweise an den Bundeswettbewerb Künstliche Intelligenz (https://www.bw-ki.de/). Außerdem bewerben wir InnoTruck Webinare, mit denen wir unser Kursprogramm vor Wettbewerbsbeginn abgeglichen und aneinander angepasst haben (https://www.innotruck.de/initiative-innotruck-startseite).
STARTUPTEENS (https://www.startupteens.de/) haben mit uns die Kursreihe Künstliche Intelligenz entwickelt. Außerdem arbeiten wir mit weiteren Vereinen zum Wissensaustausch zusammen, so z.B. der German Mittelstand und dem Bündnis gegen Cybermobbing (https://www.buendnis-gegen-cybermobbing.de/).

Darüber hinaus haben wir die eduhacktory als Projektwerkstatt an der TU Berlin entwickelt, an der Student*innen aller Studiengänge teilnehmen können. Ziel dieser Projektwerkstatt ist es, die digitale Lehre in Deutschland und Berlin aus einem interdisziplinären Blickwinkel zu reflektieren und sie mit innovativen Ideen voran zu bringen. Dabei werden Studierende und Fachleute aus der Informatik mit Schüler*innen und Lehrer*innen sowie ehrenamtlichen Unterstützer*innen zusammengebracht. In der BYTE Challenge wirken sie bei der Betreuung der Schüler*innen mit und machen sich ein Bild von deren individuellen Erfahrungen. Interviews mit Lehrer*innen sorgen zudem für einen anderen Blickwinkel auf die konkrete Situation an den Schulen.

3 Vorläufige Ergebnisse der BYTE Challenge 2021

Die BYTE Challenge findet zum ersten Mal vom 18. März bis zum 18. Juni 2021 statt. Daher können hier nur Zwischenergebnisse dargestellt werden.

3.1 Anmeldestatistik

Die vorliegende statistische Vorauswertung bezieht sich auf den Stand vom 17.3., den ursprünglichen Anmeldeschluss. Es lagen 763 durch die Eltern bestätigte Anmeldungen

vor. Während der Anmeldung haben wir mit optionalen Feldern die bei den Diagrammen angegebenen Fragen gestellt.

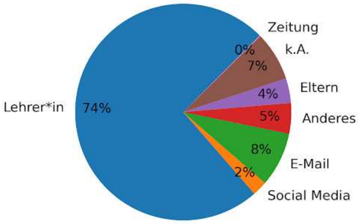


Abb. 1: Wie hast du von der BYTE Challenge erfahren?

Es zeigt sich, dass die Mehrheit der Teilnehmenden über ihre Lehrer*innen von der BYTE Challenge erfahren haben. Es wurde jedoch nicht erhoben, auf welchen Kanälen die Lehrer*innen ihrerseits von der BYTE Challenge erfahren haben. Unsere Videos auf TikTok erhalten meist um 200 Views, einzelne bis zu 650 Views. Auf Instagram haben wir 516 Abonennten und auf Twitter 134 Follower.

Auf die Frage „Hast Du schon einmal an einem Informatikwettbewerb teilgenommen?“ antworteten 64% „Nein“, 29% „Ja“ und „8%“ machten keine Angabe.
Auf „Hast Du bereits Scratch-Erfahrungen?“ antworteten 49% „Nein“, 43% „Ja“ und „7%“ machten keine Angabe.
Das Ziel, überwiegend Schüler*innen ohne Vorerfahrungen anzusprechen, wurde demnach erreicht.

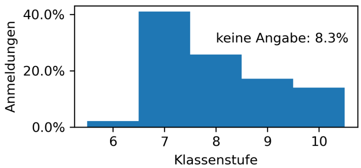


Abb. 2: In welcher Klassenstufe bist Du?

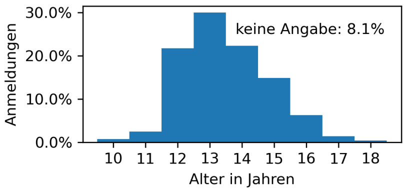


Abb. 3: Wie alt bist Du?

Abb. 2 und 3 zeigen, die Zielgruppe der Klassenstufe 7-10 ist mit besonderer Häufung in Klasse 7 und 8 getroffen. Unter den Teilnehmenden befinden sich Schüler*innen aller Schulformen von Klasse 6 bis 10, deutschlandweit.

Abb. 4 und 5 zeigen, dass die Teilnehmenden aus ganz Deutschland kommen, jedoch eine große Spreizung zwischen den Bundesländern besteht.

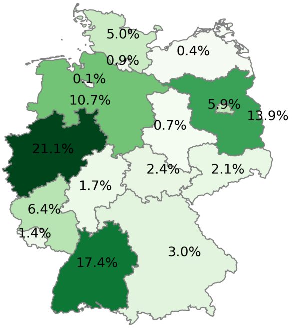


Abb. 4: Anteil der Teilnehmenden nach Bundesland

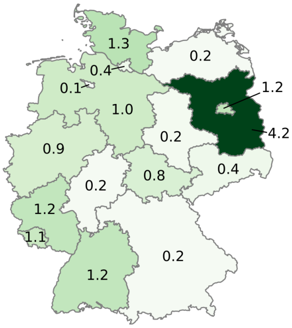


Abb. 5: Teilnehmende pro 100.000 Einwohner

3.2 Rückmeldungen von Lehrkräften

Wir haben den 80 bei uns registrierten Lehrer*innen bislang zwei digitale Zusammenkünfte angeboten, dazu kommen individuelle Videokonferenzen und Schriftwechsel mit einzelnen unterstützenden Lehrkräften. Dabei waren die Rückmeldungen durchaus heterogen, folgendes wurde jedoch mehrfach geäußert:

Lehrkräfte an Gymnasien finden die Inhalte der BYTE Challenge teilweise zu wenig anspruchsvoll.
Lehrkräfte der anderen Oberschul-Formen äußerten jedoch, dass das Niveau passend oder teilweise zu anspruchsvoll sei. Zudem wurde beschrieben, dass an Nicht-Gymnasien kaum Digital-Kenntnisse vorhanden seien und teilweise auch im Lehrkörper Defizite bestünden. Schwierigkeiten wurden vorwiegend bei E-Mail, Datei- und Accountverwaltung genannt. Zudem wurde die mangelhafte Infrastruktur und Ausrüstung der Schulen insbesondere für den Distanzunterricht beklagt.

Bei allen Schulformen ist die Gruppenanmeldung von Klassen/ Kursen durch Lehrer*innen sehr gefragt, oft verbunden mit dem Wunsch, eine Integration in den Unterricht vorzunehmen.

Insgesamt fällt das Feedback überwiegend positiv aus.

3.3 Feedback der Teilnehmenden

Am Ende der Kurse im LMS haben wir jeweils Fragebögen für (freiwilliges) Feedback integriert. Diese beinhalteten bei den Seminaren und den C+ Kursen Multiple-Choice-Fragen zum Gesamteindruck des Kurses, der Schwierigkeit, dem Interesse am Thema und Verbesserungsvorschlägen mit den Antwortmöglichkeiten „Super“, „Geht so“, „Nicht

so gut“ bzw. „Es gibt noch Dinge zu verbessern“. Bei den Scratch-Kursen haben wir im gleichen Schema Fragen zu den gestellten Aufgaben, Beispielen und der (inhaltlichen) Qualität der Videos oder Texte gestellt. Durchschnittlich gaben 65% der Kursteilnehmenden Feedback.

Vor allem bei den ersten Scratch-Kursen und einigen C+ Kursen wurde der Inhalt von einigen Teilnehmenden als „zu leicht“ eingestuft. In den Scratch-Kursen aus Phase 2 steigt der Anteil derer, die den Kurs als zu schwer einstufen (24% in Phase 2 Woche 1, 28% in Phase 2 Woche 3 und 13% in Phase 2 Woche 4 im Vergleich zu 13% in Phase 1). Dies ist vermutlich auf die unterschiedliche Vor- sowie Programmiererfahrung unter den Teilnehmenden zurückzuführen.
In Phase 2 fand ein Wechsel von Video-Tutorials mit alternativem Transkript hin zu reinen Text-Bild-Anleitungen. Von vielen Teilnehmenden wurde im Feedback angesprochen, dass sie Videos vermisst haben. Auch einige ungenaue Aufgabenformulierungen wurden angesprochen, diese wurden während des laufenden Wettbewerbs bearbeitet.

Das Feedback zu den Seminaren ist ebenfalls überwiegend positiv. Den Gesamteindruck gaben bei dem „Ausbildung oder Studium“-Seminar 60% als „super“ an, bei dem Seminar „Freiwilligendienst“ waren es 90%. Die meisten Seminare wurden vom Schwierigkeitsgrad als „genau richtig“ eingestuft (87% im Seminar „Ausbildung und Studium, 95% im Seminar „Freiwilligendienst“). Mit 8% hat das Seminar „Digitale Souveränität“ am häufigsten die Einstufung „zu schwer“ erhalten, während 15% dieses Seminar zu leicht fanden. Zu fast allen Seminaren gab es das Feedback, dass diese zu lang seien und es teilweise zu technischen Schwierigkeiten kam.

In den C+ Kursen fällt das Feedback ähnlich aus. So stuften 3% den Kurs „Computerbestandteile“ als zu schwer ein, 79% als genau richtig und 19% als zu leicht. In späteren und komplexeren Kursen wie „Künstliche Intelligenz 1“ fanden 8% den Kurs zu schwer, 84% genau richtig und 8% zu leicht. Nur sehr wenige Teilnehmende stuften die Themen der C+ Kurse als „langweilig“ ein (maximal 2% beim Kurs „Künstliche Intelligenz 2“). Auch hier wünschten sich die Teilnehmenden in einigen Kursen mehr Videos statt Texte und bemängelten im Kurs „Internet“ die zu leichten Quizfragen. Insgesamt fiel das Feedback überwiegend positiv aus.

3.4 Kritische Betrachtung

Die Entwicklung der Fragen an die Teilnehmenden war im ersten Durchgang 2021 sekundär. Zudem fehlen bei rund 200 durch Lehrkräfte nachgemeldeten Teilnehmenden die Antworten auf die Umfrage bei der Anmeldung vollständig. Die Statistiken können deswegen nur Tendenzen aufzeigen. Für ein umfassendes Qualitätsmanagement und die weitere wissenschaftliche Begleitung der BYTE Challenge ist daher auch die Weiterentwicklung der Fragen erforderlich.

4 Ausblick

Die BYTE Challenge soll auch 2022 wieder stattfinden. Die Erfahrungen aus diesem Jahr sollen folgendermaßen einfließen:

Von den vier Kategorien nach Seegerer; Romeike [SR18] vermittelt die BYTE Challenge bislang nur wenig aus der Kategorie (G2) Fluency. Das deckt sich mit den Rückmeldungen von Lehrkräften an Nicht-Gymnasien, die einen Mangel an Fähigkeiten der Schüler*innen im Umgang mit grundlegenden PC-Anwendungen beklagen. Daher sollen Inhalte auf dem Einstiegsniveau der Computernutzung integriert werden.

Dem Wunsch nach Klassenanmeldungen im Rahmen des Unterrichts wollen wir nachkommen. Einerseits durch eine Modularisierung des Kursangebotes, sodass die Lehrer*innen nur die für ihre Klasse passenden Kurse heraussuchen und bearbeiten können. Andererseits indem wir das Datenschutzkonzept so erweitern, dass BYTE als Auftragsverarbeiter der Schulen auftreten kann, dann muss nicht mehr separat das Einverständnis der Erziehungsberechtigten eingeholt werden. Die Teilnahmemöglichkeit für einzelne Schüler*innen soll dennoch erhalten bleiben, um weiterhin auf den Abbau von Benachteiligungen hinzuwirken.

Darüber hinaus soll eine Erweiterung der Themen über die Informatik hinaus auf den MINT-Bereich erprobt werden, wofür wir ebenfalls Potential vermuten [We21].

Literatur

[19a] Hochschul-Bildungs-Report 2020, Für morgen befähigen, Jahresbericht 2019, Report, Essen: Stifterverband für die Deutsche Wissenschaft e.V., 2019, URL: https://www.hochschulbildungsreport2020.de/download/file/fid/163.

[19b] MINT – Berufe, Bericht, Nürnberg: Bundesagentur für Arbeit, Aug. 2019, URL: https://statistik.arbeitsagentur.de/DE/Statischer-Content/Statistiken/Themen-im-Fokus/Berufe/Generische-Publikationen/Broschuere-MINT.pdf?__blob=publicationFile.

[20] HANDLUNGSRAHMEN FÜR DAS SCHULJAHR 2020/21, Anlage zum Brief an die Schulleitungen vom 4. August 2020, Berlin-Mitte: Senatsverwaltung für Bildung, Jugend und Familie, 4. Aug. 2020.

[BCW21] Beames, J.; Christensen, H.; Werner-Seidler, A.: School teachers: the forgotten frontline workers of Covid-19. Australasian Psychiatry/, Apr. 2021, URL: https://doi.org/10.1177/10398562211006145.

[Ko16] Kossak, H.-C.: Lernen leicht gemacht Gut vorbereitet und ohne Prüfungsangst zum Erfolg. Heidelberg, Carl Auer, 2016, ISBN: 978-3-8497-0125-3.

[Mo21a] Mohi, A.; Gniesmer, S.; Ranjbar, M.; Kakkassery, V.; Grisanti, S.; Neppert, B.; Kurz, M.; Lüke, J.; Lüke, M.; Müller, M.; Lommatzsch, C.; Grisanti, S.: Digitale Lehre 2020: Studenten schätzen die Aufmerksamkeit während einer Onlinevorlesung gleichwertig zu einer Präsenzvorlesung ein. Der Ophthalmologe/, Springer Link, März 2021, URL: https://doi.org/10.1007/s00347-021-01344-1.

[Mo21b] Molwitz, I.; Othman, A.; Brendlin, A.; Afat, S.; Barkhausen, J.; Reinartz, S. D.: Digitale Lehre mit, durch und nach COVID-19. Der Radiologe 61/1, Springer Link, S. 64–66, Jan. 2021, URL: https://doi.org/10.1007/s00117-020-00794-z.

[Mü17] Müller, D.: Berufswahl Informatiklehrkraft. In: Informatische Bildung zum Verstehen und Gestalten der digitalen Welt. Gesellschaft für Informatik, Bonn, S. 127–136, 2017.

[SH21] Schwarz, R.; Hellmig, L.: Informatikunterricht in Deutschland – eine Übersicht. Informatik Spektrum/, Springer Link, 6. Apr. 2021, URL: https://doi.org/10.1007/s00287-021-01349-9.

[SLR19] Seegerer, S.; Lindner, A.; Romeike, R.: AI Unplugged – Wir ziehen Künstlicher Intelligenz den Stecker. In (Pasternak, A., Hrsg.): Informatik für alle. Gesellschaft für Informatik, Bonn, S. 325–334, 2019.

[SR18] Seegerer, S.; Romeike, R.: Was jeder über Informatik lernen sollte - Eine Analyse von Hochschulkursen für Studierende anderer Fachrichtungen. In (Bergner, N.; Röpke, R.; Schroeder, U.; Krömker, D., Hrsg.): Hochschuldidaktik der Informatik - HDI 2018 - 8. Fachtagung des GI-Fachbereichs Informatik und Ausbildung/Didaktik der Informatik, Frankfurt, Germany, September 12-13, 2018. Universitätsverlag Potsdam, S. 13–28, 2018, URL: https://dblp.org/rec/conf/hdi/SeegererR18.bib.

[St19] Stanat, P.; Schipolowski, S.; Mahler, N.; Weirich, S.; Henschel, S.: IQB-Bildungstrend 2018 Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I im zweiten Ländervergleich. Waxmann Verlag GmbH, 2019.

[We21] Weinhandl, R.; Lavicza, Z.; Houghton, T.; Hohenwarter, M.: A look over students' shoulders when learning mathematics in home-schooling. International Journal of Mathematical Education/, Taylor and Francis, Apr. 2021, URL: https%E2%80%8B://doi.org/10.1080/0020739X.2021.1912423.

[Wi21] Wilkens, A.: Informatik wird an Schleswig-Holsteins Schulen Pflichtfach. Heise Online/, Heise Medien, Mai 2021, URL: https://www.heise.de/news/Informatik-wird-an-Schleswig-Holsteins-Schulen-Pflichtfach-6037135.html.

Distance Decay Effect and Spatial Interaction during the COVID-19 pandemic

Nicolas Wolz [1], Manning Xu [2], Tiantian Wang [3]

Abstract:

In computational communication science, social network data can be used to analyze trends in the communication behavior of people. For this purpose, a data set containing English Tweets was provided by the University of Technology Ilmenau, which was collected during the beginning of the COVID-19 pandemic by the Database Systems Research Group at University Heidelberg. The goal was to find hidden patterns within the data to show if and how the pandemic influenced our communication. This paper looks at the Distance Decay Effect, which says that near things are more related to each other than distant things, and therefore communication should get more sparse the greater the distance between users. Modeling the data with a Gravity Model shows that this relationship is true for the data provided, therefore reproducing earlier research on this topic. We were not successful in finding any clear trend showing that the strength of the Distance Decay Effect changed over the course of the first weeks of the pandemic.

Keywords: Distance Decay Effect; Gravity Model; COVID-19; Twitter

1 Introduction

An increasing demand for understanding the spatial connection characteristics including spatial and temporal perspectives of information diffusion on social media is observed in various practical scenarios [Su17]. For instance, the government needs to contain the diffusion of rumors on social media, to identify critical geographic areas and key time windows where rumors originate from, become viral and fade out, which can further help them disseminate the truth to users in those critical areas and at optimal time spots [Ca12] [SOM10] [Su17]. Some research shows that investigating the relationship between cyberspace and real space, using big data and social media data, can help better understand human activities [HTC18] [Ju15].

Twitter has become a feasible social media platform to explore global human communication patterns as well as city-scale human communication. It has been widely used as a tool

[1] TU Ilmenau, Student / Databases and Information Systems Group, Ehrenbergstraße 29, 98693 Ilmenau, Germany nicolas.wolz@tu-ilmenau.de

[2] TU Ilmenau, Student / Department of Economic Sciences and Media, Computational Communication Research Group, Ehrenbergstraße 29, 98693 Ilmenau, Germany manning.xu@tu-ilmenau.de

[3] TU Ilmenau, Student / Department of Economic Sciences and Media, Computational Communication Research Group, Ehrenbergstraße 29, 98693 Ilmenau, Germany tiantian.wang@tu-ilmenau.de

to understand group dynamics from information dissemination on online social networks [LG10]. During the situation of the COVID-19 pandemic, more importantly, Twitter provides the opportunity for researchers to explore the role social media plays, especially in a global health crisis [CE10]. Thus, Twitter has become one of the centers of the social infrastructure and is a technology that allows us to stay connected, even during the crisis [CLF20].

COVID-19 has been characterized as a pandemic on March 11. At that point, the virus has affected 114 countries, which led to unnecessary suffering and death [WH]. Some measures like social distancing, quarantines, travel bans, and business closures are changing the structure of societies worldwide. Due to being forced out of public spaces, a lot of people communicate about the pandemic in social media, like Twitter [CLF20]. The numbers of the conversation around COVID-19 have continued to expand [Ab20]. However, the influence of physical distance in digital interaction is still a topic of research [HTC18]. Some researchers like Han, Tsou, and Clarke [HTC15] find that the individuals who live in nearby places interact more with each other than the people living more distant from each other. This phenomenon is called Distance Decay Effect. To describe this effect, researchers use the gravity model [Yu17], which describes the expected rate of interaction given a distance between places, based on the data the model was trained on.

This paper aims to find the Distance Decay Effect during the breakout time of COVID-19 pandemic, using the Gravity Model and Twitter data from March 2020, extracted by the Database Systems Research Group at University Heidelberg [4]. Moreover, we want to investigate whether the COVID-19 pandemic changed this communication pattern.

2 Theoretical Considerations & Research Questions

2.1 Theoretical background

Spatial connection was implied in Tobler's first law of geography (Tobler, 1970), the concept is "near things are more related than distant things", and covers a broader range of connection than "interaction" [Yu17]. Extant research gave spatial interaction different definitions. MacLachlan [Qu] defined spatial interaction as a dynamic flow process that articulates one location with another. It is a general concept that may refer to the movement of human beings such as intra-urban commuters or intercontinental migrants but may also refer to traffic in goods such as raw materials or to flows of intangibles such as information. Other researchers define spatial interaction in a broad context as actual or potential flow among places, with any type of connection among places [HTC18].

During the COVID-19 pandemic, people produced social media content on Twitter. The place of origin and the coordinates of users' locations while posting are recorded and

[4] https://dbs.ifi.uni-heidelberg.de

are partly available for research. It is also recorded if a tweet is a reply to another tweet. Therefore, in the context of this paper, a spatial interaction is defined as a tweet replying to another tweet. The distance of this interaction is calculated using the coordinates (latitude, longitude) associated with the respective tweets. The whole study revolves around the question how many people from inside a so-called central entity interact with the outside world and vice versa. This central entity is defined as the set of tweets posted within a given radius of a central coordinate. We chose the city centers of several major US cities (New York, Los Angeles, Seattle, Chicago, Houston) as central coordinates and 100 km radius to include the whole metropolitan area of those cities. Tweets from outside the central entity are grouped in "distance from this central city"categories. By looking inside the US on a city scale, we have a large language-homogenous area for the distance decay effect to occur with less language bias than observed on a global scale.

In order to analyze the data, the commonly used Gravity Model is leveraged. It is used in several research with regard to spatial interaction due to its effectiveness in predicting the degree of interaction, the simplicity of its equation, and its ability to deal with flows in both directions [HFG12]. Distance decay is inherent in spatial gravity models, but the slope and range of the distance decay vary depending on the type of human interaction [HTC18]. The model can be applied as a qualitative conceptual tool or it may be operationalized in different ways using quantitative data [Qu]. We shall employ a simple bivariate regression model to estimate model parameters to quantify the distance decay effect on interaction.

2.2 Study design and research questions

In general, this paper investigates how the COVID-19 Pandemic affected people's communication behavior. To answer this question, the study analyzes the provided Twitter data set in an observational approach. Two research interests were developed to structure the research.

Research interest 1: If the distance from the central entity is shorter, the amount of spatial interaction is higher.

Based on the research of distance decay and spatial interaction by Han, Tsou, and Clarke [HTC15], our research interest is phrased as a hypothesis. This part of the research is conducted confirmatory.

Research interest 2: We will continue to approach the data explorational to see if there has been a change in the spatial interaction behavior over time. Temporally, this research focuses on data collected in March. During this month, both Europe and the US have seen a strong increase in COVID-19 cases and some of the most important events of the pandemic happened in this time period in respective calendar weeks (CW):

It could be expected that the strength of the Distance Decay Effect changed during the course of the COVID-19 pandemic, especially with lockdowns enforced all over the word. This research aims to explore if any general trend can be revealed.

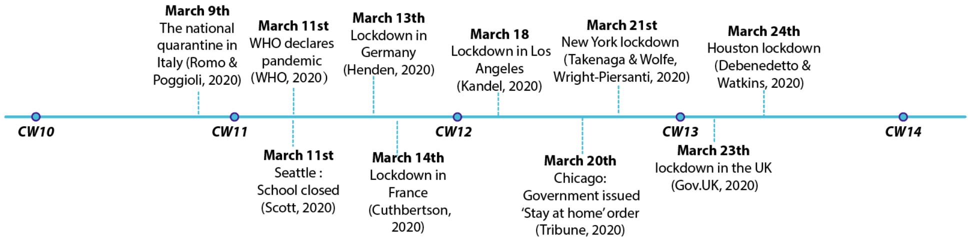


Fig. 1: Timeline of the lockdown time in European countries and five US cities.

3 Methodology

This chapter focuses on the main ideas and the model used in order to be able to understand the conclusions drawn from the results. First, the definition of terms, denotations and observed variables will be introduced. Secondly, the process of finding a Distance Decay Effect in the data is split in two distinct steps: data filtering and information extraction; building, evaluating and visualizing of Distance Decay Effect and Gravity Model.

3.1 Definition of terms, denotations and observed variables

As mentioned, a spatial interaction is defined as a tweet replying to another tweet. In order to calculate the number of those interactions, the attributes "ID" and "in_Reply_to_status_id" of the given Twitter data set are used. The number of spatial interactions is denoted as I_ij (I:big letter i).

The central entity is the middle point of our model. Like already mentioned, it represents the metropolitan area of a city. To find out if a Tweet belongs to the central entity, the coordinates found in the attribute "bounding_box" are used. Model variables belonging to the central entity are denoted with the index i. On the other hand, the outside world will be denoted with index j. Tweets get categorized based on the distance from the central entity.

The distance between the central entity i and a place in the outside world j is calculated using the coordinates found in "bounding_box". For this, the haversine formula [Ha] is used to calculate the distance between two points on a sphere with given longitudes and latitudes. It is denoted as D_ij. The conceptual size of a group is the total number of tweets in it. This measurement is used to normalize the data, since it is obvious that the pure size of some places, like Los Angeles, will increase the number of interactions. In order to remove this effect from the results, normalization is needed. The letter P is used as a denotation for the conceptual size.

3.2 Data Filtering and Information Extraction

The first step consists of reading the big data set, which contains about 100 Million Tweets, including attributes mentioned in the previous subsection, among others. It gets condensed down to a much smaller data sample with a higher information density, containing a small number of key observations. For that, the source data is iterated two times. This results in a small .json file containing all the relevant information needed for the second step, data analysis (see next subsection).

The first iteration is used to extract tweets originating from the central entity within a given time frame. The items of columns "ID" (set A) and "in_Reply_to_status_id" (set B) are stored in sets in order to work with them during the second iteration.

During the second iteration, each tweet is categorized/grouped, based on its distance from the central entity. In order to find spatial interactions, according to the definition corresponding to replies, the items of columns "ID" (set C) and "in_Reply_to_status_id" (set D) of each category are stored in sets as well. The number of interactions is calculated by adding $|A \cap D|$ and $|B \cap C|$. Additionally, the conceptual sizes of each group gets counted as well. After some data transformation, the output of this step looks like shown in figure 2 (note that just the head of the data frame is shown here, the complete data frame contains more entries).

	conceptualSize	spatialInteractions	distance
100	757716	91	100
200	826967	64	200
300	1354098	131	300
400	503594	30	400
500	388781	13	500

Fig. 2: Structure of the intermediate results

3.3 Data Analysis

3.3.1 Distance Decay Effect and Gravity Model

The second step consists of building, evaluating and visualizing the gravity model. As described in Yuan, Liu and Wei's [Yu17] research, this is a simple and effective method to model the Distance Decay Effect based on observations.

The observations, which are the output of the first step (see figure 2) and thus the input for this step, consist of the number of spatial interactions I_ij between our central entity i and

$$I_{ij} = K \frac{P_i P_j}{D_{ij}^{\beta}} \quad (1)$$

Equation 1: Gravity model formula

the outside world places j, the respective conceptual sizes P_i and P_j of those places, and the distances D_ij between them.

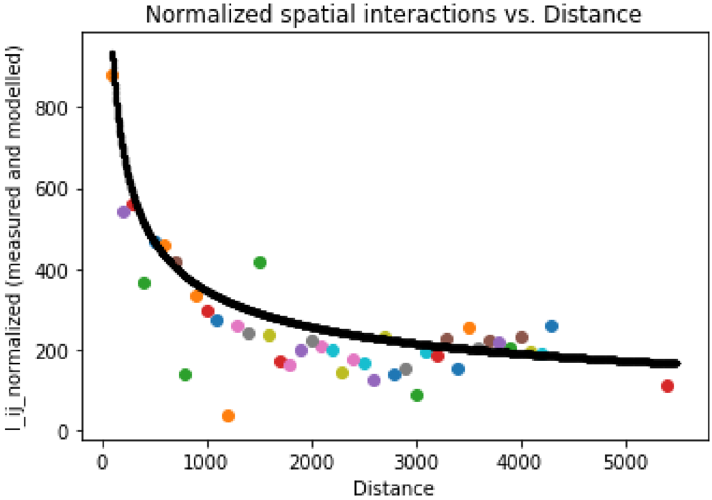


Fig. 3: Gravity model curve and observed data points

As seen in figure 3 and figure 4, we can plot the gravity model curve and a correlation graph between modelled and observed data, if we find values for the constant K, which is a scaling factor, and for β, which is the distance friction coefficient. In order to determine those two missing values, we do a so-called model fitting.

3.3.2 Model fitting and evaluation

To find the distance friction coefficient β, we iterate through a lot of potential values (e.g. 0 to 3 in very small steps) and evaluate each resulting model (see figure 5). The β with the highest evaluation score, called R_squared, will be chosen for our final model. R_squared is calculated by squaring the Pearson Correlation between the model and the observation (see figure 4). The β with the highest score is called "best matching β".

Since K is a constant scaling factor and does not influence correlation and therefore the quality of our model, a rough estimation of a realistic value is sufficient. It would be possible

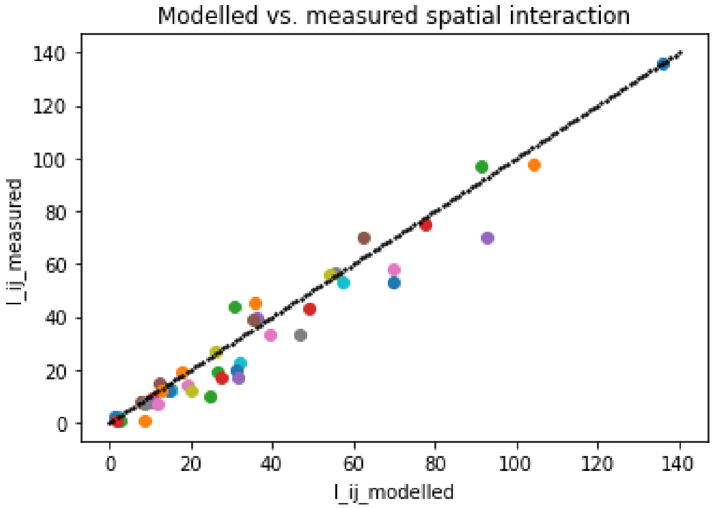


Fig. 4: Correlation between measured and modelled spatial interaction

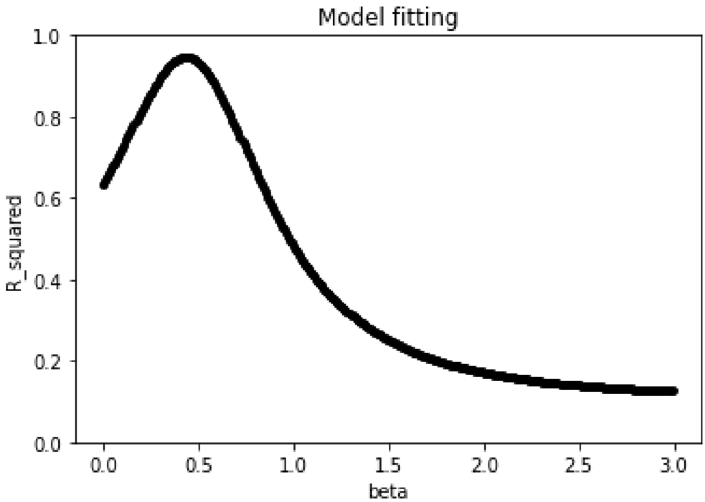


Fig. 5: Iteration over many values for β and evaluation of each resulting model

to do another round of model fitting, this time with fixed β and variable K, using e.g. least square optimization, to find the perfect value. For the purpose of this research this was not necessary, since only the best matching β and its respective R_squared value are relevant. A realistic value for K is just needed for plots like figure 3 and figure 4. For those, it is calculated using the values of just one data point in order to have a rough estimation, which is not necessarily the best fitting value.

4 Results

As mentioned before, the central entities represent the metropolitan areas around a central coordinate in the middle of the major US cities of New York, Los Angeles, Seattle, Chicago and Houston. For each city and time frame, a separate model is being built.

Regarding research interest 1, analysis was conducted using data from the whole month of March. The model fitting process produces the values seen in figure 6. The Gravity Model produces the expected curves and correlation diagrams as seen in figure 3 and figure 4.

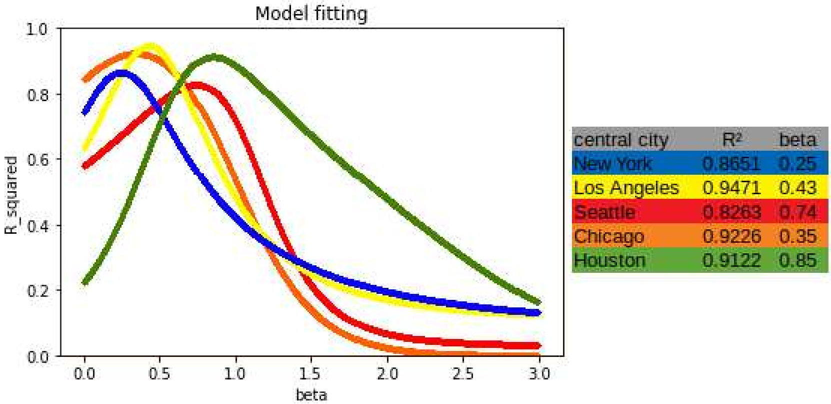


Fig. 6: Model fitting for five major cities in the US

Therefore, a distance decay effect inside the US can be clearly observed. If the hypothesis of research interest 1 was false, the results would show very low or unrealistically high values for the correlation R_squared or unrealistic values for β, e.g. 0.

In order to explore research interest 2, the data was analyzed for specific weeks (CW10 to CW13). As it can be seen in figure 7 (right), R_squared improved significantly from CW10 to CW11 for every city, which is in line with the growth of the number of tweets of each city present in the data set for those weeks. However, this does not mean that a high number of tweets automatically results in a higher R_squared score since this correlation does not continue in the following weeks CW12 and CW13.

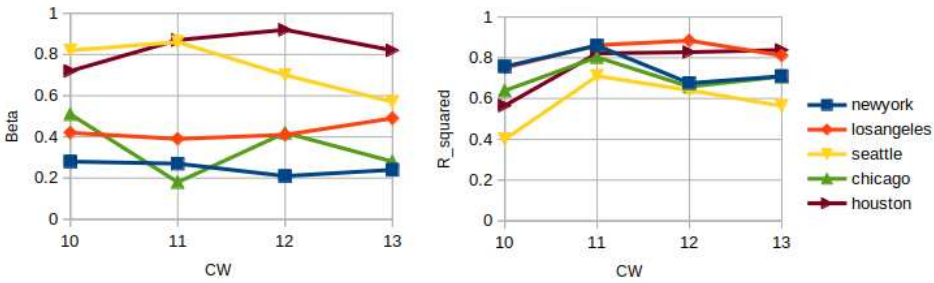


Fig. 7: Development of β and R_squared for CW10 to CW13

Figure 7 (left) shows the development of β. New York and Los Angeles only show a slight fluctuation within this time frame. Besides that, Chicago reached a low β value in CW11, shortly before the lockdown time (March 20th). However, the values for Chicago generally show a great fluctuation. Furthermore, Seattle indicates a downwards trend from CW11 to CW13, while Houston shows an upward trend from CW10 to CW12.

Overall, there is no uniform and clear trend to be seen for all cities for the value of β, given our data set and methodology. This could mean that the strength of the distance decay effect was not affected by the events of the pandemic during this time frame in the US.

5 Discussion

Initially, this study tried to analyze the data on a global scale with whole countries as entities in mind. Unfortunately, this approach was not successful because of limitations which will be discussed in the next subsection. However, with the adapted US city approach, the findings are in line with Han, Tsou, Ming-Hsiang, Clarke's (2018) [HTC18] research. As is notable in figure 8, there is a high correlation (0.98) between the results of this research and theirs. However, the definition of spatial interaction and therefore the absolute numbers differ. In Han, Tsou, Ming-Hsiang, Clarke's (2018) [HTC18] research, "following" was one way to measure spatial interaction. In this research, only "replying" was considered spatial interaction.

As for research interest 2, the results do not meet the previous expectations, as no correlation between Distance Decay Effect and the events of COVID-19 pandemic was found. As mentioned, this might indicate that there is no connection and the communication patterns were not influenced by the pandemic. But it might also be the case that the limitations, which will be discussed in the next subsection, are the reason for absence of evidence. Exploring the development of spatial interaction over a longer period of time, using a bigger and possibly better data set and defining more means of spatial interaction could help to reveal those trends and correlations, if they exist.

Fig. 8: Correlation (0.98) between our findings and Han, Tsou, Ming-Hsiang, Clarke (2018) [HTC18] [Figure 7, "Follow"]

5.1 Limitations

Regarding limitations, one of the main problems of this study was the strong language bias present in the data set. Since the data set contains only tweets labelled as English language, communication in and between non-English speaking countries is very sparse. The countries Great-Britain, United States, Ireland and Canada dominate the data set and therefore the model fitting process. Since model fitting is done using linear correlation between modelled and measured interactions, the data points with the biggest number of interactions strongly dictate the results, as seen in figure 9. Based on this finding, we conclude that it is not possible to research the Distance Decay effect on a global-scale, using our methodology and the given English-biased data set. In order to conduct more research on this topic on a global scale, a non-biased data set has to be collected.

Another limitation of this research is the definition of spatial interaction. Since only replies to a tweet are considered, other means of interaction are ignored. Besides the obvious liking and retweeting, also physical movement or mentioning of places could be considered spatial interaction [HTC18].

For some models, the R_square values are quite low (see figure 7, CW10). In some cases, this correlates with a low number of tweets and therefore a low number of spatial interactions, which makes the performance of the gravity model more susceptible to noise and randomness.

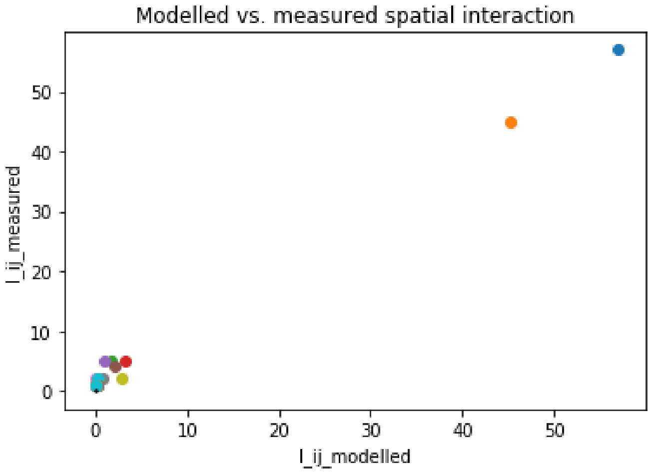


Fig. 9: Correlation graph for central entity Germany for the time from CW10 to CW13

5.2 Outlook

Further research on this topic could try to tackle mentioned problems in order to verify the findings of this research or find new trends and effects. The general methodology can be used on other data sets, possibly containing also non-English Tweets. Complementing the methodology with further definitions of spatial interaction and analyzing a longer time scope might also help to produce new insights.

6 Acknowledgements

We would like to thank both the University of Technology Ilmenau and the Heidelberg University for making this paper possible. The presented results have been achieved during the research seminar „Data Science Social Media“, supervised by Aliya Andrich, Prof. Emese Domahidi, Prof. Kai-Uwe Sattler and Dr. Nadine Steinmetz. The Twitter data set has been retrieved with the help of the Database Systems Research Group at Heidelberg University.

Bibliography

[Ab20] Abbas, Ansar; Eliyana, Anis; Ekowati, Dian; Saud, Muhammad; Raza, Ali; Wardani, Ratna: Data set on coping strategies in the digital age: The role of psychological well-being and social capital among university students in Java Timor, Surabaya, Indonesia. Data in Brief, 30:105583, 2020.

[Ca12] Cao, Nan; Lin, Yu-Ru; Sun, Xiaohua; Lazer, David; Liu, Shixia; Qu, Huamin: Whisper: Tracing the Spatiotemporal Process of Information Diffusion in Real Time. IEEE Transactions on Visualization and Computer Graphics, 18(12):2649–2658, 2012.

[CE10] Chew, Cynthia; Eysenbach, Gunther: Pandemics in the Age of Twitter: Content Analysis of Tweets During the 2009 H1N1 Outbreak. PloS one, 5:e14118, 11 2010.

[CLF20] Chen, Emily; Lerman, Kristina; Ferrara, Emilio: Tracking Social Media Discourse About the COVID-19 Pandemic: Development of a Public Coronavirus Twitter Data Set. JMIR public health and surveillance, 6(2), 2020.

[Ha] Haversine formula. https://pypi.org/project/haversine/. Accessed: 2020-07-21.

[HFG12] Hardy, Darren; Frew, James; Goodchild, Michael F.: Volunteered geographic information production as a spatial process. International Journal of Geographical Information Science, 26(7):1191–1212, 2012.

[HTC15] Han, Su Yeon; Tsou, Ming-Hsiang; Clarke, Keith C.: Do Global Cities Enable Global Views? Using Twitter to Quantify the Level of Geographical Awareness of U.S. Cities. PLOS ONE, 10(7):1–23, 07 2015.

[HTC18] Han, Su Yeon; Tsou, Ming-Hsiang; Clarke, Keith C.: Revisiting the death of geography in the era of Big Data: the friction of distance in cyberspace and real space. International Journal of Digital Earth, 11(5):451–469, 2018.

[Ju15] Jurdak, Raja; Zhao, Kun; Liu, Jiajun; Abou Jaoude, Maurice; Cameron, Mark; Newth, David: Understanding Human Mobility from Twitter. PLoS ONE, 10, 07 2015.

[LG10] Lerman, Kristina; Ghosh, Rumi: , Information Contagion: an Empirical Study of the Spread of News on Digg and Twitter Social Networks, 2010.

[Qu] Quantitative Models for Geographical Analysis. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.471.6867&rep=rep1&type=pdf. Accessed: 2020-06-11.

[SOM10] Sakaki, Takeshi; Okazaki, Makoto; Matsuo, Yutaka: Earthquake Shakes Twitter Users: Real-Time Event Detection by Social Sensors. In: Proceedings of the 19th International Conference on World Wide Web. WWW '10, Association for Computing Machinery, New York, NY, USA, S. 851–860, 2010.

[Su17] Sun, Guodao; Tang, Tan; Peng, Tai-Quan; Liang, Ronghua; Wu, Yingcai: SocialWave: Visual Analysis of Spatio-Temporal Diffusion of Information on Social Media. ACM Trans. Intell. Syst. Technol., 9(2), Oktober 2017.

[WH] WHO Emergencies Press Conference on coronavirus disease outbreak-11 March 2020. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/media-resources/press-briefing. Accessed: 2020-03-11.

[Yu17] Yuan, Yihong: Exploring the Spatial Decay Effect in Mass Media and Location-Based Social Media: A Case Study of China. In (Griffith, Daniel A.; Chun, Yongwan; Dean, Denis J., Hrsg.): Advances in Geocomputation. Springer International Publishing, Cham, S. 133–142, 2017.

Autor:innenverzeichnis

GI-Edition Lecture Notes in Informatics

P-1 Gregor Engels, Andreas Oberweis, Albert Zündorf (Hrsg.): Modellierung 2001.

P-2 Mikhail Godlevsky, Heinrich C. Mayr (Hrsg.): Information Systems Technology and its Applications, ISTA'2001.

P-3 Ana M. Moreno, Reind P. van de Riet (Hrsg.): Applications of Natural Lan-guage to Information Systems, NLDB'2001.

P-4 H. Wörn, J. Mühling, C. Vahl, H.-P. Meinzer (Hrsg.): Rechner- und sensorgestützte Chirurgie; Workshop des SFB 414.

P-5 Andy Schürr (Hg.): OMER – Object-Oriented Modeling of Embedded Real-Time Systems.

P-6 Hans-Jürgen Appelrath, Rolf Beyer, Uwe Marquardt, Heinrich C. Mayr, Claudia Steinberger (Hrsg.): Unternehmen Hochschule, UH'2001.

P-7 Andy Evans, Robert France, Ana Moreira, Bernhard Rumpe (Hrsg.): Practical UML-Based Rigorous Development Methods – Countering or Integrating the extremists, pUML'2001.

P-8 Reinhard Keil-Slawik, Johannes Magenheim (Hrsg.): Informatikunterricht und Medienbildung, INFOS'2001.

P-9 Jan von Knop, Wilhelm Haverkamp (Hrsg.): Innovative Anwendungen in Kommunikationsnetzen, 15. DFN Arbeitstagung.

P-10 Mirjam Minor, Steffen Staab (Hrsg.): 1st German Workshop on Experience Management: Sharing Experiences about the Sharing Experience.

P-11 Michael Weber, Frank Kargl (Hrsg.): Mobile Ad-Hoc Netzwerke, WMAN 2002.

P-12 Martin Glinz, Günther Müller-Luschnat (Hrsg.): Modellierung 2002.

P-13 Jan von Knop, Peter Schirmbacher and Viljan Mahni_ (Hrsg.): The Changing Universities – The Role of Technology.

P-14 Robert Tolksdorf, Rainer Eckstein (Hrsg.): XML-Technologien für das Semantic Web – XSW 2002.

P-15 Hans-Bernd Bludau, Andreas Koop (Hrsg.): Mobile Computing in Medicine.

P-16 J. Felix Hampe, Gerhard Schwabe (Hrsg.): Mobile and Collaborative Business 2002.

P-17 Jan von Knop, Wilhelm Haverkamp (Hrsg.): Zukunft der Netze –Die Verletzbarkeit meistern, 16. DFN Arbeitstagung.

P-18 Elmar J. Sinz, Markus Plaha (Hrsg.): Modellierung betrieblicher Informationssysteme – MobIS 2002.

P-19 Sigrid Schubert, Bernd Reusch, Norbert Jesse (Hrsg.): Informatik bewegt – Informatik 2002 – 32. Jahrestagung der Gesellschaft für Informatik e.V. (GI) 30.Sept.-3. Okt. 2002 in Dortmund.

P-20 Sigrid Schubert, Bernd Reusch, Norbert Jesse (Hrsg.): Informatik bewegt – Informatik 2002 – 32. Jahrestagung der Gesellschaft für Informatik e.V. (GI) 30.Sept.-3. Okt. 2002 in Dortmund (Ergänzungs-band).

P-21 Jörg Desel, Mathias Weske (Hrsg.): Promise 2002: Prozessorientierte Methoden und Werkzeuge für die Entwicklung von Informationssystemen.

P-22 Sigrid Schubert, Johannes Magenheim, Peter Hubwieser, Torsten Brinda (Hrsg.): Forschungsbeiträge zur "Didaktik der Informatik" – Theorie, Praxis, Evaluation.

P-23 Thorsten Spitta, Jens Borchers, Harry M. Sneed (Hrsg.): Software Management 2002 – Fortschritt durch Beständigkeit

P-24 Rainer Eckstein, Robert Tolksdorf (Hrsg.): XMIDX 2003 – XML-Technologien für Middleware – Middleware für XML-Anwendungen

P-25 Key Pousttchi, Klaus Turowski (Hrsg.): Mobile Commerce – Anwendungen und Perspektiven – 3. Workshop Mobile Commerce, Universität Augsburg, 04.02.2003

P-26 Gerhard Weikum, Harald Schöning, Erhard Rahm (Hrsg.): BTW 2003: Datenbanksysteme für Business, Technologie und Web

P-27 Michael Kroll, Hans-Gerd Lipinski, Kay Melzer (Hrsg.): Mobiles Computing in der Medizin

P-28 Ulrich Reimer, Andreas Abecker, Steffen Staab, Gerd Stumme (Hrsg.): WM 2003: Professionelles Wissensmanagement – Er-fahrungen und Visionen

P-29 Antje Düsterhöft, Bernhard Thalheim (Eds.): NLDB'2003: Natural Language Processing and Information Systems

P-30 Mikhail Godlevsky, Stephen Liddle, Heinrich C. Mayr (Eds.): Information Systems Technology and its Applications

P-31 Arslan Brömme, Christoph Busch (Eds.): BIOSIG 2003: Biometrics and Electronic Signatures, Proceedings of the Special Interest Group on Biometrics and Electronic Signatures,
24. July 2003 in Darmstadt, Germany

P-32 Peter Hubwieser (Hrsg.): Informatische Fachkonzepte im Unterricht – INFOS 2003

P-33 Andreas Geyer-Schulz, Alfred Taudes (Hrsg.): Informationswirtschaft: Ein Sektor mit Zukunft

P-34 Klaus Dittrich, Wolfgang König, Andreas Oberweis, Kai Rannenberg, Wolfgang Wahlster (Hrsg.): Informatik 2003 – Innovative Informatikanwendungen (Band 1)

P-35 Klaus Dittrich, Wolfgang König, Andreas Oberweis, Kai Rannenberg, Wolfgang Wahlster (Hrsg.): Informatik 2003 – Innovative Informatikanwendungen (Band 2)

P-36 Rüdiger Grimm, Hubert B. Keller, Kai Rannenberg (Hrsg.): Informatik 2003 – Mit Sicherheit Informatik

P-37 Arndt Bode, Jörg Desel, Sabine Rathmayer, Martin Wessner (Hrsg.): DeLFI 2003: e-Learning Fachtagung Informatik

P-38 E.J. Sinz, M. Plaha, P. Neckel (Hrsg.): Modellierung betrieblicher Informationssysteme – MobIS 2003

P-39 Jens Nedon, Sandra Frings, Oliver Göbel (Hrsg.): IT-Incident Management & IT-Forensics – IMF 2003

P-40 Michael Rebstock (Hrsg.): Modellierung betrieblicher Informationssysteme – MobIS 2004

P-41 Uwe Brinkschulte, Jürgen Becker, Dietmar Fey, Karl-Erwin Großpietsch, Christian Hochberger, Erik Maehle, Thomas Runkler (Edts.): ARCS 2004 – Organic and Pervasive Computing

P-42 Key Pousttchi, Klaus Turowski (Hrsg.): Mobile Economy – Transaktionen und Prozesse, Anwendungen und Dienste

P-43 Birgitta König-Ries, Michael Klein, Philipp Obreiter (Hrsg.): Persistance, Scalability, Transactions – Database Mechanisms for Mobile Applications

P-44 Jan von Knop, Wilhelm Haverkamp, Eike Jessen (Hrsg.): Security, E-Learning. E-Services

P-45 Bernhard Rumpe, Wofgang Hesse (Hrsg.): Modellierung 2004

P-46 Ulrich Flegel, Michael Meier (Hrsg.): Detection of Intrusions of Malware & Vulnerability Assessment

P-47 Alexander Prosser, Robert Krimmer (Hrsg.): Electronic Voting in Europe – Technology, Law, Politics and Society

P-48 Anatoly Doroshenko, Terry Halpin, Stephen W. Liddle, Heinrich C. Mayr (Hrsg.): Information Systems Technology and its Applications

P-49 G. Schiefer, P. Wagner, M. Morgenstern, U. Rickert (Hrsg.): Integration und Datensicherheit – Anforderungen, Konflikte und Perspektiven

P-50 Peter Dadam, Manfred Reichert (Hrsg.): INFORMATIK 2004 – Informatik verbindet (Band 1) Beiträge der 34. Jahrestagung der Gesellschaft für Informatik e.V. (GI), 20.-24. September 2004 in Ulm

P-51 Peter Dadam, Manfred Reichert (Hrsg.): INFORMATIK 2004 – Informatik verbindet (Band 2) Beiträge der 34. Jahrestagung der Gesellschaft für Informatik e.V. (GI), 20.-24. September 2004 in Ulm

P-52 Gregor Engels, Silke Seehusen (Hrsg.): DELFI 2004 – Tagungsband der 2. e-Learning Fachtagung Informatik

P-53 Robert Giegerich, Jens Stoye (Hrsg.): German Conference on Bioinformatics – GCB 2004

P-54 Jens Borchers, Ralf Kneuper (Hrsg.): Softwaremanagement 2004 – Outsourcing und Integration

P-55 Jan von Knop, Wilhelm Haverkamp, Eike Jessen (Hrsg.): E-Science und Grid Adhoc-Netze Medienintegration

P-56 Fernand Feltz, Andreas Oberweis, Benoit Otjacques (Hrsg.): EMISA 2004 – Informationssysteme im E-Business und E-Government

P-57 Klaus Turowski (Hrsg.): Architekturen, Komponenten, Anwendungen

P-58 Sami Beydeda, Volker Gruhn, Johannes Mayer, Ralf Reussner, Franz Schweiggert (Hrsg.): Testing of Component-Based Systems and Software Quality

P-59 J. Felix Hampe, Franz Lehner, Key Pousttchi, Kai Ranneberg, Klaus Turowski (Hrsg.): Mobile Business – Processes, Platforms, Payments

P-60 Steffen Friedrich (Hrsg.): Unterrichtskonzepte für inforrmatische Bildung

P-61 Paul Müller, Reinhard Gotzhein, Jens B. Schmitt (Hrsg.): Kommunikation in verteilten Systemen

P-62 Federrath, Hannes (Hrsg.): „Sicherheit 2005“ – Sicherheit – Schutz und Zuverlässigkeit

P-63 Roland Kaschek, Heinrich C. Mayr, Stephen Liddle (Hrsg.): Information Systems – Technology and ist Applications

P-64 Peter Liggesmeyer, Klaus Pohl, Michael Goedicke (Hrsg.): Software Engineering 2005

P-65 Gottfried Vossen, Frank Leymann, Peter Lockemann, Wolffried Stucky (Hrsg.): Datenbanksysteme in Business, Technologie und Web

P-66 Jörg M. Haake, Ulrike Lucke, Djamshid Tavangarian (Hrsg.): DeLFI 2005: 3. deutsche e-Learning Fachtagung Informatik

P-67 Armin B. Cremers, Rainer Manthey, Peter Martini, Volker Steinhage (Hrsg.): INFORMATIK 2005 – Informatik LIVE (Band 1)

P-68 Armin B. Cremers, Rainer Manthey, Peter Martini, Volker Steinhage (Hrsg.): INFORMATIK 2005 – Informatik LIVE (Band 2)

P-69 Robert Hirschfeld, Ryszard Kowalcyk, Andreas Polze, Matthias Weske (Hrsg.): NODe 2005, GSEM 2005

P-70 Klaus Turowski, Johannes-Maria Zaha (Hrsg.): Component-oriented Enterprise Application (COAE 2005)

P-71 Andrew Torda, Stefan Kurz, Matthias Rarey (Hrsg.): German Conference on Bioinformatics 2005

P-72 Klaus P. Jantke, Klaus-Peter Fähnrich, Wolfgang S. Wittig (Hrsg.): Marktplatz Internet: Von e-Learning bis e-Payment

P-73 Jan von Knop, Wilhelm Haverkamp, Eike Jessen (Hrsg.): "Heute schon das Morgen sehen"

P-74 Christopher Wolf, Stefan Lucks, Po-Wah Yau (Hrsg.): WEWoRC 2005 – Western European Workshop on Research in Cryptology

P-75 Jörg Desel, Ulrich Frank (Hrsg.): Enterprise Modelling and Information Systems Architecture

P-76 Thomas Kirste, Birgitta König-Riess, Key Pousttchi, Klaus Turowski (Hrsg.): Mobile Informationssysteme – Potentiale, Hindernisse, Einsatz

P-77 Jana Dittmann (Hrsg.): SICHERHEIT 2006

P-78 K.-O. Wenkel, P. Wagner, M. Morgenstern, K. Luzi, P. Eisermann (Hrsg.): Land- und Ernährungswirtschaft im Wandel

P-79 Bettina Biel, Matthias Book, Volker Gruhn (Hrsg.): Softwareengineering 2006

P-80 Mareike Schoop, Christian Huemer, Michael Rebstock, Martin Bichler (Hrsg.): Service-Oriented Electronic Commerce

P-81 Wolfgang Karl, Jürgen Becker, Karl-Erwin Großpietsch, Christian Hochberger, Erik Maehle (Hrsg.): ARCS´06

P-82 Heinrich C. Mayr, Ruth Breu (Hrsg.): Modellierung 2006

P-83 Daniel Huson, Oliver Kohlbacher, Andrei Lupas, Kay Nieselt and Andreas Zell (eds.): German Conference on Bioinformatics

P-84 Dimitris Karagiannis, Heinrich C. Mayr, (Hrsg.): Information Systems Technology and its Applications

P-85 Witold Abramowicz, Heinrich C. Mayr, (Hrsg.): Business Information Systems

P-86 Robert Krimmer (Ed.): Electronic Voting 2006

P-87 Max Mühlhäuser, Guido Rößling, Ralf Steinmetz (Hrsg.): DELFI 2006: 4. e-Learning Fachtagung Informatik

P-88 Robert Hirschfeld, Andreas Polze, Ryszard Kowalczyk (Hrsg.): NODe 2006, GSEM 2006

P-90 Joachim Schelp, Robert Winter, Ulrich Frank, Bodo Rieger, Klaus Turowski (Hrsg.): Integration, Informationslogistik und Architektur

P-91 Henrik Stormer, Andreas Meier, Michael Schumacher (Eds.): European Conference on eHealth 2006

P-92 Fernand Feltz, Benoît Otjacques, Andreas Oberweis, Nicolas Poussing (Eds.): AIM 2006

P-93 Christian Hochberger, Rüdiger Liskowsky (Eds.): INFORMATIK 2006 – Informatik für Menschen, Band 1

P-94 Christian Hochberger, Rüdiger Liskowsky (Eds.): INFORMATIK 2006 – Informatik für Menschen, Band 2

P-95 Matthias Weske, Markus Nüttgens (Eds.): EMISA 2005: Methoden, Konzepte und Technologien für die Entwicklung von dienstbasierten Informationssystemen

P-96 Saartje Brockmans, Jürgen Jung, York Sure (Eds.): Meta-Modelling and Ontologies

P-97 Oliver Göbel, Dirk Schadt, Sandra Frings, Hardo Hase, Detlef Günther, Jens Nedon (Eds.): IT-Incident Mangament & IT-Forensics – IMF 2006

P-98 Hans Brandt-Pook, Werner Simonsmeier und Thorsten Spitta (Hrsg.): Beratung in der Softwareentwicklung – Modelle, Methoden, Best Practices

P-99 Andreas Schwill, Carsten Schulte, Marco Thomas (Hrsg.): Didaktik der Informatik

P-100 Peter Forbrig, Günter Siegel, Markus Schneider (Hrsg.): HDI 2006: Hochschuldidaktik der Informatik

P-101 Stefan Böttinger, Ludwig Theuvsen, Susanne Rank, Marlies Morgenstern (Hrsg.): Agrarinformatik im Spannungsfeld zwischen Regionalisierung und globalen Wertschöpfungsketten

P-102 Otto Spaniol (Eds.): Mobile Services and Personalized Environments

P-103 Alfons Kemper, Harald Schöning, Thomas Rose, Matthias Jarke, Thomas Seidl, Christoph Quix, Christoph Brochhaus (Hrsg.): Datenbanksysteme in Business, Technologie und Web (BTW 2007)

P-104 Birgitta König-Ries, Franz Lehner, Rainer Malaka, Can Türker (Hrsg.) MMS 2007: Mobilität und mobile Informationssysteme

P-105 Wolf-Gideon Bleek, Jörg Raasch, Heinz Züllighoven (Hrsg.) Software Engineering 2007

P-106 Wolf-Gideon Bleek, Henning Schwentner, Heinz Züllighoven (Hrsg.) Software Engineering 2007 – Beiträge zu den Workshops

P-107 Heinrich C. Mayr, Dimitris Karagiannis (eds.) Information Systems Technology and its Applications

P-108 Arslan Brömme, Christoph Busch, Detlef Hühnlein (eds.) BIOSIG 2007: Biometrics and Electronic Signatures

P-109 Rainer Koschke, Otthein Herzog, Karl-Heinz Rödiger, Marc Ronthaler (Hrsg.) INFORMATIK 2007 Informatik trifft Logistik Band 1

P-110 Rainer Koschke, Otthein Herzog, Karl-Heinz Rödiger, Marc Ronthaler (Hrsg.) INFORMATIK 2007 Informatik trifft Logistik Band 2

P-111 Christian Eibl, Johannes Magenheim, Sigrid Schubert, Martin Wessner (Hrsg.) DeLFI 2007: 5. e-Learning Fachtagung Informatik

P-112 Sigrid Schubert (Hrsg.) Didaktik der Informatik in Theorie und Praxis

P-113 Sören Auer, Christian Bizer, Claudia Müller, Anna V. Zhdanova (Eds.) The Social Semantic Web 2007 Proceedings of the 1st Conference on Social Semantic Web (CSSW)

P-114 Sandra Frings, Oliver Göbel, Detlef Günther, Hardo G. Hase, Jens Nedon, Dirk Schadt, Arslan Brömme (Eds.) IMF2007 IT-incident management & IT-forensics Proceedings of the 3rd International Conference on IT-Incident Management & IT-Forensics

P-115 Claudia Falter, Alexander Schliep, Joachim Selbig, Martin Vingron and Dirk Walther (Eds.) German conference on bioinformatics GCB 2007

P-116 Witold Abramowicz, Leszek Maciszek (Eds.) Business Process and Services Computing 1st International Working Conference on Business Process and Services Computing BPSC 2007

P-117 Ryszard Kowalczyk (Ed.) Grid service engineering and manegement The 4th International Conference on Grid Service Engineering and Management GSEM 2007

P-118 Andreas Hein, Wilfried Thoben, Hans-Jürgen Appelrath, Peter Jensch (Eds.) European Conference on ehealth 2007

P-119 Manfred Reichert, Stefan Strecker, Klaus Turowski (Eds.) Enterprise Modelling and Information Systems Architectures Concepts and Applications

P-120 Adam Pawlak, Kurt Sandkuhl, Wojciech Cholewa, Leandro Soares Indrusiak (Eds.) Coordination of Collaborative Engineering - State of the Art and Future Challenges

P-121 Korbinian Herrmann, Bernd Bruegge (Hrsg.) Software Engineering 2008 Fachtagung des GI-Fachbereichs Softwaretechnik

P-122 Walid Maalej, Bernd Bruegge (Hrsg.) Software Engineering 2008 - Workshopband Fachtagung des GI-Fachbereichs Softwaretechnik

P-123 Michael H. Breitner, Martin Breunig, Elgar Fleisch, Ley Pousttchi, Klaus Turowski (Hrsg.)
Mobile und Ubiquitäre Informationssysteme – Technologien, Prozesse, Marktfähigkeit
Proceedings zur 3. Konferenz Mobile und Ubiquitäre Informationssysteme (MMS 2008)

P-124 Wolfgang E. Nagel, Rolf Hoffmann, Andreas Koch (Eds.)
9th Workshop on Parallel Systems and Algorithms (PASA)
Workshop of the GI/ITG Speciel Interest Groups PARS and PARVA

P-125 Rolf A.E. Müller, Hans-H. Sundermeier, Ludwig Theuvsen, Stephanie Schütze, Marlies Morgenstern (Hrsg.)
Unternehmens-IT: Führungsinstrument oder Verwaltungsbürde
Referate der 28. GIL Jahrestagung

P-126 Rainer Gimnich, Uwe Kaiser, Jochen Quante, Andreas Winter (Hrsg.)
10th Workshop Software Reengineering (WSR 2008)

P-127 Thomas Kühne, Wolfgang Reisig, Friedrich Steimann (Hrsg.)
Modellierung 2008

P-128 Ammar Alkassar, Jörg Siekmann (Hrsg.)
Sicherheit 2008
Sicherheit, Schutz und Zuverlässigkeit
Beiträge der 4. Jahrestagung des Fachbereichs Sicherheit der Gesellschaft für Informatik e.V. (GI)
2.-4. April 2008
Saarbrücken, Germany

P-129 Wolfgang Hesse, Andreas Oberweis (Eds.)
Sigsand-Europe 2008
Proceedings of the Third AIS SIGSAND European Symposium on Analysis, Design, Use and Societal Impact of Information Systems

P-130 Paul Müller, Bernhard Neumair, Gabi Dreo Rodosek (Hrsg.)
1. DFN-Forum Kommunikations-technologien Beiträge der Fachtagung

P-131 Robert Krimmer, Rüdiger Grimm (Eds.)
3rd International Conference on Electronic Voting 2008
Co-organized by Council of Europe, Gesellschaft für Informatik and E-Voting. CC

P-132 Silke Seehusen, Ulrike Lucke, Stefan Fischer (Hrsg.)
DeLFI 2008:
Die 6. e-Learning Fachtagung Informatik

P-133 Heinz-Gerd Hegering, Axel Lehmann, Hans Jürgen Ohlbach, Christian Scheideler (Hrsg.)
INFORMATIK 2008
Beherrschbare Systeme – dank Informatik
Band 1

P-134 Heinz-Gerd Hegering, Axel Lehmann, Hans Jürgen Ohlbach, Christian Scheideler (Hrsg.)
INFORMATIK 2008
Beherrschbare Systeme – dank Informatik
Band 2

P-135 Torsten Brinda, Michael Fothe, Peter Hubwieser, Kirsten Schlüter (Hrsg.)
Didaktik der Informatik – Aktuelle Forschungsergebnisse

P-136 Andreas Beyer, Michael Schroeder (Eds.)
German Conference on Bioinformatics GCB 2008

P-137 Arslan Brömme, Christoph Busch, Detlef Hühnlein (Eds.)
BIOSIG 2008: Biometrics and Electronic Signatures

P-138 Barbara Dinter, Robert Winter, Peter Chamoni, Norbert Gronau, Klaus Turowski (Hrsg.)
Synergien durch Integration und Informationslogistik
Proceedings zur DW2008

P-139 Georg Herzwurm, Martin Mikusz (Hrsg.)
Industrialisierung des Software-Managements
Fachtagung des GI-Fachausschusses Management der Anwendungsentwick-lung und -wartung im Fachbereich Wirtschaftsinformatik

P-140 Oliver Göbel, Sandra Frings, Detlef Günther, Jens Nedon, Dirk Schadt (Eds.)
IMF 2008 - IT Incident Management & IT Forensics

P-141 Peter Loos, Markus Nüttgens, Klaus Turowski, Dirk Werth (Hrsg.)
Modellierung betrieblicher Informations-systeme (MobIS 2008)
Modellierung zwischen SOA und Compliance Management

P-142 R. Bill, P. Korduan, L. Theuvsen, M. Morgenstern (Hrsg.)
Anforderungen an die Agrarinformatik durch Globalisierung und Klimaveränderung

P-143 Peter Liggesmeyer, Gregor Engels, Jürgen Münch, Jörg Dörr, Norman Riegel (Hrsg.)
Software Engineering 2009
Fachtagung des GI-Fachbereichs Softwaretechnik

P-144 Johann-Christoph Freytag, Thomas Ruf, Wolfgang Lehner, Gottfried Vossen (Hrsg.)
Datenbanksysteme in Business, Technologie und Web (BTW)

P-145 Knut Hinkelmann, Holger Wache (Eds.)
WM2009: 5th Conference on Professional Knowledge Management

P-146 Markus Bick, Martin Breunig, Hagen Höpfner (Hrsg.)
Mobile und Ubiquitäre Informationssysteme – Entwicklung, Implementierung und Anwendung
4. Konferenz Mobile und Ubiquitäre Informationssysteme (MMS 2009)

P-147 Witold Abramowicz, Leszek Maciaszek, Ryszard Kowalczyk, Andreas Speck (Eds.)
Business Process, Services Computing and Intelligent Service Management
BPSC 2009 · ISM 2009 · YRW-MBP 2009

P-148 Christian Erfurth, Gerald Eichler, Volkmar Schau (Eds.)
9th International Conference on Innovative Internet Community Systems
I^2CS 2009

P-149 Paul Müller, Bernhard Neumair, Gabi Dreo Rodosek (Hrsg.)
2. DFN-Forum Kommunikationstechnologien
Beiträge der Fachtagung

P-150 Jürgen Münch, Peter Liggesmeyer (Hrsg.)
Software Engineering
2009 - Workshopband

P-151 Armin Heinzl, Peter Dadam, Stefan Kirn, Peter Lockemann (Eds.)
PRIMIUM
Process Innovation for Enterprise Software

P-152 Jan Mendling, Stefanie Rinderle-Ma, Werner Esswein (Eds.)
Enterprise Modelling and Information Systems Architectures
Proceedings of the 3rd Int'l Workshop EMISA 2009

P-153 Andreas Schwill, Nicolas Apostolopoulos (Hrsg.)
Lernen im Digitalen Zeitalter
DeLFI 2009 – Die 7. E-Learning Fachtagung Informatik

P-154 Stefan Fischer, Erik Maehle Rüdiger Reischuk (Hrsg.)
INFORMATIK 2009
Im Focus das Leben

P-155 Arslan Brömme, Christoph Busch, Detlef Hühnlein (Eds.)
BIOSIG 2009:
Biometrics and Electronic Signatures
Proceedings of the Special Interest Group on Biometrics and Electronic Signatures

P-156 Bernhard Koerber (Hrsg.)
Zukunft braucht Herkunft
25 Jahre »INFOS – Informatik und Schule«

P-157 Ivo Grosse, Steffen Neumann, Stefan Posch, Falk Schreiber, Peter Stadler (Eds.)
German Conference on Bioinformatics 2009

P-158 W. Claupein, L. Theuvsen, A. Kämpf, M. Morgenstern (Hrsg.)
Precision Agriculture Reloaded – Informationsgestützte Landwirtschaft

P-159 Gregor Engels, Markus Luckey, Wilhelm Schäfer (Hrsg.)
Software Engineering 2010

P-160 Gregor Engels, Markus Luckey, Alexander Pretschner, Ralf Reussner (Hrsg.)
Software Engineering 2010 – Workshopband
(inkl. Doktorandensymposium)

P-161 Gregor Engels, Dimitris Karagiannis Heinrich C. Mayr (Hrsg.)
Modellierung 2010

P-162 Maria A. Wimmer, Uwe Brinkhoff, Siegfried Kaiser, Dagmar Lück-Schneider, Erich Schweighofer, Andreas Wiebe (Hrsg.)
Vernetzte IT für einen effektiven Staat
Gemeinsame Fachtagung Verwaltungsinformatik (FTVI) und Fachtagung Rechtsinformatik (FTRI) 2010

P-163 Markus Bick, Stefan Eulgem, Elgar Fleisch, J. Felix Hampe, Birgitta König-Ries, Franz Lehner, Key Pousttchi, Kai Rannenberg (Hrsg.)
Mobile und Ubiquitäre Informationssysteme
Technologien, Anwendungen und Dienste zur Unterstützung von mobiler Kollaboration

P-164 Arslan Brömme, Christoph Busch (Eds.)
BIOSIG 2010: Biometrics and Electronic Signatures Proceedings of the Special Interest Group on Biometrics and Electronic Signatures

P-165 Gerald Eichler, Peter Kropf, Ulrike Lechner, Phayung Meesad, Herwig Unger (Eds.)
10th International Conference on Innovative Internet Community Systems (I^2CS) – Jubilee Edition 2010 –

P-166 Paul Müller, Bernhard Neumair, Gabi Dreo Rodosek (Hrsg.)
3. DFN-Forum Kommunikationstechnologien Beiträge der Fachtagung

P-167 Robert Krimmer, Rüdiger Grimm (Eds.)
4th International Conference on Electronic Voting 2010
co-organized by the Council of Europe, Gesellschaft für Informatik and E-Voting.CC

P-168 Ira Diethelm, Christina Dörge, Claudia Hildebrandt, Carsten Schulte (Hrsg.)
Didaktik der Informatik
Möglichkeiten empirischer Forschungsmethoden und Perspektiven der Fachdidaktik

P-169 Michael Kerres, Nadine Ojstersek Ulrik Schroeder, Ulrich Hoppe (Hrsg.)
DeLFI 2010 - 8. Tagung der Fachgruppe E-Learning der Gesellschaft für Informatik e.V.

P-170 Felix C. Freiling (Hrsg.)
Sicherheit 2010
Sicherheit, Schutz und Zuverlässigkeit

P-171 Werner Esswein, Klaus Turowski, Martin Juhrisch (Hrsg.)
Modellierung betrieblicher Informationssysteme (MobIS 2010)
Modellgestütztes Management

P-172 Stefan Klink, Agnes Koschmider Marco Mevius, Andreas Oberweis (Hrsg.)
EMISA 2010
Einflussfaktoren auf die Entwicklung flexibler, integrierter Informationssysteme
Beiträge des Workshops der GI-Fachgruppe EMISA (Entwicklungsmethoden für Informationssysteme und deren Anwendung)

P-173 Dietmar Schomburg, Andreas Grote (Eds.)
German Conference on Bioinformatics 2010

P-174 Arslan Brömme, Torsten Eymann, Detlef Hühnlein, Heiko Roßnagel, Paul Schmücker (Hrsg.)
perspeGKtive 2010
Workshop „Innovative und sichere Informationstechnologie für das Gesundheitswesen von morgen“

P-175 Klaus-Peter Fähnrich, Bogdan Franczyk (Hrsg.)
INFORMATIK 2010
Service Science – Neue Perspektiven für die Informatik
Band 1

P-176 Klaus-Peter Fähnrich, Bogdan Franczyk (Hrsg.)
INFORMATIK 2010
Service Science – Neue Perspektiven für die Informatik
Band 2

P-177 Witold Abramowicz, Rainer Alt, Klaus-Peter Fähnrich, Bogdan Franczyk, Leszek A. Maciaszek (Eds.)
INFORMATIK 2010
Business Process and Service Science – Proceedings of ISSS and BPSC

P-178 Wolfram Pietsch, Benedikt Krams (Hrsg.)
Vom Projekt zum Produkt
Fachtagung des GI-Fachausschusses Management der Anwendungsentwicklung und -wartung im Fachbereich Wirtschafts-informatik (WI-MAW), Aachen, 2010

P-179 Stefan Gruner, Bernhard Rumpe (Eds.)
FM+AM`2010
Second International Workshop on Formal Methods and Agile Methods

P-180 Theo Härder, Wolfgang Lehner, Bernhard Mitschang, Harald Schöning, Holger Schwarz (Hrsg.)
Datenbanksysteme für Business, Technologie und Web (BTW)
14. Fachtagung des GI-Fachbereichs „Datenbanken und Informationssysteme“ (DBIS)

P-181 Michael Clasen, Otto Schätzel, Brigitte Theuvsen (Hrsg.)
Qualität und Effizienz durch informationsgestützte Landwirtschaft, Fokus: Moderne Weinwirtschaft

P-182 Ronald Maier (Hrsg.)
6th Conference on Professional Knowledge Management
From Knowledge to Action

P-183 Ralf Reussner, Matthias Grund, Andreas Oberweis, Walter Tichy (Hrsg.)
Software Engineering 2011
Fachtagung des GI-Fachbereichs Softwaretechnik

P-184 Ralf Reussner, Alexander Pretschner, Stefan Jähnichen (Hrsg.)
Software Engineering 2011
Workshopband
(inkl. Doktorandensymposium)

P-185 Hagen Höpfner, Günther Specht, Thomas Ritz, Christian Bunse (Hrsg.)
MMS 2011: Mobile und ubiquitäre Informationssysteme Proceedings zur 6. Konferenz Mobile und Ubiquitäre Informationssysteme (MMS 2011)

P-186 Gerald Eichler, Axel Küpper, Volkmar Schau, Hacène Fouchal, Herwig Unger (Eds.)
11th International Conference on Innovative Internet Community Systems (I^2CS)

P-187 Paul Müller, Bernhard Neumair, Gabi Dreo Rodosek (Hrsg.)
4. DFN-Forum Kommunikations-technologien, Beiträge der Fachtagung 20. Juni bis 21. Juni 2011 Bonn

P-188 Holger Rohland, Andrea Kienle, Steffen Friedrich (Hrsg.)
DeLFI 2011 – Die 9. e-Learning Fachtagung Informatik der Gesellschaft für Informatik e.V. 5.–8. September 2011, Dresden

P-189 Thomas, Marco (Hrsg.)
Informatik in Bildung und Beruf
INFOS 2011
14. GI-Fachtagung Informatik und Schule

P-190 Markus Nüttgens, Oliver Thomas, Barbara Weber (Eds.)
Enterprise Modelling and Information Systems Architectures (EMISA 2011)

P-191 Arslan Brömme, Christoph Busch (Eds.)
BIOSIG 2011
International Conference of the Biometrics Special Interest Group

P-192 Hans-Ulrich Heiß, Peter Pepper, Holger Schlingloff, Jörg Schneider (Hrsg.)
INFORMATIK 2011
Informatik schafft Communities

P-193 Wolfgang Lehner, Gunther Piller (Hrsg.)
IMDM 2011

P-194 M. Clasen, G. Fröhlich, H. Bernhardt, K. Hildebrand, B. Theuvsen (Hrsg.)
Informationstechnologie für eine nachhaltige Landbewirtschaftung
Fokus Forstwirtschaft

P-195 Neeraj Suri, Michael Waidner (Hrsg.)
Sicherheit 2012
Sicherheit, Schutz und Zuverlässigkeit
Beiträge der 6. Jahrestagung des Fachbereichs Sicherheit der Gesellschaft für Informatik e.V. (GI)

P-196 Arslan Brömme, Christoph Busch (Eds.)
BIOSIG 2012
Proceedings of the 11th International Conference of the Biometrics Special Interest Group

P-197 Jörn von Lucke, Christian P. Geiger, Siegfried Kaiser, Erich Schweighofer, Maria A. Wimmer (Hrsg.)
Auf dem Weg zu einer offenen, smarten und vernetzten Verwaltungskultur
Gemeinsame Fachtagung Verwaltungsinformatik (FTVI) und Fachtagung Rechtsinformatik (FTRI) 2012

P-198 Stefan Jähnichen, Axel Küpper, Sahin Albayrak (Hrsg.)
Software Engineering 2012
Fachtagung des GI-Fachbereichs Softwaretechnik

P-199 Stefan Jähnichen, Bernhard Rumpe, Holger Schlingloff (Hrsg.)
Software Engineering 2012
Workshopband

P-200 Gero Mühl, Jan Richling, Andreas Herkersdorf (Hrsg.)
ARCS 2012 Workshops

P-201 Elmar J. Sinz Andy Schürr (Hrsg.)
Modellierung 2012

P-202 Andrea Back, Markus Bick, Martin Breunig, Key Pousttchi, Frédéric Thiesse (Hrsg.)
MMS 2012:Mobile und Ubiquitäre Informationssysteme

P-203 Paul Müller, Bernhard Neumair, Helmut Reiser, Gabi Dreo Rodosek (Hrsg.)
5. DFN-Forum Kommunikations-technologien
Beiträge der Fachtagung

P-204 Gerald Eichler, Leendert W. M. Wienhofen, Anders Kofod-Petersen, Herwig Unger (Eds.)
12th International Conference on Innovative Internet Community Systems (I2CS 2012)

P-205 Manuel J. Kripp, Melanie Volkamer, Rüdiger Grimm (Eds.)
5th International Conference on Electronic Voting 2012 (EVOTE2012)
Co-organized by the Council of Europe, Gesellschaft für Informatik and E-Voting.CC

P-206 Stefanie Rinderle-Ma, Mathias Weske (Hrsg.)
EMISA 2012
Der Mensch im Zentrum der Modellierung

P-207 Jörg Desel, Jörg M. Haake, Christian Spannagel (Hrsg.)
DeLFI 2012: Die 10. e-Learning Fachtagung Informatik der Gesellschaft für Informatik e.V.
24.–26. September 2012

P-208 Ursula Goltz, Marcus Magnor,
Hans-Jürgen Appelrath, Herbert Matthies,
Wolf-Tilo Balke, Lars Wolf (Hrsg.)
INFORMATIK 2012

P-209 Hans Brandt-Pook, André Fleer, Thorsten
Spitta, Malte Wattenberg (Hrsg.)
Nachhaltiges Software Management

P-210 Erhard Plödereder, Peter Dencker,
Herbert Klenk, Hubert B. Keller,
Silke Spitzer (Hrsg.)
Automotive – Safety & Security 2012
Sicherheit und Zuverlässigkeit für
automobile Informationstechnik

P-211 M. Clasen, K. C. Kersebaum, A.
Meyer-Aurich, B. Theuvsen (Hrsg.)
Massendatenmanagement in der
Agrar- und Ernährungswirtschaft
Erhebung - Verarbeitung - Nutzung
Referate der 33. GIL-Jahrestagung
20. – 21. Februar 2013, Potsdam

P-212 Arslan Brömme, Christoph Busch (Eds.)
BIOSIG 2013
Proceedings of the 12th International
Conference of the Biometrics
Special Interest Group
04.–06. September 2013
Darmstadt, Germany

P-213 Stefan Kowalewski,
Bernhard Rumpe (Hrsg.)
Software Engineering 2013
Fachtagung des GI-Fachbereichs
Softwaretechnik

P-214 Volker Markl, Gunter Saake, Kai-Uwe
Sattler, Gregor Hackenbroich, Bernhard Mit
schang, Theo Härder, Veit Köppen (Hrsg.)
Datenbanksysteme für Business,
Technologie und Web (BTW) 2013
13. – 15. März 2013, Magdeburg

P-215 Stefan Wagner, Horst Lichter (Hrsg.)
Software Engineering 2013
Workshopband
(inkl. Doktorandensymposium)
26. Februar – 1. März 2013, Aachen

P-216 Gunter Saake, Andreas Henrich,
Wolfgang Lehner, Thomas Neumann,
Veit Köppen (Hrsg.)
Datenbanksysteme für Business,
Technologie und Web (BTW) 2013 –
Workshopband
11. – 12. März 2013, Magdeburg

P-217 Paul Müller, Bernhard Neumair, Helmut
Reiser, Gabi Dreo Rodosek (Hrsg.)
6. DFN-Forum Kommunikations-
technologien
Beiträge der Fachtagung
03.–04. Juni 2013, Erlangen

P-218 Andreas Breiter, Christoph Rensing (Hrsg.)
DeLFI 2013: Die 11 e-Learning
Fachtagung Informatik der Gesellschaft
für Informatik e.V. (GI)
8. – 11. September 2013, Bremen

P-219 Norbert Breier, Peer Stechert,
Thomas Wilke (Hrsg.)
Informatik erweitert Horizonte
INFOS 2013
15. GI-Fachtagung Informatik und Schule
26. – 28. September 2013

P-220 Matthias Horbach (Hrsg.)
INFORMATIK 2013
Informatik angepasst an Mensch,
Organisation und Umwelt
16. – 20. September 2013, Koblenz

P-221 Maria A. Wimmer, Marijn Janssen,
Ann Macintosh, Hans Jochen Scholl,
Efthimios Tambouris (Eds.)
Electronic Government and
Electronic Participation
Joint Proceedings of Ongoing Research of
IFIP EGOV and IFIP ePart 2013
16. – 19. September 2013, Koblenz

P-222 Reinhard Jung, Manfred Reichert (Eds.)
Enterprise Modelling
and Information Systems Architectures
(EMISA 2013)
St. Gallen, Switzerland
September 5. – 6. 2013

P-223 Detlef Hühnlein, Heiko Roßnagel (Hrsg.)
Open Identity Summit 2013
10. – 11. September 2013
Kloster Banz, Germany

P-224 Eckhart Hanser, Martin Mikusz, Masud
Fazal-Baqaie (Hrsg.)
Vorgehensmodelle 2013
Vorgehensmodelle – Anspruch und
Wirklichkeit
20. Tagung der Fachgruppe
Vorgehensmodelle im Fachgebiet
Wirtschaftsinformatik (WI-VM) der
Gesellschaft für Informatik e.V.
Lörrach, 2013

P-225 Hans-Georg Fill, Dimitris Karagiannis,
Ulrich Reimer (Hrsg.)
Modellierung 2014
19. – 21. März 2014, Wien

P-226 M. Clasen, M. Hamer, S. Lehnert,
B. Petersen, B. Theuvsen (Hrsg.)
IT-Standards in der Agrar- und
Ernährungswirtschaft Fokus: Risiko- und
Krisenmanagement
Referate der 34. GIL-Jahrestagung
24. – 25. Februar 2014, Bonn

P-227 Wilhelm Hasselbring,
Nils Christian Ehmke (Hrsg.)
Software Engineering 2014
Fachtagung des GI-Fachbereichs Softwaretechnik
25. – 28. Februar 2014
Kiel, Deutschland

P-228 Stefan Katzenbeisser, Volkmar Lotz, Edgar Weippl (Hrsg.)
Sicherheit 2014
Sicherheit, Schutz und Zuverlässigkeit
Beiträge der 7. Jahrestagung des Fachbereichs Sicherheit der Gesellschaft für Informatik e.V. (GI)
19. – 21. März 2014, Wien

P-229 Dagmar Lück-Schneider, Thomas Gordon, Siegfried Kaiser, Jörn von Lucke,Erich Schweighofer, Maria A.Wimmer, Martin G. Löhe (Hrsg.)
Gemeinsam Electronic Government ziel(gruppen)gerecht gestalten und organisieren
Gemeinsame Fachtagung Verwaltungsinformatik (FTVI) und Fachtagung Rechtsinformatik (FTRI) 2014, 20.-21. März 2014 in Berlin

P-230 Arslan Brömme, Christoph Busch (Eds.)
BIOSIG 2014
Proceedings of the 13th International Conference of the Biometrics Special Interest Group
10. – 12. September 2014 in Darmstadt, Germany

P-231 Paul Müller, Bernhard Neumair, Helmut Reiser, Gabi Dreo Rodosek (Hrsg.)
7. DFN-Forum Kommunikationstechnologien
16. – 17. Juni 2014
Fulda

P-232 E. Plödereder, L. Grunske, E. Schneider, D. Ull (Hrsg.)
INFORMATIK 2014
Big Data – Komplexität meistern
22. – 26. September 2014
Stuttgart

P-233 Stephan Trahasch, Rolf Plötzner, Gerhard Schneider, Claudia Gayer, Daniel Sassiat, Nicole Wöhrle (Hrsg.)
DeLFI 2014 – Die 12. e-Learning Fachtagung Informatik
der Gesellschaft für Informatik e.V.
15. – 17. September 2014
Freiburg

P-234 Fernand Feltz, Bela Mutschler, Benoît Otjacques (Eds.)
Enterprise Modelling and Information Systems Architectures
(EMISA 2014)
Luxembourg, September 25-26, 2014

P-235 Robert Giegerich,
Ralf Hofestädt,
Tim W. Nattkemper (Eds.)
German Conference on Bioinformatics 2014
September 28 – October 1
Bielefeld, Germany

P-236 Martin Engstler, Eckhart Hanser, Martin Mikusz, Georg Herzwurm (Hrsg.)
Projektmanagement und Vorgehensmodelle 2014
Soziale Aspekte und Standardisierung
Gemeinsame Tagung der Fachgruppen Projektmanagement (WI-PM) und Vorgehensmodelle (WI-VM) im Fachgebiet Wirtschaftsinformatik der Gesellschaft für Informatik e.V., Stuttgart 2014

P-237 Detlef Hühnlein, Heiko Roßnagel (Hrsg.)
Open Identity Summit 2014
4.–6. November 2014
Stuttgart, Germany

P-238 Arno Ruckelshausen, Hans-Peter Schwarz, Brigitte Theuvsen (Hrsg.)
Informatik in der Land-, Forst- und Ernährungswirtschaft
Referate der 35. GIL-Jahrestagung
23. – 24. Februar 2015, Geisenheim

P-239 Uwe Aßmann, Birgit Demuth, Thorsten Spitta, Georg Püschel, Ronny Kaiser (Hrsg.)
Software Engineering & Management 2015
17.-20. März 2015, Dresden

P-240 Herbert Klenk, Hubert B. Keller, Erhard Plödereder, Peter Dencker (Hrsg.)
Automotive – Safety & Security 2015
Sicherheit und Zuverlässigkeit für automobile Informationstechnik
21.–22. April 2015, Stuttgart

P-241 Thomas Seidl, Norbert Ritter, Harald Schöning, Kai-Uwe Sattler, Theo Härder, Steffen Friedrich, Wolfram Wingerath (Hrsg.)
Datenbanksysteme für Business, Technologie und Web (BTW 2015)
04. – 06. März 2015, Hamburg

P-242 Norbert Ritter, Andreas Henrich, Wolfgang Lehner, Andreas Thor, Steffen Friedrich, Wolfram Wingerath (Hrsg.)
Datenbanksysteme für Business, Technologie und Web (BTW 2015) – Workshopband
02. – 03. März 2015, Hamburg

P-243 Paul Müller, Bernhard Neumair, Helmut Reiser, Gabi Dreo Rodosek (Hrsg.)
8. DFN-Forum Kommunikationstechnologien
06.–09. Juni 2015, Lübeck

P-244 Alfred Zimmermann, Alexander Rossmann (Eds.)
Digital Enterprise Computing (DEC 2015)
Böblingen, Germany June 25-26, 2015

P-245 Arslan Brömme, Christoph Busch , Christian Rathgeb, Andreas Uhl (Eds.)
BIOSIG 2015
Proceedings of the 14th International Conference of the Biometrics Special Interest Group
09.–11. September 2015
Darmstadt, Germany

P-246 Douglas W. Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.)
INFORMATIK 2015
28.9.-2.10. 2015, Cottbus

P-247 Hans Pongratz, Reinhard Keil (Hrsg.)
DeLFI 2015 – Die 13. E-Learning Fachtagung Informatik der Gesellschaft für Informatik e.V. (GI)
1.–4. September 2015
München

P-248 Jens Kolb, Henrik Leopold, Jan Mendling (Eds.)
Enterprise Modelling and Information Systems Architectures
Proceedings of the 6th Int. Workshop on Enterprise Modelling and Information Systems Architectures, Innsbruck, Austria
September 3-4, 2015

P-249 Jens Gallenbacher (Hrsg.)
Informatik allgemeinbildend begreifen
INFOS 2015 16. GI-Fachtagung Informatik und Schule
20.–23. September 2015

P-250 Martin Engstler, Masud Fazal-Baqaie, Eckhart Hanser, Martin Mikusz, Alexander Volland (Hrsg.)
Projektmanagement und Vorgehensmodelle 2015
Hybride Projektstrukturen erfolgreich umsetzen
Gemeinsame Tagung der Fachgruppen Projektmanagement (WI-PM) und Vorgehensmodelle (WI-VM) im Fachgebiet Wirtschaftsinformatik der Gesellschaft für Informatik e.V., Elmshorn 2015

P-251 Detlef Hühnlein, Heiko Roßnagel, Raik Kuhlisch, Jan Ziesing (Eds.)
Open Identity Summit 2015
10.–11. November 2015
Berlin, Germany

P-252 Jens Knoop, Uwe Zdun (Hrsg.)
Software Engineering 2016
Fachtagung des GI-Fachbereichs Softwaretechnik
23.–26. Februar 2016, Wien

P-253 A. Ruckelshausen, A. Meyer-Aurich, T. Rath, G. Recke, B. Theuvsen (Hrsg.)
Informatik in der Land-, Forst- und Ernährungswirtschaft
Fokus: Intelligente Systeme – Stand der Technik und neue Möglichkeiten
Referate der 36. GIL-Jahrestagung
22.-23. Februar 2016, Osnabrück

P-254 Andreas Oberweis, Ralf Reussner (Hrsg.)
Modellierung 2016
2.–4. März 2016, Karlsruhe

P-255 Stefanie Betz, Ulrich Reimer (Hrsg.)
Modellierung 2016 Workshopband
2.–4. März 2016, Karlsruhe

P-256 Michael Meier, Delphine Reinhardt, Steffen Wendzel (Hrsg.)
Sicherheit 2016
Sicherheit, Schutz und Zuverlässigkeit
Beiträge der 8. Jahrestagung des Fachbereichs Sicherheit der Gesellschaft für Informatik e.V. (GI)
5.–7. April 2016, Bonn

P-257 Paul Müller, Bernhard Neumair, Helmut Reiser, Gabi Dreo Rodosek (Hrsg.)
9. DFN-Forum Kommunikationstechnologien
31. Mai – 01. Juni 2016, Rostock

P-258 Dieter Hertweck, Christian Decker (Eds.)
Digital Enterprise Computing (DEC 2016)
14.–15. Juni 2016, Böblingen

P-259 Heinrich C. Mayr, Martin Pinzger (Hrsg.)
INFORMATIK 2016
26.–30. September 2016, Klagenfurt

P-260 Arslan Brömme, Christoph Busch, Christian Rathgeb, Andreas Uhl (Eds.)
BIOSIG 2016
Proceedings of the 15th International Conference of the Biometrics Special Interest Group
21.–23. September 2016, Darmstadt

P-261 Detlef Rätz, Michael Breidung, Dagmar Lück-Schneider, Siegfried Kaiser, Erich Schweighofer (Hrsg.)
Digitale Transformation: Methoden, Kompetenzen und Technologien für die Verwaltung
Gemeinsame Fachtagung Verwaltungsinformatik (FTVI) und Fachtagung Rechtsinformatik (FTRI) 2016
22.–23. September 2016, Dresden

P-262 Ulrike Lucke, Andreas Schwill, Raphael Zender (Hrsg.)
DeLFI 2016 – Die 14. E-Learning Fachtagung Informatik
der Gesellschaft für Informatik e.V. (GI)
11.–14. September 2016, Potsdam

P-263 Martin Engstler, Masud Fazal-Baqaie, Eckhart Hanser, Oliver Linssen, Martin Mikusz, Alexander Volland (Hrsg.)
Projektmanagement und Vorgehensmodelle 2016
Arbeiten in hybriden Projekten: Das Sowohl-als-auch von Stabilität und Dynamik
Gemeinsame Tagung der Fachgruppen Projektmanagement (WI-PM) und Vorgehensmodelle (WI-VM) im Fachgebiet Wirtschaftsinformatik der Gesellschaft für Informatik e.V., Paderborn 2016

P-264 Detlef Hühnlein, Heiko Roßnagel, Christian H. Schunck, Maurizio Talamo (Eds.)
Open Identity Summit 2016
der Gesellschaft für Informatik e.V. (GI)
13.–14. October 2016, Rome, Italy

P-265 Bernhard Mitschang, Daniela Nicklas,Frank Leymann, Harald Schöning, Melanie Herschel, Jens Teubner, Theo Härder, Oliver Kopp, Matthias Wieland (Hrsg.)
Datenbanksysteme für Business, Technologie und Web (BTW 2017)
6.–10. März 2017, Stuttgart

P-266 Bernhard Mitschang, Norbert Ritter, Holger Schwarz, Meike Klettke, Andreas Thor, Oliver Kopp, Matthias Wieland (Hrsg.)
Datenbanksysteme für Business, Technologie und Web (BTW 2017) Workshopband
6.–7. März 2017, Stuttgart

P-267 Jan Jürjens, Kurt Schneider (Hrsg.)
Software Engineering 2017
21.–24. Februar 2017, Hannover

P-268 A. Ruckelshausen, A. Meyer-Aurich, W. Lentz, B. Theuvsen (Hrsg.)
Informatik in der Land-, Forst- und Ernährungswirtschaft
Fokus: Digitale Transformation – Wege in eine zukunftsfähige Landwirtschaft
Referate der 37. GIL-Jahrestagung
06.–07. März 2017, Dresden

P-269 Peter Dencker, Herbert Klenk, Hubert Keller, Erhard Plödereder (Hrsg.)
Automotive – Safety & Security 2017
30.–31. Mai 2017, Stuttgart

P-270 Arslan Brömme, Christoph Busch, Antitza Dantcheva, Christian Rathgeb, Andreas Uhl (Eds.)
BIOSIG 2017
20.–22. September 2017, Darmstadt

P-271 Paul Müller, Bernhard Neumair, Helmut Reiser, Gabi Dreo Rodosek (Hrsg.)
10. DFN-Forum Kommunikations-technologien
30. – 31. Mai 2017, Berlin

P-272 Alexander Rossmann, Alfred Zimmermann (eds.)
Digital Enterprise Computing (DEC 2017)
11.–12. Juli 2017, Böblingen

P-273 Christoph Igel, Carsten Ullrich, Martin Wessner (Hrsg.)
BILDUNGSRÄUME
DeLFI 2017
Die 15. e-Learning Fachtagung Informatik der Gesellschaft für Informatik e.V. (GI)
5. bis 8. September 2017, Chemnitz

P-274 Ira Diethelm (Hrsg.)
Informatische Bildung zum Verstehen und Gestalten der digitalen Welt
13.–15. September 2017, Oldenburg

P-275 Maximilian Eibl, Martin Gaedke (Hrsg.)
INFORMATIK 2017
25.–29. September 2017, Chemnitz

P276 Alexander Volland, Martin Engstler, Masud Fazal-Baqaie, Eckhart Hanser, Oliver Linssen, Martin Mikusz (Hrsg.)
Projektmanagement und Vorgehensmodelle 2017
Die Spannung zwischen dem Prozess und den Menschen im Projekt
Gemeinsame Tagung der Fachgruppen Projektmanagement und Vorgehensmodelle im Fachgebiet Wirtschaftsinformatik der Gesellschaft für Informatik e.V. in Kooperation mit der Fachgruppe IT-Projektmanagement der GPM e.V., Darmstadt 2017

P-277 Lothar Fritsch, Heiko Roßnagel, Detlef Hühnlein (Hrsg.)
Open Identity Summit 2017
5.–6. October 2017, Karlstad, Sweden

P-278 Arno Ruckelshausen, Andreas Meyer-Aurich, Karsten Borchard, Constanze Hofacker, Jens-Peter Loy, Rolf Schwerdtfeger, Hans-Hennig Sundermeier, Helga Floto, Brigitte Theuvsen (Hrsg.)
Informatik in der Land-, Forst- und Ernährungswirtschaft
Referate der 38. GIL-Jahrestagung
26.–27. Februar 2018, Kiel

P-279 Matthias Tichy, Eric Bodden, Marco Kuhrmann, Stefan Wagner, Jan-Philipp Steghöfer (Hrsg.)
Software Engineering und Software Management 2018
5.–9. März 2018, Ulm

P-280 Ina Schaefer, Dimitris Karagiannis, Andreas Vogelsang, Daniel Méndez, Christoph Seidl (Hrsg.)
Modellierung 2018
21.–23. Februar 2018, Braunschweig

P-281 Hanno Langweg, Michael Meier, Bernhard C. Witt, Delphine Reinhardt (Hrsg.)
Sicherheit 2018
Sicherheit, Schutz und Zuverlässigkeit
25.–27. April 2018, Konstanz

P-282 Arslan Brömme, Christoph Busch, Antitza Dantcheva, Christian Rathgeb, Andreas Uhl (Eds.)
BIOSIG 2018
Proceedings of the 17th International Conference of the Biometrics Special Interest Group
26.–28. September 2018
Darmstadt, Germany

P-283 Paul Müller, Bernhard Neumair, Helmut Reiser, Gabi Dreo Rodosek (Hrsg.)
11. DFN-Forum Kommunikations-technologien
27.–28. Juni 2018, Günzburg

P-284 Detlef Krömker, Ulrik Schroeder (Hrsg.)
DeLFI 2018 – Die 16. E-Learning Fachtagung Informatik
10.–12. September 2018, Frankfurt a. M.

P-285 Christian Czarnecki, Carsten Brockmann, Eldar Sultanow, Agnes Koschmider, Annika Selzer (Hrsg.)
Workshops der INFORMATIK 2018 - Architekturen, Prozesse, Sicherheit und Nachhaltigkeit
26.–27. September 2018, Berlin

P-286 Martin Mikusz, Alexander Volland, Martin Engstler, Masud Fazal-Baqaie, Eckhart Hanser, Oliver Linssen (Hrsg.)
Projektmanagement und Vorgehensmodelle 2018
Der Einfluss der Digitalisierung auf Projektmanagementmethoden und Entwicklungsprozesse
Düsseldorf 2018

P-287 A. Meyer-Aurich, M. Gandorfer, N. Barta, A. Gronauer, J. Kantelhardt, H. Floto (Hrsg.)
Informatik in der Land-, Forst- und Ernährungswirtschaft
Fokus: Digitalisierung für landwirtschaftliche Betriebe in kleinstrukturierten Regionen – ein Widerspruch in sich?
Referate der 39. GIL-Jahrestagung
18.–19. Februar 2019, Wien

P-288 Arno Pasternak (Hrsg.)
Informatik für alle
18. GI-Fachtagung
Informatik und Schule
16.-18. September 2019 in Dortmund

P-289 Torsten Grust, Felix Naumann, Alexander Böhm, Wolfgang Lehner, Jens Teubner, Meike Klettke, Theo Härder, Erhard Rahm, Andreas Heuer, Holger Meyer (Hrsg.)
Datenbanksysteme für Business, Technologie und Web (BTW 2019)
4.–8. März 2019 in Rostock

P-290 Holger Meyer, Norbert Ritter, Andreas Thor, Daniela Nicklas, Andreas Heuer, Meike Klettke (Hrsg.)
Datenbanksysteme für Business, Technologie und Web (BTW 2019)
Workshopband
4.–8. März 2019 in Rostock

P-291 Michael Räckers, Sebastian Halsbenning, Detlef Rätz, David Richter, Erich Schweighofer (Hrsg.)
Digitalisierung von Staat und Verwaltung
Gemeinsame Fachtagung Verwaltungsinformatik (FTVI) und Fachtagung Rechtsinformatik (FTRI) 2019
6.–7. März 2019 in Münster

P-292 Steffen Becker, Ivan Bogicevic,Georg Herzwurm, Stefan Wagner (Hrsg.)
Software Engineering and Software Management 2019
18.–22. Februar 2019 in Stuttgart

P-293 Heiko Roßnagel, Sven Wagner, Detlef Hühnlein (Hrsg.)
Open Identity Summit 2019
28.–29. März 2019
Garmisch-Partenkirchen

P-294 Klaus David, Kurt Geihs, Martin Lange, Gerd Stumme (Hrsg.)
INFORMATIK 2019
50 Jahre Gesellschaft für Informatik – Informatik für Gesellschaft
23.–26. September 2019 in Kassel

P-295 Claude Draude, Martin Lange, Bernhard Sick (Hrsg.)
INFORMATIK 2019
50 Jahre Gesellschaft für Informatik – Informatik für Gesellschaft
Workshop-Beiträge
23.–26. September 2019 in Kassel

P-296 Arslan Brömme, Christoph Busch, Antitza Dantcheva, Christian Rathgeb, Andreas Uhl (Eds.)
BIOSIG 2019
Proceedings of the 18th International Conference of the Biometrics Special Interest Group
18.–20. September 2019
Darmstadt, Germany

P-297 Niels Pinkwart, Johannes Konert (Hrsg.)
DELFI 2019 –Die 17. Fachtagung Bildungstechnologien
16.–19. September 2019 in Berlin

P-298 Oliver Linssen, Martin Mikusz, Alexander Volland, Enes Yigitbas, Martin Engstler, Masud Fazal-Baqaie, Marco Kuhrmann (Hrsg.)
Projektmanagement und Vorgehensmodelle 2019 –Neue Vorgehensmodelle in Projekten – Führung, Kulturen und Infrastrukturen im Wandel
1Gemeinsame Tagung der Fachgruppen Projektmanagement (WI-PM), Vorgehensmodelle (WI-VM) und Software Produktmanagement (WI-ProdM) im Fachgebiet Wirtschaftsinformatik der Gesellschaft für Informatik e.V.
in Kooperation mit der Fachgruppe IT-Projektmanagement der GPM e.V., Lörrach 2019

P-299 M. Gandorfer, A. Meyer-Aurich, H. Bernhardt, F. X. Maidl, G. Fröhlich, H. Floto (Hrsg.)
Informatik in der Land-, Forst- und Ernährungswirtschaft
Fokus: Digitalisierung für Mensch, Umwelt und Tier
Referate der 40. GIL-Jahrestagung
17.–18. Februar 2020,
Campus Weihenstephan

P-300 Michael Felderer, Wilhelm Hasselbring, Rick Rabiser, Reiner Jung (Hrsg.)
Software Engineering 2020
24.–28. Februar 2020
Innsbruck, Austria

P-301 Delphine Reinhardt, Hanno Langweg, Bernhard C. Witt, Mathias Fischer (Hrsg.)
Sicherheit 2020
Sicherheit, Schutz und Zuverlässigkeit
17.–20. März 2020, Göttingen

P-302 Dominik Bork, Dimitris Karagiannis, Heinrich C. Mayr (Hrsg.)
Modellierung 2020
19.–21. Februar 2020, Wien

P-303 Peter Heisig, Ronald Orth, Jakob Michael Schönborn, Stefan Thalmann (Hrsg.)
Wissensmanagement in digitalen Arbeitswelten: Aktuelle Ansätze und Perspektiven
18.–20.03.2019, Potsdam

P-304 Heinrich C. Mayr, Stefanie Rinderle-Ma, Stefan Strecker (Hrsg.)
40 Years EMISA
Digital Ecosystems of the Future: Methodology, Techniques and Applications
May 15.–17. 2019
Tutzing am Starnberger See

P-305 Heiko Roßnagel, Christian H. Schunck, Sebastian Mödersheim, Detlef Hühnlein (Hrsg.)
Open Identity Summit 2020
26.–27. May 2020, Copenhagen

P-306 Arslan Brömme, Christoph Busch, Antitza Dantcheva, Kiran Raja, Christian Rathgeb, Andreas Uhl (Eds.)
BIOSIG 2020
Proceedings of the 19th International Conference of the Biometrics Special Interest Group
16.–18. September 2020
International Digital Conference

P-307 Ralf H. Reussner, Anne Koziolek, Robert Heinrich (Hrsg.)
INFORMATIK 2020
Back to the Future
28. September – 2. Oktober 2020, Karlsruhe

P-308 Raphael Zender, Dirk Ifenthaler, Thiemo Leonhardt, Clara Schumacher (Hrsg.)
DELFI 2020 –
Die 18. Fachtagung Bildungstechnologien der Gesellschaft für Informatik e.V.
14.–18. September 2020
Online

P-309 A. Meyer-Aurich, M. Gandorfer, C. Hoffmann, C. Weltzien, S. Bellingrath-Kimura, H. Floto (Hrsg.)
Informatik in der Land-, Forst- und Ernährungswirtschaft
Referate der 41. GIL-Jahrestagung
08.–09. März 2021, Leibniz-Institut für Agrartechnik und Bioökonomie e.V., Potsdam

P-310 Anne Koziolek, Ina Schaefer, Christoph Seidl (Hrsg.)
Software Engineering 2021
22.–26. Februar 2021,
Braunschweig/Virtuell

P-311 Kai-Uwe Sattler, Melanie Herschel, Wolfgang Lehner (Hrsg.)
Datenbanksysteme für Business, Technologie und Web (BTW 2021)
Tagungsband
13.–17. September 2021,
Dresden

P-312 Heiko Roßnagel, Christian H. Schunck, Sebastian Mödersheim (Hrsg.)
Open Identity Summit 2021
01.–02. Juni 2021, Copenhagen

P-313 Ludger Humbert (Hrsg.)
Informatik – Bildung von Lehrkräften in allen Phasen
19. GI-Fachtagung Informatik und Schule
8.–10. September 2021 Wuppertal

GI-Edition Lecture Notes in Informatics – Seminars

S-1 Johannes Magenheim, Sigrid Schubert (Eds.):
Informatics and Student Assessment
Concepts of Empirical Research and Standardisation of Measurement in the Area of Didactics of Informatics

S-2 Gesellschaft für Informatik (Hrsg.)
Informationstage 2005
Fachwissenschaftlicher Informatik-Kongress

S-3 Gesellschaft für Informatik (Hrsg.)
Informationstage 2006
Fachwissenschaftlicher Informatik-Kongress

S-4 Hans Hagen, Andreas Kerren, Peter Dannenmann (Eds.)
Visualization of Large and Unstructured Data Sets
First workshop of the DFG's International Research Training Group "Visualization of Large and Unstructured Data Sets – Applications in Geospatial Planning, Modeling and Engineering"

S-5 Gesellschaft für Informatik (Hrsg.)
Informationstage 2007
Fachwissenschaftlicher Informatik-Kongress

S-6 Gesellschaft für Informatik (Hrsg.)
Informationstage 2008
Fachwissenschaftlicher Informatik-Kongress

S-7 Hans Hagen, Martin Hering-Bertram, Christoph Garth (Eds.)
Visualization of Large and Unstructured Data Sets

S-8 Gesellschaft für Informatik (Hrsg.)
Informatiktage 2009
Fachwissenschaftlicher Informatik-Kongress

S-9 Gesellschaft für Informatik (Hrsg.)
Informatiktage 2010
Fachwissenschaftlicher Informatik-Kongress

S-10 Gesellschaft für Informatik (Hrsg.)
Informatiktage 2011
Fachwissenschaftlicher Informatik-Kongress

S-11 Gesellschaft für Informatik (Hrsg.)
Informatiktage 2012
Fachwissenschaftlicher Informatik-Kongress

S-12 Gesellschaft für Informatik (Hrsg.)
Informatiktage 2013
Fachwissenschaftlicher Informatik-Kongress

S-13 Gesellschaft für Informatik (Hrsg.)
Informatiktage 2014
Fachwissenschaftlicher Informatik-Kongress

S-14 Gesellschaft für Informatik (Hrsg.)
SKILL 2018
Studierendenkonferenz Informatik

S-15 Gesellschaft für Informatik (Hrsg.)
SKILL 2019
Studierendenkonferenz Informatik

S-16 Gesellschaft für Informatik (Hrsg.)
SKILL 2020
Studierendenkonferenz Informatik

S-17 Gesellschaft für Informatik (Hrsg.)
SKILL 2021
Studierendenkonferenz Informatik

The titles can be purchased at:

Köllen Druck + Verlag GmbH
Ernst-Robert-Curtius-Str. 14 · D-53117 Bonn
Fax: +49 (0)228/9898222
E-Mail: druckverlag@koellen.de

Gesellschaft für Informatik e.V. (GI)

publishes this series in order to make available to a broad public recent findings in informatics (i.e. computer science and information systems), to document conferences that are organized in cooperation with GI and to publish the annual GI Award dissertation.

Broken down into
- seminars
- proceedings
- dissertations
- thematics

current topics are dealt with from the vantage point of research and development, teaching and further training in theory and practice. The Editorial Committee uses an intensive review process in order to ensure high quality contributions.

The volumes are published in German or English.

Information: http://www.gi.de/service/publikationen/lni/

ISSN 1614-3213
ISBN 978-3-88579-751-7