

Adaptive Informationsvisualisierung: Der Einsatz von Machine-Learning zur Entscheidungsunterstützung

Franziska Marx ^{1, 2}

Abstract: In meiner Dissertation beschäftige ich mich mit der Frage, inwiefern die für die datengestützte Entscheidungsfindung relevanten Informationsvisualisierungen durch Machine-Learning-Algorithmen optimiert werden können. Ausgehend von der Problematik, dass es vielen Entscheidungsträgern schwerfällt, statistische Daten korrekt zu lesen, ist es notwendig, für die Erfüllung der Aufgabe relevante Informationen optimal zu visualisieren. Dabei muss sowohl das Wissen des Entscheidungsträgers als auch der Kontext der dargestellten Informationen berücksichtigt werden. Durch den Einsatz von Machine-Learning-Konzepten sollen die daraus gewonnenen Erkenntnisse sukzessive zur Verbesserung des Visualisierungssystems beitragen und den Entscheidungsträger gezielt unterstützen. Durch die Teilnahme am Doktoranden-Symposium erhoffe ich mir einen intensiven Austausch zur besseren Verankerung meiner interdisziplinären Fragestellung sowie die Möglichkeit der Weitergabe von bereits gewonnenen Erkenntnissen.

Keywords: Informationsvisualisierung, Decision Support Systems, Machine Learning, Statistical Literacy

1 Forschungsvorhaben

1.1 Motivation

Durch die im Zuge der *Datafication* vorhandenen Fülle an digitalen Daten, deren hohe Verfügbarkeit sowie die weltweite Vernetzung ist es heute ein Leichtes, Daten miteinander zu verknüpfen und in einem Decision Support System (DSS) einzusetzen. Diese Systeme haben zum Ziel, für die Entscheidungsfindung notwendige Daten effizient analysieren zu können und häufig durch geeignete Visualisierung zu präsentieren, sodass eine schnelle, flexible und einfache Einsicht in die Daten erlangt werden kann. Während Unternehmen Decision Support Systems zur Steigerung der Wettbewerbsfähigkeit einsetzen, soll beispielsweise durch den Einsatz in Krankenhäusern die Wartezeit für Patienten³ verkürzt werden oder in Schulen die individuelle Förderung einzelner Schüler ermöglicht werden. Die Einbettung in grafische Benutzungsoberflächen oder Web-Applikationen ermöglicht eine Interaktion mit den Daten, wodurch diese individuell selektierbar werden. Allerdings kann ein solches System nur soweit unterstützend wirken, wie die für den Benutzer sichtbaren Daten zugänglich und verständlich sind. Zur Interpretation und späteren Entschei-

¹ Institut für Informationsmanagement Bremen GmbH (ifib), Am Fallturm 1, 28359 Bremen, fmarx@ifib.de

² Universität Bremen, Bibliothekstraße 1, 28359 Bremen, franziska.marx@uni-bremen.de

³ Aus Gründen des besseren Leseflusses wird im vorliegenden Beitrag die weibliche Formen nicht explizit angeführt. An dieser Stelle wird jedoch ausdrücklich darauf hingewiesen, dass sich alle personenbezogenen Formulierungen grundsätzlich gleichermaßen auf Frauen und Männer beziehen.

dungsfindung ist es somit von Belang, inwiefern der Nutzer mit dem Kontext der Daten vertraut ist, aber auch, ob er diese interpretieren kann.

1.2 Fragestellung

In meiner Dissertation möchte ich mich der Frage widmen, inwiefern Machine-Learning-Algorithmen zur Optimierung von Decision Support Systems beitragen können, sodass die Darstellung von Informationen den Nutzer in seiner Entscheidung bestmöglich unterstützt. Dazu soll zunächst das Themengebiet "Entscheidungsfindung" aus verschiedenen Perspektiven und Disziplinen, wie Psychologie oder Betriebswirtschaftslehre, betrachtet werden, um herauszufinden, inwiefern die Messung der Qualität einer Entscheidung möglich ist und wie diese unterstützt werden kann. Daneben soll durch die Analyse von Machine-Learning-Konzepten die Frage beantwortet werden, welche Algorithmen für die Unterstützung von Entscheidungsfindung am Besten geeignet sind.

2 Stand der Forschung

Zur Beantwortung meiner Forschungsfrage erscheint es mir zunächst notwendig, das Themengebiet von Decision Support Systems näher zu beleuchten. Die Forschung in diesem Bereich hat in den letzten Jahren verstärkt die Benutzbarkeit und Verständlichkeit solcher Systeme in den Vordergrund gerückt, wodurch die Darstellungsform der Informationen diskutiert wurde. Auf der anderen Seite spielt der Einsatz von künstlichen Intelligenzen (*Artificial Intelligence*) und selbst lernenden Systemen (*Machine Learning*) eine immer größere Rolle, wodurch ein DSS komplexe Handlungen erfassen kann und aus vergangenen Entscheidungen automatisch lernt. Eine Herausforderung stellt hierbei die Fähigkeit des Entscheidungsträgers dar, statistische Daten zu lesen und auszuwerten, die so genannte *Statistical Literacy*. Forschungsbedarf sehe ich in der Verknüpfung dieser Themengebiete, da sich die bisherigen Forschungen entweder auf die Generierung von verständlichen Visualisierungen im Rahmen von DSS konzentrieren oder aber auf den Einsatz von Machine-Learning-Konzepten in diesem Feld, wobei die adaptive Gestaltung hier auf die Prozesse und nicht auf die Eigenschaften des Entscheidungsträger bezogen ist. In meiner Dissertation werde ich eine Brücke zwischen den einzelnen Themenbereichen schlagen und die *Statistical Literacy* des Nutzers sowie den Kontext der Entscheidung fokussieren, um daraus die bestmögliche grafische Darstellung zu gewinnen. Im Folgenden wird daher der aktuelle Stand der Forschung meiner thematischen Schwerpunkte kurz erläutert.

2.1 Decision Support Systems

Little beschreibt 1970 [Li70] mit dem Begriff des "Decision Calculus" ein "set of procedures for processing data and judgments to assist a manager in his decision making" und argumentiert, dass ein solches System erfolgreich ist, wenn es simpel, robust, einfach zu bedienen, adaptiv, in den wichtigsten Punkten vollständig und die Kommunikation mit

dem System einfach ist. 1971 etabliert Morton den Begriff des "Decision Support Systems", das mit Computern und analytischen Modellen Managern helfen soll, Probleme zu lösen [Mo71]. Keen betont, dass "a final system can be developed only through an adaptive process of learning and evolution" und fordert eine stärkere Verknüpfung von System, Benutzer und Entwickler, während der Nutzer den adaptiven Designprozess bestimmt [Ke80]. Um die Entscheidungsfindung zu erleichtern, wird zunehmend die Ausgestaltung der Datenpräsentation in Form von Visualisierungen untersucht (vgl. [Te99] [SHS02]). Yu et al. fokussieren vor allem die Möglichkeit der Interaktion des Nutzers mit den dargestellten Informationen [YDZ09], während Zhang die Visualisierung durch die Anwendung von abstrakten Regeln optimiert [Zh98]. Tsoukalas et al. stellen die Einbindung von Machine-Learning-Algorithmen zur Verbesserung von DSS vor [TAT15].

Diese Entwicklung zeigt, wie sich das Verständnis des Begriffs Decision Support System über die Jahre vom reinen Management-Tool für einzelne Prozesse zu einem vielschichtigen Werkzeug gewandelt hat, bei dem der Nutzer immer mehr in den Fokus des Systemdesigns rückt. Meine Arbeit soll diesen Trend aufgreifen und durch die Verknüpfung mit weiteren Techniken zugunsten der Benutzbarkeit und Verständlichkeit weiter entwickeln, sodass ein DSS entsteht, welches das Vorwissen und die Fähigkeiten des Benutzers berücksichtigt.

2.2 Machine Learning

Machine Learning, die künstliche Generierung von Wissen aus Erfahrung, findet seit Jahren verstärkt Anwendung in DSS, um dort bestimmte Muster zu erkennen und zu beurteilen (vgl. Turban, Part IV[TA05]). Die Grundlagenforschung des Machine Learning entwickelt dabei stets neue Algorithmen, die sich für die Lösung von Problemen unterschiedlich gut eignen. Auch in diesem Forschungsbereich wird der Nutzen für eine verbesserte Nutzerinteraktion in den letzten Jahren verstärkt erforscht, z. B. in Bezug auf die Verarbeitung natürlicher Sprache [ZL08] oder die Visualisierung von Daten [Ar15]. Eine Herausforderung stellt hierbei die Wahl des richtigen Algorithmus sowie dessen Integration in das DSS dar [MN06]. Durch die Kooperation mit Wissenschaftlern aus dem Bereich der Grundlagenforschung im Machine Learning möchte ich einen für mein Vorhaben optimalen Algorithmus auswählen und in das DSS integrieren.

2.3 Informationsvisualisierung

Die Visualisierung von Informationen zur besseren Verständlichkeit ist keineswegs eine Erfindung der modernen Gesellschaft, wie Landkarten aus der Steinzeit beweisen. Die menschliche Wahrnehmung in Bezug auf visuelle Informationen ist dabei gut erforscht (z. B. Few [Fe09], Miller [Mi56]), wird aber häufig nicht berücksichtigt. Tufte prangert an, dass bei der grafischen Darstellung von quantitativen Daten oftmals gestalterische Aspekte im Vordergrund stehen, welche die Korrektheit der Daten - teils unbeabsichtigt - verfälschen und die Aussage damit zur "Lüge" abwandeln [Tu01]. Er entwickelte daraufhin Grundregeln für die Visualisierung von Informationen, insbesondere von statisti-

schen Daten, die ich in meiner Dissertation mit den für das Lesen statistischer Daten notwendigen Kompetenzen verknüpfen möchte. Durch die Implementierung eines Machine-Learning-Systems soll dies sukzessive ausgebaut werden.

2.4 Statistical Literacy

Bereits 1951 machte Walker auf die gesellschaftliche Problematik aufmerksam, welche die Unfähigkeit statistische Daten zu lesen mit sich bringt [Wa51]. Die durch Sozialforscher oder Organisationen erhobenen Daten zur Erlangung von mehr Transparenz und Veranschaulichung von Sachverhalten wurden missverstanden oder durch deren Autoren falsch eingesetzt, sodass sie oftmals das Gegenteil bewirkten. Wallmann betont, dass Statistical Literacy keineswegs nur im wissenschaftlichen oder politischen Kontext relevant ist, sondern für alle Teile einer Gesellschaft im täglichen Leben eine bedeutende Rolle spielt [Wa93]. Die fehlende Statistical Literacy beschränkt sich dabei nicht etwa auf bildungsferne Bevölkerungsschichten, sondern ist auch in Berufsgruppen wie Ärzten [We15] oder Lehrern [PCG13] zu finden. Um dem entgegenzuwirken untersuchten Chick und Pierce die konkreten Kompetenzen, die eine Lehrkraft braucht, um Schul-Statistiken zu lesen und formulierten daraus Empfehlungen für gezielte Schulungen [CP13], während Carvalho und Solomon die Notwendigkeit der Individualisierung von Statistiken und die Berücksichtigung des Kontexts des Lesers herausstellen [CS12]. Ausgehend von dem Wissen über die Visualisierung von Informationen kann hier angesetzt werden, um fehlende Kompetenzen im Umgang mit Statistiken kontextabhängig zu unterstützen. Dabei ist noch zu untersuchen, wie die Statistical Literacy der Zielgruppe des zu entwickelnden DSS, also u.a. Hochschulleitungen oder Studieninteressierte, eingeschätzt werden kann.

3 Forschungsplan

Im Zuge meiner Dissertation werde ich zunächst auf Basis vorhandener Forschungsergebnisse die grundlegenden Anforderungen an das zu entwickelnde System herausarbeiten, die sich auf die zu unterstützenden Kompetenzen und die Darstellungsform der Informationen beziehen. Im zweiten Schritt werde ich die Besonderheiten benennen, die der gewählte Informationskontext mit sich bringt. Aktuell sieht mein Forschungsvorhaben vor, Hochschul-Ranking-Daten zu verwenden, da hier eine große Anzahl von Datensätzen vorhanden ist, zwischen denen interessante Wechselwirkungen auftreten und die das notwendige Maß an Komplexität aufweisen. Meine bisherigen Recherchen haben zudem ergeben, dass es im deutschsprachigen Raum keine Anwendung gibt, die diese Daten unter Berücksichtigung des Nutzers (z. B. ob Hochschul-Direktor oder Studienanfänger) zur Verfügung stellt, obwohl ein DSS im Hochschulbereich beispielsweise für die Studiengangswahl oder organisatorische Entscheidungen von Vorteil wäre. Anschließend werde ich verschiedene Konzepte des Machine Learning betrachten sowie deren wesentliche Gemeinsamkeiten und Unterschiede herausarbeiten. Daraus soll abgeleitet werden, welcher Algorithmus am besten für das geplante System geeignet ist. Im nächsten Schritt werde ich ein DSS auf Basis der vorherigen Erkenntnisse entwickeln.

Daneben muss eine Metrik ausgewählt oder angepasst werden, mit der die Qualität der Entscheidung bestimmt werden kann. Hierbei ist insbesondere eine Analyse der aus Psychologie oder BWL bereits bestehenden Metriken erforderlich.

Abschließend soll das entwickelte System mittels vergleichender Nutzertests in zwei Schritten evaluiert werden. Im ersten Schritt werde ich die Statistical Literacy von Testnutzern sowie deren Hintergrundwissen in einem bestimmten Bereich durch einen Fragebogen erheben. Anschließend werden jeder Testperson Visualisierungen von Informationen gezeigt, auf Basis derer sie Entscheidungen treffen muss. Die Ergebnisse aus diesem Schritt werden als Trainingsdaten in das Machine Learning System übergeben. Im zweiten Schritt werden eine weitere Testgruppe sowie eine Kontrollgruppe bestimmt. Beide Gruppen durchlaufen denselben Testzyklus wie in Schritt 1, wobei die Testgruppe angepasste Visualisierungen auf Basis der Erkenntnisse aus Schritt 1 gezeigt bekommt, während die Kontrollgruppe auf Basis der unangepassten Visualisierungen Entscheidungen treffen soll. Die Ergebnisse der beiden Gruppen werden anschließend gegenüber gestellt.

4 Stand der Arbeit

Seit Anfang dieses Jahres beschäftige ich mich mit der Konkretisierung meines Dissertationsthemas. Zu diesem Zweck habe ich begonnen, die Schnittmengen und Abgrenzungen der vorhergehend aufgeführten Themen herauszuarbeiten. Hier gilt es besonders zu berücksichtigen, dass eine Entscheidung meist abhängig von dem Themengebiet getroffen wird, in dem sie angesiedelt ist. Daraus ergibt sich die Akquise von Daten aus einem konkreten Anwendungsfeld, aus dem Trainingsdaten für die initiale Eingabe in das Machine-Learning-System gewonnen werden können. Daneben möchte ich eine Kooperation mit Wissenschaftlern erreichen, die sich mit der Grundlagenforschung von Machine-Learning-Algorithmen beschäftigen und deren Vorgehen für die Auswahl eines bestimmten Algorithmus näher beleuchten. Von dem Doktoranden-Symposium der INFORMATIK 2016 erhoffe ich mir einen intensiven Austausch mit Doktoranden, deren Forschungsvorhaben in einem oder mehreren der oben genannten Bereiche angesiedelt ist sowie die Exploration neuer Schnittmengen mit Themengebieten, denen ich bisher weniger Beachtung geschenkt habe. Zudem würde ich meine geplante methodische Vorgehensweise und bisherigen Erfahrungen gerne mit weiteren jungen Wissenschaftlern teilen und zur Diskussion stellen, da dies in der interdisziplinären Forschung eine Besondere Herausforderung darstellt.

Literaturverzeichnis

- [Ar15] Archambault, Daniel; Bunte, Kerstin; Carreira-Perpiñán, Miguel; Ebert, David; Ertl, Thomas; Zupan, Blaz: Machine Learning Meets Visualization: A Roadmap for Scalable Data Analytics. In: Bridging Information Visualization with Machine Learning. Dagstuhl, S. 7–12, 2015.
- [CP13] Chick, Helen; Pierce, Robyn: The Statistical Literacy Needed to Interpret School Assessment Data. Mathematics Teacher Education and Development, 15(2), 2013.
- [CS12] Carvalho, Carolina; Solomon, Yvette: Supporting statistical literacy: What do culturally relevant/realistic tasks show us about the nature of pupil engagement with statistics? International Journal of Educational Research, 55:57–65, 2012.

- [Fe09] Few, Stephen: Now you see it: simple visualization techniques for quantitative analysis. Analytics Press, Oakland, 2009.
- [Ke80] Keen, Peter G. W.: Adaptive Design for Decision Support Systems. ACM SIGOA Newsletter, 1(4-5):15–25, September 1980.
- [Li70] Little, John D. C.: Models and Managers: The Concept of a Decision Calculus. The Institute of Management Sciences, 16(8):466 – 485, April 1970.
- [Mi56] Miller, George A.: The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychological Review, 63(2):81–97, 1956.
- [MN06] Maniyar, Dharmesh M.; Nabney, Ian T.: Visual Data Mining: Integrating Machine Learning with Information Visualization. ACM, Philadelphia, 2006.
- [Mo71] Morton, Michael S. Scott: Management Decision Systems: Computer-Based Support of Decision Making. Harvard University Press, Cambridge, 1971.
- [PCG13] Pierce, Robyn; Chick, Helen; Gordon, Ian: Teachers’ perceptions of the factors influencing their engagement with statistical reports on student achievement data. Australian Journal of Education, 57(3):237–255, November 2013.
- [SHS02] Shen-Hsieh, Angela; Schindl, Mark: Data Visualization for Strategic Decision Making. In: Case Studies of the CHI2002, AIGA Experience Design FORUM. CHI ’02, ACM, New York, S. 1–17, 2002.
- [TA05] Turban, Efraim; Aronson, Jay E.: Decision Support Systems and Intelligent Systems. Pearson/Prentice Hall, 2005.
- [TAT15] Tsoukalas, Athanasios; Albertson, Timothy; Tagkopoulos, Ilias: From Data to Optimal Decision Making: A Data-Driven, Probabilistic Machine Learning Approach to Decision Support for Patients With Sepsis. JMIR Medical Informatics, 3(1):e11, Februar 2015.
- [Te99] Tegarden, David P.: Business Information Visualization. Communications of the Association for Information Systems, 1(1):2–38, Januar 1999.
- [Tu01] Tufte, Edward R.: The visual display of quantitative information. Graphics Press, Cheshire, 2nd ed. Auflage, 2001.
- [Wa51] Walker, Helen M.: Statistical Literacy in the Social Sciences. The American Statistician, 5(1):6–12, 1951.
- [Wa93] Wallman, Katherine K.: Enhancing Statistical Literacy: Enriching Our Society. Journal of the American Statistical Association, 88(421):1–8, 1993.
- [We15] Wegwarth, Odette: Statistical literacy in medicine. Dissertation, Freie Universität Berlin, Freie Universität Berlin, Germany, 2015.
- [YDZ09] Yu, Shidong; Deng, Liguu; Zhang, Yong: Visualization User Interface for Decision Support Systems. In: Ninth International Conference on Hybrid Intelligent Systems, 2009. HIS ’09. Jgg. 1, Shenyang, S. 63–66, August 2009.
- [Zh98] Zhang, Ping: An image construction method for visualizing managerial data. Decision Support Systems, 23(4):371–387, Oktober 1998.
- [ZL08] Zhao, Li; Li, F.: Statistical Machine Learning in Natural Language Understanding: Object Constraint Language Translator for Business Process. In: IEEE International Symposium on Knowledge Acquisition and Modeling Workshop, 2008. KAM Workshop 2008. S. 1056–1059, Dezember 2008.