

Argumentative Schreibunterstützung durch maschinelle Sprachverarbeitung¹

Christian Stab²

Abstract: Die automatische Analyse von geschriebenen Argumenten ist wegen der hohen Ambiguität und Vagheit von natürlicher Sprache eine große Herausforderung. Zum einen erfordert das Trainieren von maschinellen Lernverfahren hochqualitative Korpora, und zum anderen besteht die automatische Erkennung von Argumentationsstrukturen aus mehreren voneinander abhängigen Analyseschritten. In dieser Dissertation wird die Anwendbarkeit von theoretischen Argumentationsmodellen auf argumentative Aufsätze erstmals untersucht und ein Korpus zur ganzheitlichen Erkennung von Argumentationsstrukturen erstellt. Es wird ein neues Modell zur Erkennung von Argumentationsstrukturen vorgestellt, das die Funktion von Argumentkomponenten und argumentativen Relationen global optimiert und die Erkennungsraten gegenüber lokalen Modellen verbessert. Zusätzlich wird ein Ansatz zur automatischen Qualitätsbewertung vorgestellt, mit dem unzureichend begründete Argumente in argumentativen Aufsätzen mit hoher Präzision erkannt werden.

1 Einleitung

Das Schreiben von argumentativen Aufsätzen ist eine effektive Methode, Argumentationsfähigkeiten zu lehren. Durch die Analyse kontroverser Meinungen lernen Schüler_innen, Argumente zur Begründung ihres eigenen Standpunkts zu formulieren, Gegenargumente zu berücksichtigen und logische Fehler zu vermeiden. Wegen des enormen Korrekturaufwands können Lehrkräfte jedoch nur eine geringe Anzahl an argumentativen Schreibaufgaben vergeben, die in der Praxis nicht ausreichen, um Schüler_innen ausreichend auszubilden [BB11]. Computergestützte Assistenz- oder Bewertungssysteme [SB13] liefern zwar Rückmeldungen zu Grammatik, Diskursstrukturen, und lexikalischem Umfang, sind aber noch nicht in der Lage natürlichsprachliche Argumente automatisch zu erkennen und zu bewerten. „Argument Mining“ – ein junges Forschungsfeld der automatischen Sprachverarbeitung – hat das Potential diese Lücke zu schließen und neue intelligente Schreibhilfen zu ermöglichen, die automatisch konstruktive Rückmeldungen zu natürlichsprachlichen Argumenten generieren.

Die automatische Erkennung und Bewertung von natürlichsprachlichen Argumenten unterliegt jedoch den folgenden Herausforderungen: (1) Methoden zur automatischen Analyse von Argumenten basieren – wie viele andere Verfahren der automatischen Sprachverarbeitung – auf überwachten maschinellen Lernmethoden. Diese lernen die Erkennung und Bewertung von Argumenten anhand von bekannten Argumenten in manuell annotierten Korpora. Wegen der Ambiguität und Vagheit von Texten ist eine eindeutige Interpretation

¹ Englischer Titel der Dissertation: „Argumentative Writing Support by means of Natural Language Processing“

² Technische Universität Darmstadt, Ubiquitous Knowledge Processing Lab (UKP-TUDA), Department of Computer Science, Hochschulstraße 10, 64289 Darmstadt, <nachname>@ukp.informatik.tu-darmstadt.de

von natürlichsprachlichen Argumenten selbst für Menschen eine schwierige Aufgabe. Zu Beginn dieser Arbeit war es weitestgehend unbekannt, ob existierende theoretische Argumentationsmodelle von menschlichen Annotatoren mit ausreichender Übereinstimmung auf argumentative Aufsätze angewandt werden können, um qualitativ hochwertige Trainingsdaten zu erstellen. (2) Die meisten existierenden Methoden zur automatischen Argumentextraktion adressieren spezifische Teilaufgaben wie die Identifikation von Argumentkomponenten, die Klassifikation der argumentativen Funktion von Argumentkomponenten (bspw. als Behauptung oder Prämisse), oder die Erkennung von argumentativen Relationen. Diese Aufgaben sind jedoch nicht unabhängig voneinander. Beispielsweise lässt sich die Funktion einer Argumentkomponente erst bestimmen, wenn die argumentative Diskursstruktur bekannt ist und umgekehrt. Daher ist die unabhängige Modellierung dieser Teilaufgaben nicht ausreichend, um konsistente Argumentationsstrukturen zu erkennen. (3) Die Bewertung von Argumenten ist eine höchst subjektive Aufgabe. Die Qualität eines Arguments ist das Produkt einer Vielzahl von Kriterien, die nicht nur von persönlichen Präferenzen sondern auch vom Vorwissen eines Individuums abhängig sind [Th73]. Zum Beispiel unterliegt die Qualität eines Arguments dem Grad des Vertrauens in den Argumentierenden (Ethos), den durch das Argument angesprochenen Emotionen (Pathos), der logischen Korrektheit der Argumentation (Logos) und der Situation in der das Argument geäußert wurde (Kairos) [SN13]. Die hohe Subjektivität stellt eine wesentliche Herausforderung für die Entwicklung von automatischen Methoden zur Qualitätsbewertung von Argumenten dar. Basierend auf diesen Herausforderungen adressiert diese Dissertation die folgenden drei Forschungsfragen (engl. „research questions“):

RQ1 Annotation von Argumentationsstrukturen: Es sollte untersucht werden, ob sich theoretische Argumentationsmodelle auf studentische Aufsätze anwenden lassen. In diesem Zusammenhang wollen wir evaluieren, ob und in welchem Ausmaß menschliche Annotatoren bei der Erkennung von Argumentationsstrukturen übereinstimmen und ob es möglich ist, Trainingsdaten von hoher Qualität zu erstellen.

RQ2 Automatische Erkennung von Argumentationsstrukturen: Die zweite Forschungsfrage adressierte die automatische Erkennung von Argumentationsstrukturen. Es sollte untersucht werden, welche linguistischen Eigenschaften für die Teilschritte der Argumenterkennung effektiv sind und ob die gemeinsame Modellierung der Analyseschritte die Genauigkeit verbessert.

RQ3 Automatische Qualitätsbewertung von Argumenten: Mit dieser Forschungsfrage sollte beantwortet werden, welche Qualitätskriterien verwendet werden können, um objektives Feedback zu generieren, ob diese Kriterien auf echte Argumente anwendbar sind, und mit welcher Genauigkeit sie automatisch bewertet werden können.

Mit der Beantwortung dieser Forschungsfragen trägt diese Dissertation zu einem besseren Verständnis von natürlichsprachlichen Argumenten bei. Zum einen ermöglicht die Untersuchung der Anwendbarkeit von theoretischen (meist normativen) Argumentationsmodellen auf studentische Aufsätze eine empirische Beurteilung der Machbarkeit von automatischen Methoden zur Argumentanalyse. Andererseits leistet diese Dissertation methodische Beiträge zur automatischen Erkennung und Bewertung von Argumenten in Texten,

welche die Entwicklung von neuen Basistechnologien für innovative Suchverfahren zur Entscheidungsunterstützung ermöglichen.

2 Annotation von Argumentationsstrukturen

Die erste Herausforderung bestand in der Definition eines geeigneten Annotationsschemas. Dazu wurden in einem ersten Schritt theoretische Ansätze aus Philosophie und der (informellen) Logik untersucht.³ Um einen möglichst hohen Detailgrad der automatischen Analyse zu ermöglichen, lag der Schwerpunkt hierbei auf *monologischen Argumentationsmodellen*, die die Mikrostruktur und Komponenten von Argumenten formalisieren [BMB10]. Nach diesen Modellen besteht ein Argument aus mehreren Komponenten [Go10]: Die *Behauptung* ist eine kontroverse Aussage und die zentrale Komponente eines Arguments. Die *Prämissen* sind Aussagen zur Begründung oder Widerlegung der Behauptung. Die *Konsequenzrelation* verbindet die Prämisse(n) mit der Behauptung und legt die Art der Schlussfolgerung fest.

Der Vergleich existierender Argumentationsmodelle zeigte, dass *Argumentdiagramme* wegen ihrer Ähnlichkeit zu existierenden Diskursanalyseansätzen am Besten geeignet sind. Argumentdiagramme sind eine Methode aus der informellen Logik mit der unstrukturierte Argumente in natürlicher Sprache für die weitere Analyse in eine strukturierte Repräsentation überführt werden [Go10]. Ein Argumentdiagramm besteht aus Knoten und gerichteten Kanten. Jeder Knoten entspricht einer Komponente des Arguments (eine Aussage in natürlicher Sprache). Eine Kante stellt eine gerichtete argumentative Beziehung von der Ursprungskomponente zur Zielkomponente dar (bspw. „begründet“ oder „widerlegt“). Im Gegensatz zu anderen normativen Argumentmodellen, wie beispielsweise dem Toulmin Modell [To58] oder den Argumentationsschemata von Walton [WRM08], können mit Argumentdiagrammen auch komplexe Argumentationsketten und geschachtelte Widerlegungen modelliert werden. Durch die explizite Modellierung von argumentativen Relationen können zudem mehrere Argumente in einem Text leicht voneinander getrennt werden.

2.1 Annotationsschema

Das vorgeschlagene Annotationsschema modelliert die Argumentstruktur eines Textes als Baum, wobei der Wurzelknoten die argumentative *Kernaussage* (engl. „major claim“) enthält. Die eigentlichen Argumente werden mit Knoten vom Typ Behauptung und Prämisse modelliert. Gerichtete argumentative Relationen vom Typ begründet und widerlegt modellieren die Zusammenhänge zwischen den Argumentkomponenten. Der folgende Auszug eines Aufsatzes zum Thema Auslandssemester illustriert das vorgeschlagene Annotationsschema (Kernaussagen sind fett gedruckt, Behauptungen unterstrichen, und Prämissen gewellt unterstrichen):

Beispiel 1: „Studieren im Ausland ist ein häufig diskutiertes Thema. Ich denke, jeder sollte während des Studiums ein Semester im Ausland verbringen“_{Kernaussage}. [Ein Auslandssemester

³ vgl. Kapitel 2 in [St17].

ist eine wertvolle Erfahrung]Behauptung- [*Das Leben an einem fremden Ort wird natürlich erstmal schwierig sein*]Prämisse1, aber [*diese Schwierigkeiten werden später zur wertvollen Erfahrung*]Prämisse2. [*Man lernt außerdem auf eigenen Füßen zu stehen und nicht von anderen abhängig zu sein*]Prämisse3.“

Der erste Satz leitet das Thema ein und enthält keinen argumentativen Inhalt. Der zweite Satz enthält die Kernaussage und die Meinung des Autors zum Thema Auslandssemester. Der dritte Satz ist die Behauptung des ersten Arguments zur Untermauerung der Kernaussage. Der vierte Satz enthält zwei Prämissen von denen Prämisse₁ die Behauptung und Prämisse₂ Prämisse₁ widerlegt. Der letzte Satz enthält eine weitere Prämisse welche die Begründung direkt belegt. Abbildung 1 zeigt die Argumentationsstruktur des Beispiels.

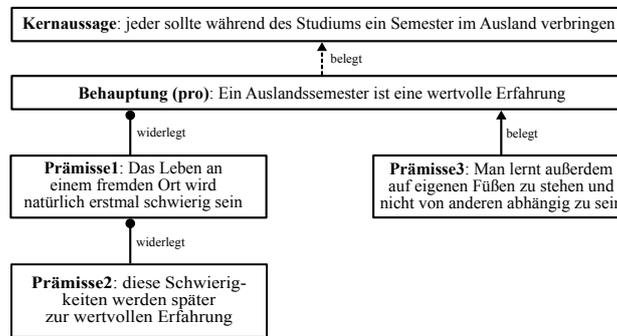


Abb. 1: Argumentationsstruktur von Beispiel 1.

2.2 Annotationsstudie und Erstellung der Trainingsdaten

Um die Anwendbarkeit des Annotationsschemas zu evaluieren wurde eine Annotationsstudie durchgeführt. Das Korpus für diese Studie ist eine Sammlung von 402 englischen Aufsätzen, die einem Onlineforum entnommen wurden.⁴ Die Aufsätze wurden zufällig aus dem Bereich „Schreibfeedback“ entnommen, in dem Schüler_innen Rückmeldungen zu ihren Aufsätzen anfordern. Die Aufsätze wurden manuell überprüft, um einen argumentativen Schreibstil zu gewährleisten. Insgesamt enthält das Korpus 7,116 Sätze mit 147,271 Wörtern.

Um eine hohe Datenqualität zu gewährleisten, wurde eine 31 Seiten umfassende Annotationsrichtlinie erarbeitet. Basierend auf dieser Richtlinie, annotierten drei Wissenschaftler aus dem Bereich der Sprachtechnologie unabhängig voneinander eine zufällige Auswahl von 80 Aufsätzen. Die Übereinstimmung zwischen den Annotatoren wurde mit Interrater-Reliabilität-Metriken systematisch evaluiert. Die Ergebnisse zeigten eine substantielle Übereinstimmung bei der Annotation von Argumentkomponenten. Unter Berücksichtigung der Komponententypen und deren Grenzen im Text erreichten die Annotatoren eine Übereinstimmung von $\alpha_U = .767$.⁵ Die gemessene Interrater-Reliabilität der

⁴ www.essaysforum.com

⁵ α_U ist ein Interrater-Reliabilität-Metrik die sowohl die Kategorie (Kernaussage, Behauptung und Prämisse) als auch die Grenzen der markierten Textabschnitte auf Wortebene berücksichtigt.

argumentativen Relationen beträgt $\kappa = .708$ für den Relationstyp *belegt* und $\kappa = .737$ für den Relationstyp *widerlegt*. Diese Ergebnisse zeigen, dass das vorgeschlagene Annotationschema mit hoher Übereinstimmung auf argumentative Aufsätze anwendbar ist und es möglich ist Trainingsdaten mit hoher Qualität zu erstellen. Ausgehend von diesen Ergebnissen wurde das komplette Korpus annotiert, womit der wissenschaftlichen Gemeinschaft erstmals Trainingsdaten für die ganzheitliche Erkennung von Argumentstrukturen in argumentative Aufsätzen zur Verfügung stand [SG14a, SG17a].

3 Automatische Erkennung von Argumentationsstrukturen

Die automatische Erkennung von Argumentationsstrukturen besteht aus mehreren Teilaufgaben. Diese umfassen die Trennung von argumentativen und nicht-argumentativen Textabschnitten und die Erkennung der Komponentengrenzen auf Wortebene (Segmentierung), die Klassifikation der argumentativen Rolle einer Komponente (z.B. als Behauptung oder Prämisse), und die Verlinkung der Argumentkomponenten mit argumentativen Relationen. Die Klassifikation von Argumentkomponenten und die Verlinkung von Argumentkomponenten sind jedoch nicht unabhängig voneinander [SG14b]. Insbesondere lässt sich die Funktion einer Argumentkomponente nur unter Berücksichtigung der argumentativen Diskursstruktur eines Textes bestimmen und umgekehrt. Um diese Herausforderung zu lösen und die wechselseitigen Informationen zu nutzen, wurde ein „Joint-Model“ entwickelt, das die Komponententypen und argumentativen Relationen global optimiert.

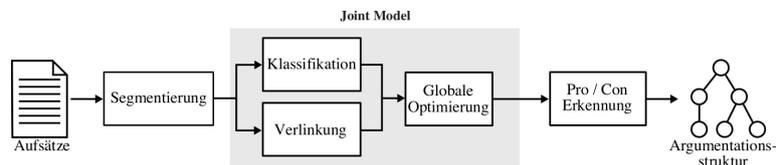


Abb. 2: Komponenten des Modells für die automatische Erkennung von Argumentationsstrukturen.

Abbildung 2 zeigt die Architektur des Modells mit den folgenden Komponenten:

1. *Segmentierung*: Die Segmentierung trennt argumentative von nicht-argumentativen Textabschnitten und erkennt die Grenzen von Argumentkomponenten auf Wortebene mit einem IOB-Schema⁶ und einem *Conditional Random Field*.
2. *Joint-Model*: Das Joint-Model baut auf zwei lokalen Klassifikatoren auf, um die Funktion von Argumentkomponenten zu bestimmen und sie mit gerichteten argumentativen Relationen zu verknüpfen. Beide Modelle basieren auf einer *Support Vector Machine* (SVM). Die Ergebnisse beider Klassifikatoren werden anschließend global mit *Integer Linear Programming* (ILP) optimiert, um einen oder mehrere Bäume in jedem Paragraphen des Textes zu finden.

⁶ Jedes Wort wird entweder als Beginn (B), innerhalb (I) oder außerhalb (O) einer Argumentkomponente gekennzeichnet.

3. *Pro/Con Erkennung*: Dieses Modell klassifiziert mit einer SVM jede Behauptung und Prämisse als “widerlegt” oder “belegt”, um zwischen verschiedenen argumentativen Relationen zu unterscheiden.

Für jedes dieser Modelle wurden spezifische linguistische Eigenschaften definiert, um optimale Ergebnisse zu erzielen. Die Auswahl dieser Eigenschaften wurde für jedes Modell systematisch durch eine Kreuzvalidierung auf den Trainingsdaten durchgeführt. Die Ergebnisse zeigten beispielsweise, dass strukturelle Eigenschaften (bspw. die Position eines Satzes) effektiv für die Trennung von argumentativen und nicht-argumentativen Textabschnitten genutzt werden können und lexikalisch-syntaktische Informationen für die Erkennung des ersten Wortes einer Argumentkomponente effektiv sind. Die genauen Details zu dieser Untersuchung sind in Kapitel 5 der Dissertation beschrieben [St17].

Die Ergebnisse der lokalen Klassifikatoren werden mit folgendem ILP-Modell global optimiert, um eine konsistente Baumstruktur zu erhalten. Für n gegebene Argumentkomponenten ist die Zielfunktion des Modells definiert als:

$$\operatorname{argmax}_{x,b} \sum_{i=1}^n \sum_{j=1}^n w_{ij} x_{ij} + b_{ij}, \quad (1)$$

wobei die Variablen $x_{ij} \in \{0, 1\}$ eine argumentative Relation zwischen den Argumentkomponenten i und j , und $b_{ij} \in \{0, 1\}$ einen direkten Pfad von i nach j kennzeichnen.⁷ Die Koeffizienten $w_{ij} \in \mathbb{R}$ sind Gewichte für jede Relation, die durch die Ergebnisse der lokalen Klassifikatoren bestimmt werden (siehe unten). Um eine Baumstruktur zu gewährleisten wird die Zielfunktion unter Berücksichtigung folgender Bedingungen gelöst.⁸ Die erste Bedingung $\forall i : \sum_{j=1}^n x_{ij} \leq 1$ verhindert, dass eine Komponente mehr als eine ausgehende Kante besitzt. Um sicherzustellen dass mindestens eine Argumentkomponente ohne ausgehende Relation existiert (Wurzelknoten), wird die Anzahl aller Relationen durch die Bedingung $\sum_{i=1}^n \sum_{j=1}^n x_{ij} \leq n - 1$ beschränkt. Die dritte Bedingung $\forall i : x_{ii} = 0$ verhindert dass eine Relation denselben Start- und Endknoten hat. Zusätzlich verhindern die drei folgenden Bedingungen (übernommen von [Kü08, p. 92]) Zyklen in der finalen Struktur. Die Bedingung $\forall i \forall j : x_{ij} - b_{ij} \leq 0$ koppelt die Hilfsvariablen b_{ij} mit den Variablen x_{ij} und gewährleistet, dass der Pfad zwischen den Argumentkomponenten i und j in b_{ij} gesetzt ist, wenn es eine direkte Relation zwischen beiden Argumentkomponenten gibt. Mit der Bedingung $\forall i \forall j \forall k : b_{ik} - b_{ij} - b_{jk} \leq -1$ werden alle Pfade deren Länge größer als eins sind an die Hilfsvariablen gekoppelt. Durch die letzte Bedingung $\forall i : b_{ii} = 0$ werden alle Pfade die in derselben Argumentkomponente beginnen und enden – und somit Zyklen – vermieden.

Um die Ergebnisse der beiden lokalen Klassifikatoren in die Zielfunktion zu integrieren, werden die Gewichte für jede Relation als $w_{ij} = \phi_r r_{ij} + \phi_c c_{ij} + \phi_{cr} cr_{ij}$ definiert. Hierbei ist $r_{ij} = 1$, wenn eine argumentative Relation zwischen den Argumentkomponenten i und j gefunden wurde und 0 andernfalls. Die Variable c_{ij} ist 1 falls Komponente j als

⁷ Jede Variable wird beim Lösen der Funktion als binär betrachtet.

⁸ Zum Lösen verwenden wir das Framework lpsolve: <http://lpsolve.sourceforge.net>

Behauptung klassifiziert wurde, sonst ist $c_{ij} = 0$. Die dritte Komponente $cr_{ij} = cs_j - cs_i$ weist Relationen von Behauptungen zu Prämissen einen höheren Wert zu. Der Wert $cs_i = \frac{rel_{in_i} - rel_{out_i} + n - 1}{rel + n - 1}$ ist die Wahrscheinlichkeit, dass Argumentkomponente i eine Behauptung ist und wird durch die lokal gefundenen Relationen bestimmt. Dabei ist $rel_{in_i} = \sum_{k=1}^n r_{ki}[i \neq k]$ die Anzahl der eingehenden Relationen von i , $rel_{out_i} = \sum_{l=1}^n r_{il}[i \neq l]$ die Anzahl der ausgehenden Relationen von i , und $rel = \sum_{k=1}^n \sum_{l=1}^n r_{kl}[k \neq l]$ die Anzahl der gefundenen Relationen. Der Wert von cs_i ist für eine Argumentkomponente mit vielen eingehenden Relationen größer als für eine Argumentkomponente mit weniger eingehenden Relationen. Jedes ϕ ist ein Parameter des ILP-Modells, um den Anteil der drei Informationen r_{ij} , c_{ij} , and cr_{ij} zu steuern. Diese Parameter wurden anhand von Kreuzvalidierung bestimmt ($\phi_r = \frac{1}{2}$ und $\phi_{cr} = \phi_c = \frac{1}{4}$). Nach der Anwendung des ILP-Modells, werden die Relationen und Komponententypen entsprechend der resultierenden x_{ij} gesetzt. Jede Komponente ohne ausgehende Kante wird als Behauptung und alle weiteren als Prämissen markiert.

Um das vorgeschlagene Modell zu evaluieren, wurde das annotierte Korpus in Trainings- (80%) und Testdaten (20%) aufgeteilt. Alle Parameter des Modells wurden durch systematische Kreuzvalidierung auf den Trainingsdaten bestimmt bevor die Modelle auf den Testdaten mit den folgenden zwei Baselines verglichen wurden: Die erste Baseline klassifiziert jede Instanz mit der häufigsten Klasse in den Daten (Mehrheits-Baseline). Die zweite Baseline (Heuristische Baseline) ist eine regelbasiertes System, welches anhand von Schreibrichtlinien für argumentative Aufsätze definiert wurde.⁹ Tabelle 1 zeigt die

	Segmentierung	Klassifikation	Verlinkung	Pro/Con	\emptyset F1
Mehrheits-Baseline	25,9	26,0	45,5	47,8	36,3
Heuristische Baseline	64,2	75,9	70,0	56,2	66,6
Lokale Klassifikatoren	† 86,7	79,4	†71,7	† 68,0	76,5
ILP Joint Model	-	†‡ 82,6	† 75,1	-	78,1

Tab. 1: Evaluationsergebnisse (Makro-F1-Metrik) der Argumenterkennung († = signifikante Verbesserung über der heuristischen Baseline; ‡ = signifikante Verbesserung über den lokalen Klassifikatoren; Wilcoxon-Test mit $\alpha = .005$).

Ergebnisse der Evaluation auf den Testdaten. Wie die Ergebnisse zeigen, übertrifft das Segmentierungsmodell die heuristische Baseline signifikant ($p = 1,65 \times 10^{-14}$). Die Evaluation des ILP-Modells zeigt, dass es erfolgreich die Ergebnisse der lokalen Klassifikation von Argumentkomponenten verbessert ($p = 7,45 \times 10^{-4}$) ohne die Ergebnisse des Modells zur Verlinkung negativ zu beeinflussen. Es erzielt 82,6 F1 für die Klassifikation von Argumentkomponenten und 75,1 F1 für die Erkennung von argumentativen Relationen. Eine genauere Untersuchung zeigt, dass das Modell die Erkennung von Behauptungen um 7,1 F1 und auch die Erkennung von verlinkten Argumentkomponenten um 7,7 F1 gegenüber den lokalen Modellen verbessert. Das Modell zur Unterscheidung von belegenden und widerlegenden Relationen verbessert die Klassifikationsgenauigkeit der heuristischen Baseline um 11,8 F1 ($p = 0.008$). Insgesamt zeigen die Ergebnisse, dass das ILP-Modell gleichzeitig die Klassifikation und die Verlinkung von Argumentkomponenten verbessert und

⁹ Die Details zu dieser Baseline sind in [SG17a] beschrieben.

somit eine vielversprechende Grundlage für die Umsetzung von argumentativen Schreibsystemen bietet.

4 Automatische Qualitätsbewertung von Argumenten

Für die automatische Evaluation der Argumentqualität wurde ein Ansatz basierend auf dem Hinlänglichkeitskriterium von [JB77] entwickelt (engl. „sufficiency criterion“).¹⁰ Ein Argument erfüllt dieses Kriterium, wenn dessen Prämissen ausreichen, um die Behauptung zu belegen. Das folgende Argument zum Thema Auslandsstudium veranschaulicht eine Verletzung des Hinlänglichkeitskriterium:

Beispiel 2): *„Einer meiner Freunde studierte Informatik an der Universität London und hat heute eine gut bezahlte Anstellung bei Google. Das zeigt, dass Studenten die im Ausland studierten besser bezahlt werden.“*

Die Prämisse dieses Beispiels bezieht sich auf ein einziges Beispiel (erster Satz), das die generelle Behauptung des Arguments belegen soll. Ein einziges Beispiel ist jedoch nicht ausreichend, um den generellen Fall zu bestätigen. Diese Art von Argument ist auch als voreilige Schlussfolgerung (engl. „hasty generalization fallacy“) bekannt [Go10].

Beispiel 3): *„Laut einer Umfrage verdienen Studenten die im Ausland studierten durchschnittlich 25% mehr pro Jahr. Daher könnte sich ein Auslandsstudium positiv auf das spätere Gehalt auswirken.“*

Die Behauptung in Beispiel 3 ist ausreichend begründet. Ausgehend von den Studienergebnissen kann man davon ausgehen, dass sich ein Auslandsstudium positiv auf das Gehalt auswirken könnte.

Um die Anwendbarkeit des Kriteriums für natürlichsprachliche Argumente zu evaluieren, annotierten drei Annotatoren 433 Argumente des in Abschnitt 2 beschriebenen Korpus. Die Ergebnisse zeigten, dass die Entscheidung der Annotatoren für 91,07% der Argumente übereinstimmte. Die hohen Werte der Interrater-Reliabilität-Metriken von $\kappa = .7672$ und $\alpha = .7672$ bestätigen die substantielle Übereinstimmung. Ausgehend von diesen Ergebnissen wurden alle 1029 Argumente des Korpus annotiert. Insgesamt sind 33,8% dieser Argumente unzureichend begründet.

Zur automatischen Erkennung von unzureichend begründeten Argumenten wurden mehrere Ansätze in einer wiederholten Kreuzvalidierung evaluiert.¹¹ Das erste Modell ist eine SVM, welches anhand manuell definierter Eigenschaften ein gegebenes Argument als „ausreichend“ oder „unzureichend“ klassifiziert [SG17b]. Das zweite Modell ist ein *Convolutional Neural Network* (CNN) mit zuvor trainierten Word Embeddings. Tabelle 2 zeigt die Ergebnisse und den Vergleich beider Systeme mit einer Mehrheits-Baseline und einer SVM mit binären lexikalische Eigenschaften (SVM-Bow). Die SVM mit manuell definierten Eigenschaften verbessert die Ergebnisse der beiden Baselines signifikant. Es erreicht

¹⁰ Kapitel 2 der Dissertation enthält einen detaillierten Vergleich von Qualitätskriterien aus Logik und Philosophie.

¹¹ Insgesamt wurden pro Modell hundert Experimente durchgeführt (5-fold CV mit 20 Wiederholungen).

	Genauigkeit	F1	F1 Unzureichend	Precision	Recall
Mehrheits-Baseline	66,2	39,8	0	0	0
SVM-bow Baseline†	78,5	75,5	66,1	70,9	62,4
SVM †‡	79,8	77,0	68,1	73,1	64,1
CNN †‡	84,3	82,7	77,0	76,2	78,4

Tab. 2: Evaluationsergebnisse der Qualitätsbewertung († = signifikante Verbesserung über der mehrheitlichen Baseline; ‡ signifikante Verbesserung über SVM-bow; Wilcoxon-Test mit $\alpha = 0.05$)

eine Genauigkeit von 79,8% und 77,0 F1. Die besten Ergebnisse erzielt das neuronale Netz. Er erreicht 82,7 F1 und eine Genauigkeit von 84,3%. Die genauere Betrachtung der Ergebnisse zeigt außerdem, dass das neuronale Netz eine deutlich bessere Trefferquote (engl. „recall“) als die SVM erzielt und sich somit besser für den Einsatz in argumentativen Schreibsystemen eignet.

5 Zusammenfassung

In dieser Dissertation wurden neue Ansätze zur automatischen Erkennung und Bewertung von natürlichsprachlichen Argumenten vorgestellt. Um die erforderlichen Korpora zu erstellen, wurden existierende Argumentationstheorien verglichen und ein Argumentationsmodell vorgestellt, welches die gesamte Argumentationsstruktur eines Dokuments als Baum modelliert. Wir zeigten erstmalig, dass menschliche Annotatoren Argumentationsstrukturen in argumentative Aufsätzen mit hoher Übereinstimmung identifizieren. Somit leistet diese Dissertation einen Beitrag für ein komplexes Problem der automatischen Sprachverarbeitung und dem jungen Forschungsbereich „Argument Mining“. Das Ergebnis dieser Annotationsstudie ist ein mit Argumentationsstrukturen annotiertes Korpus, welches von der Forschungsgemeinschaft vielfältig für weitere Arbeiten genutzt wird. Darüber hinaus stellten wir einen neuen Ansatz zur automatischen Erkennung von Argumentationsstrukturen vor. Dieser Ansatz erkennt die Grenzen von Argumentkomponenten auf Wortebene und optimiert die Funktion von Argumentkomponenten und argumentativen Relationen. Die Evaluationsergebnisse zeigen, dass dieser Ansatz nicht nur konsistente Argumentationsstrukturen erkennt, sondern im Vergleich zu mehreren Baselines auch signifikant bessere Erkennungsraten erzielt. Zusätzlich wurde ein Ansatz zur automatischen Qualitätsbewertung von Argumenten vorgestellt. Wir untersuchten erstmals die Eigenschaften von unzureichend begründeten Argumenten und stellten einen Ansatz basierend auf neuronalen Netzen vor, welcher im Vergleich mit mehreren Baselinesystemen signifikant bessere Ergebnisse erzielt.

Literaturverzeichnis

- [BB11] Butler, Jodie A.; Britt, M. Anne Britt: Investigating Instruction for Improving Revision of Argumentative Essays. *Written Communication*, 28(1):70–96, 2011.
- [BMB10] Bentahar, Jamal; Moulin, Bernard; Bélanger, Micheline: A taxonomy of argumentation models used for knowledge representation. *Artifi. Intelli. Rev.*, 33(3):211–259, 2010.

- [Go10] Govier, Trudy: A Practical Study of Argument. Wadsworth, Cengage Learning, 7th. Auflage, 2010.
- [JB77] Johnson, Ralph H.; Blair, Anthony J.: Logical Self-Defense. McGraw-Hill Ryerson, 1977.
- [Kü08] Kübler, Sandra; McDonald, Ryan; Nivre, Joakim; Hirst, Graeme: Dependency Parsing. Morgan and Claypool Publishers, 2008.
- [SB13] Shermis, Mark D.; Burstein, Jill: Handbook of Automated Essay Evaluation: Current Applications and New Directions. Routledge Chapman & Hall, 2013.
- [SG14a] Stab, Christian; Gurevych, Iryna: Annotating Argument Components and Relations in Persuasive Essays. In: Proceedings of the 25th International Conference on Computational Linguistics (COLING 2014). S. 1501–1510, 2014.
- [SG14b] Stab, Christian; Gurevych, Iryna: Identifying Argumentative Discourse Structures in Persuasive Essays. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). S. 46–56, 2014.
- [SG17a] Stab, Christian; Gurevych, Iryna: Parsing Argumentation Structures in Persuasive Essays. Computational Linguistics, 43(3):619–659, 2017.
- [SG17b] Stab, Christian; Gurevych, Iryna: Recognizing Insufficiently Supported Arguments in Argumentative Essays. In: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers. S. 980–990, 2017.
- [SN13] Schiappa, Edward; Nordin, John P.: Keeping Faith with Reason: A Theory of Practical Reason. Pearson Learning Solutions, 2013.
- [St17] Stab, Christian: Argumentative Writing Support by means of Natural Language Processing. Dissertation, Technische Universität Darmstadt, 2017.
- [Th73] Thomas, Stephen N.: Practical reasoning in natural language. Prentice-Hall, 1973.
- [To58] Toulmin, Stephen E.: The Uses of Argument. Cambridge University Press, 1958.
- [WRM08] Walton, Douglas; Reed, Chris; Macagno, Fabrizio: Argumentation Schemes. Cambridge University Press, 2008.



Christian Stab studierte Informatik an der Technischen Universität Darmstadt. Im Jahr 2009 schrieb er seine Diplomarbeit zum Thema „Interaktionsanalyse für adaptive Benutzerschnittstellen“. Nach seinem Studium arbeitete er vier Jahre im Bereich Informationsvisualisierung am Fraunhofer Institut für Graphische Datenverarbeitung. Im Oktober 2013 begann er mit seiner Dissertation am UKP-Lab der Technischen Universität Darmstadt. Seine Dissertation schloss er im Februar 2017 mit Auszeichnung ab. Er ist (Ko)-Autor von mehr als dreißig Publikationen auf renommierten, internationalen Konferenzen. Seine Forschung im Bereich „Argument Mining“ wurde im Jahr 2016 mit einem „IBM Ph.D. Fellowship Award“ ausgezeichnet. Derzeit ist er Post-Doc am UKP-Lab und koordiniert das Validierungsvorhaben „ArgumenText“, in dem die neuesten Methoden zur automatischen Argumentextraktion in industrienahen Anwendungen erprobt werden.