KI-gestützte Inhaltsanalyse von wissenschaftlichen Arbeiten mittels Fine-Tunings von Large Language Models

Samir M'Sallem
Technische Hochschule Mittelhessen
Fachbereich MND
Wilhelm-Leuschner-Str. 13
61169 Friedberg
E-Mail:
samir.m.sallem@mnd.thm.de

Prof. Dr. Harald Ritz
Technische Hochschule Mittelhessen
Fachbereich MNI
Wiesenstraße 14
35390 Gießen
E-Mail:
harald.ritz@mni.thm.de

Prof. Dr. Frank Kammer
Technische Hochschule Mittelhessen
Fachbereich MNI
Wiesenstraße 14
35390 Gießen
E-Mail:
frank.kammer@mni.thm.de

Kategorie

Masterarbeit

Schlüsselwörter

Künstliche Intelligenz, Natural Language Processing, Large Language Models, Fine-Tuning, German BERT, Named Entity Recognition, Part-of-Speech-Tagging, Text Classification, Topic Modelling, Semantic Textual Similarity, Explainable Artificial Intelligence, Kohärenz, Kohäsion, Argument Mining

Zusammenfassung

Diese Masterarbeit untersucht, inwiefern Verfahren der künstlichen Intelligenz zur automatisierten Analyse von Qualitätsmerkmalen in wissenschaftlichen Texten eingesetzt werden können. Das Ziel ist die Entwicklung eines prototypischen Systems, das deutschsprachige Arbeiten aus der Wirtschaftsinformatik hinsichtlich definitorischer Vollständigkeit, argumentativer Bezüge im Schlussteil und semantischer Kohärenz prüft. Zur methodischen Umsetzung werden vortrainierte Transformer-Sprachmodelle herangezogen und durch Transfer Learning auf die Aufgaben feinabgestimmt. Zusätzlich wird eine Themenmodellierung zur Extraktion von Themen in Kapiteln und Argumenten implementiert. Ergänzend erfolgt die Berechnung der Kosinus-Ähnlichkeit dieser Themen, um semantische Nähe von Sätzen und Kapiteln festzustellen. Die Ergebnisse zeigen, dass die anvisierten Charakteristika von wissenschaftlichen Arbeiten zuverlässig erkannt und analysiert werden können. Zukünftige Arbeiten sollten den Fokus auf eine effiziente Erklärbarkeit der Ergebnisse und einer erweiterten Datenbasis für das Modelltraining legen.

Problemstellung

In den Bildungsbereichen hängt die Qualität wissenschaftlicher Arbeiten maßgeblich von der Klarheit zentraler Begriffe, der Argumentationsstruktur, sowie der inhaltlichen Kohärenz ab. Künstliche Intelligenz besitzt die Fähigkeit große Mengen an Informationen schnell und präzise zu analysieren. Diese Masterarbeit untersucht daher, inwiefern Verfahren der künstlichen Intelligenz zur automatisierten Analyse dieser Qualitätsmerkmale in wissenschaftlichen Texten eingesetzt werden können, um Lehrende und Studierende bei der Erstellung und Evaluation von wissenschaftlichen Arbeiten zu unterstützen.

Zielsetzung

Das Ziel ist die Entwicklung eines prototypischen Systems, das deutschsprachige Arbeiten aus der Wirtschaftsinformatik hinsichtlich definitorischer Vollständigkeit, argumentativer Bezüge im Schlussteil und semantischer Kohärenz prüft. Der Arbeit liegen dabei folgende Forschungsfragen zugrunde:

1. Wie kann automatisch überprüft werden, ob in wissenschaftlichen Arbeiten verwendete Fachbegriffe definiert wurden?
2. Wie kann überprüft werden, ob Autoren im Schlussteil einer wissenschaftlichen Arbeit auf vorhergehende Theorien und Konzepte Bezug nehmen und diese argumentativ stützen?
3. Wie kann analysiert werden, ob Textteile in wissenschaftlichen Arbeiten semantisch kohärent sind?

Vorgehensweise

Diese Arbeit ist in acht Kapitel gegliedert. Das erste Kapitel leitet das Thema und die zugrunde liegenden Forschungsfragen ein. Im zweiten Kapitel werden die notwendigen sprachwissenschaftlichen Grundlagen erklärt. Anschließend werden die technischen Grundlagen dargelegt. Ausgehend von den Forschungsfragen erfolgt eine ausführliche Darlegung des Forschungsstands unter Einbeziehung thematisch verwandter Arbeiten. Die anschließende Beantwortung der Forschungsfragen erfolgt in einem mehrteiligen Verfahren. Zunächst werden die Forschungsfragen unter Berücksichtigung erarbeiteter Grundlagen und verwandter Arbeiten in softwaretechnische Systemanforderungen überführt. Weiterhin werden geeignete Technologien aufgestellt und ein vollständiges Konzept mit Architekturdiagrammen aufgestellt. Im nächsten Schritt wird die Implementierung der

notwendigen Softwareartefakte erläutert. Nachdem die Implementierung abgeschlossen ist, werden die entwickelten Sprachmodelle anhand von aufgestellten Metriken evaluiert. Zusätzlich werden die hervorgebrachten Ergebnisse entsprechend der Systemanforderungen validiert. Schlussendlich mündet die Arbeit in ein Schlusskapitel, in dem die Beantwortung der Forschungsfragen sowie ein Fazit und Ausblick im Vordergrund stehen.

Umsetzung

Zur methodischen Umsetzung werden vortrainierte Transformer-Sprachmodelle herangezogen und durch Transfer Learning auf die Aufgaben feinabgestimmt. Die Transformer-Modelle basieren auf dem Self-Attention-Mechanismus, der es ihnen erlaubt, semantische Beziehungen zwischen Wörtern auch über größere Textkontexte hinweg zu erfassen. Im Hinblick auf die Forschungsfragen ist eine kontextübergreifende Verarbeitung erforderlich, die durch Transformer sichergestellt werden können. Die bisherigen Ansätze verwandter Arbeiten zeigen auf, dass Transformer für Klassifikationsaufgaben auf Wort- und Satzebene herangezogen werden sollten. Die Arbeit umfasst die Modellierung und Erstellung von Datensätzen. Als Basis dienen drei zentrale Datenstrategien, die aus der Nutzung eines Wikipedia-Datensatzes, der Erstellung eines Datensatzes mit Hilfe von existierenden wissenschaftlichen Arbeiten und der Erstellung eines Datensatzes mittels generativer KI bestehen. Die Annotation der Datensätze erfolgte im Falle der ersten Datenstrategie manuell, indem Merkmale von Definitionen innerhalb der Texte markiert werden. Schließlich folgt die Auswahl von Sprachmodellen, die für die Klassifikation von wissenschaftlichen Texten in der Wirtschaftsinformatik geeignet sein könnten. Hierzu fällt die Auswahl unter anderem auf die Modelle deepset/gbert-base und deepset/gbert-large, da diese auf der originalen BERT-Architektur aufbauen und darüber hinaus auf großen deutschsprachigen Textkorpora trainiert wurden, weshalb diese für die Analyse von Qualitätsmerkmalen in wissenschaftlichen Texten geeignet sein könnten. Es folgt ein Fine-Tuning der selektierten Sprachmodelle auf die Zielaufgaben unter Einsatz der zuvor entwickelten Datensätze. Zusätzlich wird eine komplexe Systemarchitektur bestehend aus Frontend und Backend implementiert, die den explorativen Einsatz der entwickelten Sprachmodelle ermöglicht. Zu diesem Zweck werden die Python-Bibliotheken Flask und Streamlit herangezogen. Zusätzlich wird die Themenmodellierung zur Extraktion von Themen in Kapiteln und Argumenten implementiert. Ergänzend erfolgt die Berechnung der Kosinus-Ähnlichkeit dieser Themen, um semantische Nähe von Sätzen und Kapiteln festzustellen. Schlussendlich werden die Resultate in die Anwendung eingearbeitet, sodass Nutzer eine Analyse von PDF-Dateien und einzelnen Textpassagen vornehmen können und passende Ergebnisse erhalten. Zusätzlich existiert eine Anzeige von Explainable Artificial Intelligence (XAI) Ausgaben, um eine Nachvollziehbarkeit der durch die Modelle vorgegebenen Ergebnisse zu ermöglichen.

Ergebnisse und Fazit

Die Ergebnisse zeigen, dass die zugrundeliegenden Charakteristika von wissenschaftlichen Arbeiten zuverlässig erkannt und analysiert werden können. Eine Erkennung von Definitionen wird mit einer Genauigkeit von 96,3% durch das feinabgestimmte gbert-base-Modell erfüllt. Die Klassifikation von Textteilen erfolgt unter einer Genauigkeit von 78,5% sowie die Klassifikation von Argumenten, Gegenargumenten und Konklusionen im Fazit unter einer Genauigkeit von 96,6%. Zuletzt bringt das Fine-Tuning von gbert-large eine Erkennung von Kohärenz und Kohäsion in Texten mit einer Genauigkeit von 95,3% hervor. Damit wird deutlich, dass die in der vorliegenden Arbeit angewendete Klassifikation zur Beantwortung der Forschungsfragen zielführend ist. Gleichzeitig wird erkennbar, dass die Qualität der Resultate maßgeblich von der Beschaffenheit der Datensätze und Modelle abhängt. Eine konsequente Anwendung eines „Multi-Annotator-Ansatzes“ erscheint sinnvoll, um die Konsistenz und Ausgewogenheit der Trainingsdaten sicherzustellen. Ebenfalls stellen die erweiterte Anwendung von Data Augmentation, die Berücksichtigung von weiteren Mustern von Kohärenz und Kohäsion sowie die Einbettung von Definitionen über mehrere Sätze eine potentielle Verbesserung der Erkennungsleistung dar. Auch der Einsatz von Explainable Artificial Intelligence zur Klärung der prognostizierten Resultate durch die Sprachmodelle erweist sich als vielversprechend, jedoch zeitaufwändig. Zukünftige Arbeiten sollten daher den Fokus auf eine effiziente Erklärbarkeit der Ergebnisse legen. Perspektivisch kann der vorgestellte Ansatz dazu beitragen, ein intelligentes Tutorensystem zu entwickeln, das Studierende bei der Erstellung wissenschaftlicher Texte unterstützt und die Bewertung durch Lehrende ergänzt.

Literatur

L. J. Heinrich, R. Riedl, und A. Heinzl, „Wissenschaftscharakter der Wirtschaftsinformatik“, in Wirtschaftsinformatik: Einführung und Grundlegung, L. J. Heinrich, A. Heinzl, und R. Riedl, Hrsgg. Berlin, Heidelberg: Springer,2011, S. 47–60. Online verfügbar: https://doi.org/10.1007/978-3-642-15426-3_4 (abgerufen: 24.04.2025).

L. J. Heinrich, R. Riedl, und A. Heinzl, „Fachsprache der Wirtschaftsinformatik“, in Wirtschaftsinformatik: Einführung und Grundlegung, L. J. Heinrich, A. Heinzl, und R. Riedl, Hrsgg. Berlin, Heidelberg: Springer, 2011, S. 61–72. Online verfügbar: https://doi.org/10.1007/978-3-642-15426-3_5 (abgerufen: 24.04.2025).

R. Musan, „Kohäsion und Kohärenz“, in Linguistik im Sprachvergleich: Germanistik– Romanistik – Anglistik,

R. Klabunde, W. Mihatsch, und S. Dipper, Hrsgg. Berlin, Heidelberg: Springer Berlin Heidelberg, 2022, S. 577–593.

D. Jurafsky und J. H. Martin, Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models, 3. Aufl., 2025. Online verfügbar: https://web.stanford.edu/~jurafsky/slp3/ (abgerufen: 01.02.2025).

A. Vaswani et al., „Attention Is All You Need“, Aug. 2023. Online verfügbar: http://arxiv.org/abs/1706.03762 (abgerufen: 03.03.2025). ArXiv:1706.03762 [cs].

J. Devlin, M.-W. Chang, K. Lee, und K. Toutanova, „BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding“, Mai 2019. Online verfügbar: http://arxiv.org/abs/1810.04805 (abgerufen: 16.12.2024). ArXiv:1810.04805 [cs].

A. Storrer und S. Wellinghoff, „Automated detection and annotation of term definitions in German text corpora“, in Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC‘06). Genoa, Italy: European Language Resources Association (ELRA), Mai 2006. Online verfügbar: https://aclanthology.org/L06-1066/ (abgerufen: 23.03.2025).

A.-M. Avram, D.-C. Cercel, und C. Chiru, „UPB at SemEval-2020 Task 6: Pretrained Language Models for Definition Extraction“, in Proceedings of the Fourteenth Workshop on Semantic Evaluation, A. Herbelot, X. Zhu, A. Palmer,N. Schneider, J. May, und E. Shutova, Hrsgg. Barcelona (online): International Committee for Computational Linguistics, Dez. 2020, S. 737–745. Online verfügbar: https://aclanthology.org/2020.semeval-1.97/ (abgerufen: 13.01.2025).

S. Teufel, A. Siddharthan, und C. Batchelor, „Towards Domain-Independent Argumentative Zoning: Evidence from Chemistry and Computational Linguistics“, in Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, P. Koehn und R. Mihalcea, Hrsgg. Singapore: Association for Computational Linguistics, Aug. 2009, S. 1493–1502. Online verfügbar: https://aclanthology.org/D09-1155/ (abgerufen: 25.01.2025).

O. Kashefi, S. Chan, und S. Somasundaran, „Argument Detection in Student Essays under Resource Constraints“, in Proceedings of the 10th Workshop on Argument Mining, M. Alshomary, C.-C. Chen, S. Muresan, J. Park, und J. Romberg, Hrsgg. Singapore: Association for Computational Linguistics, Dez.2023, S. 64–75. Online verfügbar: https://aclanthology.org/2023.argmining-1.7/ (abgerufen: 09.01.2025).

A. Shen, M. Mistica, B. Salehi, H. Li, T. Baldwin, und J. Qi, „Evaluating Document Coherence Modeling“, Transactions of the Association for Computational Linguistics, Bd. 9, S. 621–640, Juli 2021. Online verfügbar: https://doi.org/10.1162/tacl_a_00388 (abgerufen: 26.11.2024).

P. Törnberg, „Best Practices for Text Annotation with Large Language Models“, Febr. 2024. Online verfügbar: http://arxiv.org/abs/2402.05129 (abgerufen:23.02.2025).

B. Chan, S. Schweter, und T. Möller, „German’s Next Language Model“, Dez. 2020. Online verfügbar: http://arxiv.org/abs/2010.10906 (abgerufen: 26.03.2025).