



Hochschule für angewandte Wissenschaften Coburg  
Fakultät Elektrotechnik und Informatik

Studiengang: Informatik

Bachelorarbeit

# Multi-Modal Feature Fusion with Cross-Attention for Tabular and Textual Data

Städler, Sebastian

Abgabe der Arbeit: 09.12.2025

Betreut durch:

Prof. Dr. Roman Rischke, Hochschule Coburg

# Inhaltsverzeichnis

<b>1</b>	<b>Einführung</b>	<b>1</b>
<b>2</b>	<b>Theoretische Grundlagen</b>	<b>2</b>
2.1	Repräsentation von Text- und Tabulardaten . . . . .	2
2.1.1	Textrepräsentation . . . . .	2
2.1.2	Tabellendatenrepräsentation . . . . .	3
2.2	Attention-Mechanismen . . . . .	3
2.3	Transformer-Architektur . . . . .	3
2.4	Multimodales Lernen und Feature Fusion . . . . .	3
<b>3</b>	<b>Verwandte Arbeiten</b>	<b>4</b>
<b>4</b>	<b>Methodik</b>	<b>5</b>
<b>5</b>	<b>Ergebnisse</b>	<b>6</b>
<b>6</b>	<b>Diskussion und Fazit</b>	<b>7</b>
	<b>Literaturverzeichnis</b>	<b>II</b>
	<b>Ehrenwörtliche Erklärung</b>	<b>III</b>

# 1 Einführung

Das ist meine Einfuehrung ...

## 2 Theoretische Grundlagen

### 2.1 Repräsentation von Text- und Tabulardaten

#### 2.1.1 Textrepräsentation

Für die effektive Verarbeitung natürlicher Sprache (NLP) und die in dieser Arbeit untersuchte Fusion von Text- und Tabellendaten ist die Transformation textueller Informationen in numerische Vektorrepräsentationen unerlässlich [AU22].

Zentrale Begriffe:

- **Text Vektorisierung:** Umwandlung von Text in numerische Vektoren, notwendig für maschinelle Lernalgorithmen [AU22].
- **Embedding:** Gelernte, dichte und niedrigdimensionale Vektorrepräsentation von Text (Wörtern, Sätzen etc.), die semantische Beziehungen abbildet (z. B. „king – man + woman  $\approx$  queen“) [MCCD13].
- **Repräsentation:** Jede Form der Textkodierung im Modell.

Historisch betrachtet basierten frühe Ansätze auf frequenzbasierten Verfahren wie dem Bag-of-Words-Modell oder der Term Frequency-Inverse Document Frequency (TF-IDF) [SB88]. Diese Methoden repräsentieren Dokumente als Vektoren von Worthäufigkeiten, ignorieren jedoch weitgehend die grammatikalische Struktur und die semantische Bedeutung der Wörter. Zudem führen sie oft zu hochdimensionalen, dünn besetzten Vektoren (sparse vectors).

Einen signifikanten Fortschritt markierte die Einführung von verteilten Wortrepräsentationen (Word Embeddings), insbesondere durch das Word2Vec-Verfahren [MCCD13]. Hierbei werden Wörter in einen dichten, niedrigdimensionalen Vektorraum projiziert, wobei diese Projektionen durch einfache neuronale Netze gelernt werden. Semantisch ähnliche Wörter liegen in diesem Vektorraum nahe beieinander. Ein wesentlicher Nachteil dieser statischen Embeddings besteht jedoch darin, dass jedem Wort unabhängig von seinem Kontext ein fixer Vektor zugewiesen wird. Polyseme Wörter, die je nach Satzkontext unterschiedliche Bedeutungen haben, können so nicht adäquat abgebildet werden.

Um dieses Defizit zu beheben, wurden kontextuelle Embeddings entwickelt. Peters et al. stellten mit ELMo (Embeddings from Language Models) einen Ansatz vor, der tiefere neuronale Netze nutzt, um kontextabhängige Wortvektoren zu generieren [PNI<sup>+</sup>18]. Den aktuellen Standard setzen jedoch Transformer-basierte Modelle wie BERT (Bidirectional Encoder Representations from Transformers), eingeführt von Devlin et al. [DCLT19]. BERT erzeugt durch ein bidirektionales Training tiefgreifende kontextuelle Repräsentationen, die das Verständnis von komplexen

Sprachstrukturen ermöglichen. Eine detailliertere Betrachtung der Transformer-Architektur erfolgt in einem späteren Kapitel dieses Theorieteils.

Für die Verarbeitung ganzer Sätze oder Dokumente, wie sie in dieser Arbeit relevant ist, stoßen reine Token-Embeddings jedoch an Grenzen. Hier kommen Techniken wie Pooling oder die Verwendung spezieller Tokens ins Spiel. Insbesondere das [CLS]-Token (Classifier Token) in Modellen wie BERT wird häufig genutzt, um eine aggregierte Repräsentation des gesamten Satzes zu erhalten. Dieses Token sammelt während der Verarbeitung durch die Transformer-Schichten kontextuelle Informationen des gesamten Inputs und kann dann als Satz-Embedding verwendet werden. Während Modifikationen wie Sentence-BERT (SBERT) [RG19] darauf abzielen, semantisch aussagekräftige Satz-Embeddings zu erzeugen, liegt der Fokus dieser Arbeit auf der direkten Nutzung von [CLS]-Tokens für die Satzrepräsentation. Diese stabilen semantischen Kodierungen auf Satzebene schaffen die notwendige Voraussetzung, um Textinformationen effizient mit anderen Modalitäten, wie tabellarischen Merkmalen, in einer multimodalen Architektur zu fusionieren.

### **2.1.2 Tabellendatenrepräsentation**

Test

## **2.2 Attention-Mechanismen**

## **2.3 Transformer-Architektur**

## **2.4 Multimodales Lernen und Feature Fusion**

## **3 Verwandte Arbeiten**

## **4 Methodik**

## **5 Ergebnisse**



## **6 Diskussion und Fazit**

## Literaturverzeichnis

- [AU22] ABUBAKAR, H. D. ; UMAR, M.: Sentiment Classification: Review of Text Vectorization Methods: Bag of Words, Tf-Idf, Word2vec and Doc2vec. In: *SLUJST* 4 (2022), Aug, Nr. 1 & 2, S. 27–33. <http://dx.doi.org/10.56471/slujst.v4i.266>. – DOI 10.56471/slujst.v4i.266
  
- [DCLT19] DEVLIN, Jacob ; CHANG, Ming-Wei ; LEE, Kenton ; TOUTANOVA, Kristina: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. (2019), May. <http://dx.doi.org/10.48550/arXiv.1810.04805>. – DOI 10.48550/arXiv.1810.04805
  
- [MCCD13] MIKOLOV, Tomas ; CHEN, Kai ; CORRADO, Greg ; DEAN, Jeffrey: Efficient Estimation of Word Representations in Vector Space. (2013), Sep. <http://dx.doi.org/10.48550/arXiv.1301.3781>. – DOI 10.48550/arXiv.1301.3781
  
- [PNI<sup>+</sup>18] PETERS, Matthew E. ; NEUMANN, Mark ; IYYER, Mohit ; GARDNER, Matt ; CLARK, Christopher ; LEE, Kenton ; ZETTLEMOYER, Luke: Deep Contextualized Word Representations. In: WALKER, Marilyn (Hrsg.) ; JI, Heng (Hrsg.) ; STENT, Amanda (Hrsg.): *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*. New Orleans, Louisiana : Association for Computational Linguistics, Jun 2018, S. 2227–2237
  
- [RG19] REIMERS, Nils ; GUREVYCH, Iryna: Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. (2019), Aug. <http://dx.doi.org/10.48550/arXiv.1908.10084>. – DOI 10.48550/arXiv.1908.10084
  
- [SB88] SALTON, Gerard ; BUCKLEY, Chris: Term-weighting approaches in automatic text retrieval. In: *Information Processing & Management* 24 (1988), Nr. 5, S. 513–523. [http://dx.doi.org/10.1016/0306-4573\(88\)90021-0](http://dx.doi.org/10.1016/0306-4573(88)90021-0). – DOI 10.1016/0306-4573(88)90021-0

**Persönliche Angaben / Personal details**

**Städler, Sebastian**

Familienname, Vorname / Surnames, given names

**19.03.2002**

Geburtsdatum / Date of birth

**Informatik**

Studiengang / Course of study

**00383022**

Matrikelnummer / Student registration number

**Eigenständigkeitserklärung*****Declaration***

Hiermit versichere ich, dass ich diese Arbeit selbständig verfasst und noch nicht anderweitig für Prüfungszwecke vorgelegt habe. Ich habe keine anderen als die angegebenen Quellen oder Hilfsmittel benutzt. Die Arbeit wurde weder in Gänze noch in Teilen von einer Künstlichen Intelligenz (KI) erstellt, es sei denn, die zur Erstellung genutzte KI wurde von der zuständigen Prüfungskommission oder der bzw. dem zuständigen Prüfenden ausdrücklich zugelassen. Wörtliche oder sinngemäße Zitate habe ich als solche gekennzeichnet.

Es ist mir bekannt, dass im Rahmen der Beurteilung meiner Arbeit Plagiatserkennungssoftware zum Einsatz kommen kann.

Es ist mir bewusst, dass Verstöße gegen Prüfungsvorschriften zur Bewertung meiner Arbeit mit „nicht ausreichend“ und in schweren Fällen auch zum Verlust sämtlicher Wiederholungsversuche führen können.

*I hereby certify that I have written this thesis independently and have not submitted it elsewhere for examination purposes. I have not used any sources or aids other than those indicated. The work has not been created in whole or in part by an artificial intelligence (AI), unless the AI used to create the work has been expressly approved by the responsible examination board or examiner. I have marked verbatim quotations or quotations in the spirit of the text as such.*

*I am aware that plagiarism detection software may be used in the assessment of my work.*

*I am aware that violations of examination regulations can lead to my work being graded as "unsatisfactory" and, in serious cases, to the loss of all repeat attempts.*

\_\_\_\_\_  
Unterschrift Studierende/Studierender / Signature student

**96450 Coburg, den 09.12.2025**

Ort, Datum / Place, date