

Multi-Task-Learning mit Transformer-Modellen für die Ad-Hoc-Suche mittels Information-Retrieval-Axiomen

Adrien Klose

`adrien.klose@student.uni-halle.de`

Martin-Luther-Universität Halle-Wittenberg
Institut für Informatik

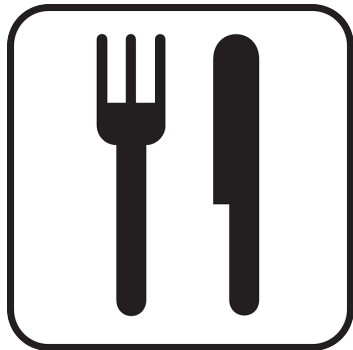
Dezember 16, 2021

Problem Ursprung

Bedürfnisse

„All human activity is prompted by desire.“

Bertrand Russel, Literatur Nobelpreisträger 1955



Problem Ursprung

Informationsbedürfnisse

Google

Bing

Ask

YAHOO!

Yandex



DuckDuckGo

Baidu 百度



<https://aeroadmin.com/articles/en/wp-content/uploads/2020/11/search-engine-logo.png>

Mahatma Ghandi Mensch

- 1 Mahatma Glück, Mahatma Pech, Mahatma Ghandi ist Teil eines Liedes von Bernd Stelter.
- 2 Mahatma Ghandi ist ein Mensch, der die indische Unabhängigkeitsbewegung maßgeblich beeinflusste.



⇒ Fehlerhafte Entscheidung

Problem Ad-Hoc-Suche

Axiomatischer Lösungsansatz

- ▶ IR-Axiom: Heuristiken für effektive Rankings
- ▶ LB1: Anzahl verschiedener Anfrageterme im Dokument
- ▶ Reranking nach LB1 [Hagen et al., CIKM'16]:

Mahatma Ghandi Mensch

- 1 **Mahatma Ghandi** ist ein **Mensch**, der die indische Unabhängigkeitsbewegung maßgeblich beeinflusste.
- 2 **Mahatma** Glück, **Mahatma** Pech, **Mahatma Ghandi** ist Teil eines Liedes von Bernd Stelter.



Problem Ad-Hoc-Suche

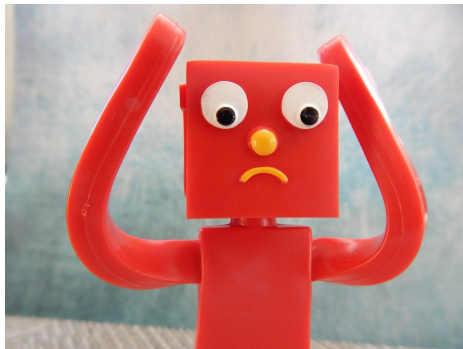
Unser axiomatischer Lösungsansatz

- ▶ Vorbild Natural-Language-Understanding-Ansatz
[Liu et al., ACL'19]
- ▶ Vortrainiertes Transformer-Modell BERT
- ▶ Multi-Task-Learning
 - ▶ IR-Axiom-Klassifikationsaufgaben
 - ▶ Ranking-Regressionsaufgabe (Pointwise)

Repräsentationen für Sprache

Japanische Suchanfrage

ニューヨークタイムズは
何ですか？



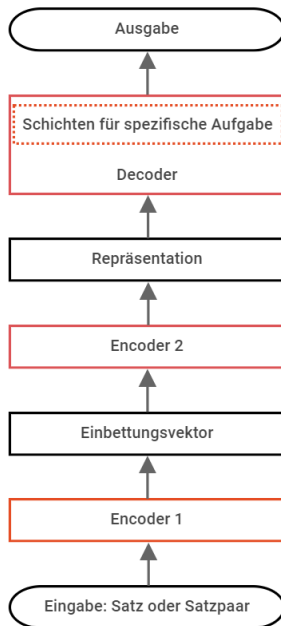
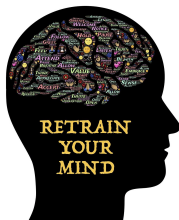
What is New York Times?



Repräsentationen für Sprache

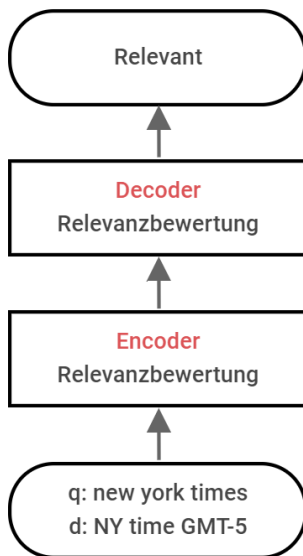
Transformer-Modell BERT

- ▶ Masked-Language-Modelling
- ▶ Next-Sentence-Prediction
- ▶ Aufgabe beeinflusst Repräsentation von Sprache



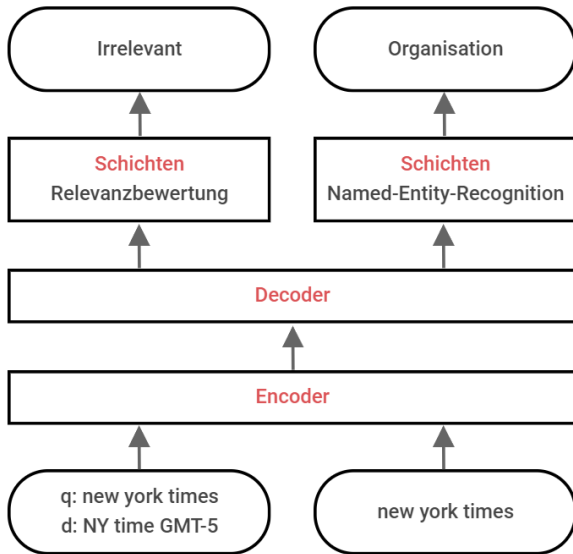
Multi-Task-Learning

Single-Task-Setup



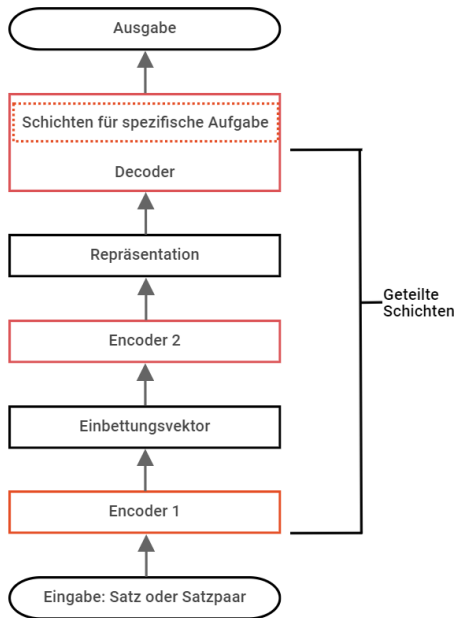
Multi-Task-Learning

Multi-Task-Setup



Multi-Task-Learning

BERT



IR-Axiom-Klassifikationsaufgaben

- ▶ IR-Axiom: Heuristiken für effektives Ranking
- ▶ Eingabe: eine Anfrage q , zwei Dokumente d_1 und d_2
- ▶ Ausgabe: Ranking-Präferenz
 - ▶ d_1 relevanter als d_2
 - ▶ d_2 relevanter als d_1
 - ▶ d_1 und d_2 gleich relevant
- ▶ Zielstellung überschneidet sich mit Ranking-Aufgabe
- ▶ Trainingsdaten unbeschränkt erstellbar

Experimente

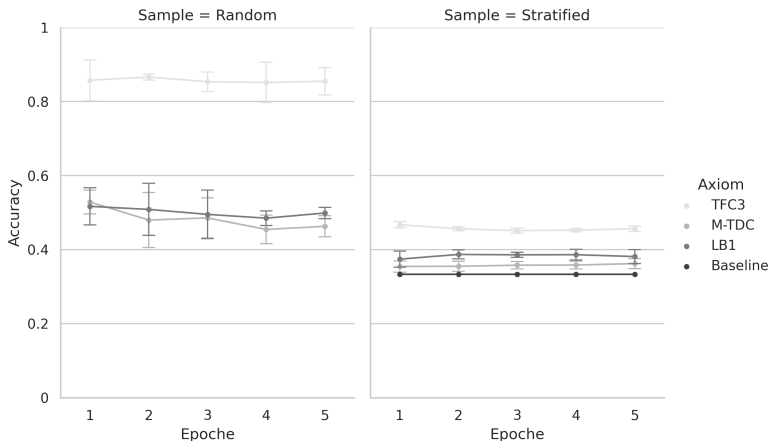
Implementation

- ▶ 20 IR-Axiome [Völske et al., ICTIR'21]
- ▶ MS MARCO Passagen Datensatz
- ▶ Test-/Trainingsdaten Erzeugung IR-Axiome:
 - ▶ 700 Anfragen Top-100 Dokumente
 - ▶ Disjunkte 100 Anfragen Top-100 Dokumente
 - ▶ Eingabereihenfolge nach Dokument-ID
 - ▶ Filter: Vorbedingung erfüllt
 - ▶ Stratifiziert (Train, Test) und zufällig (Test) Stichprobe
- ▶ BERT-Base-Uncased Modell
- ▶ MT-DNN [Liu et al., ACL'19]

Experimente

Single-Task-Setup IR-Axiome

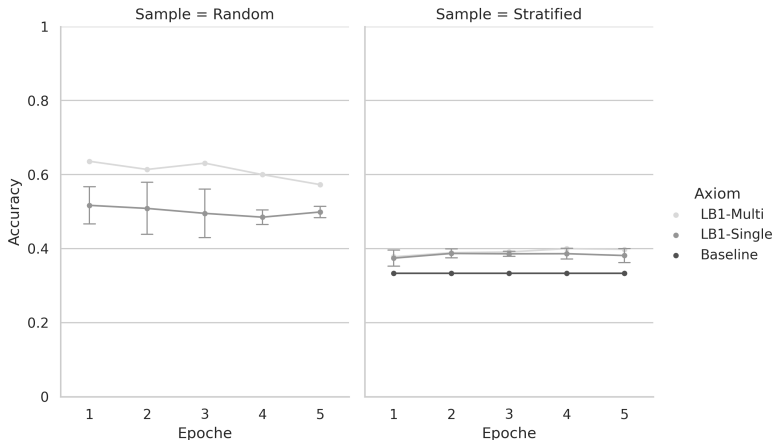
► 10.000 Trainingsdaten für jedes Axiom



Experimente

Multi-Task-Setup IR-Axiome

- ▶ 10.000 Trainingsdaten LB1-Single
- ▶ 25.000 Trainingsdaten LB1-Multi



Experimente

Ergebnisse Ranking

- ▶ Beste Ranking-Ergebnisse aus 5 Epochen
- ▶ Trainingsdaten: 100.000 Ranking, 25.000 je Axiom

T E R M	Q U E R	P R O X	TREC Deep Learning 2019			TREC Deep Learning 2020		
			NDCG@10	P@1	MRR@10	NDCG@10	P@1	MRR@10
X	X	X	0.686	0.767	0.855	0.614	0.648	0.749

- ▶ TERM: IR-Axiome LB1 und M-TDC
- ▶ QUER: IR-Axiome REG und ANTI-REG
- ▶ PROX: IR-Axiome PROX4 und PROX5

Experimente

Ergebnisse Ranking

- ▶ Beste Ranking-Ergebnisse aus 5 Epochen
- ▶ Trainingsdaten: 100.000 Ranking, 25.000 je Axiom

T E R M	Q U E R	P R O X	TREC Deep Learning 2019			TREC Deep Learning 2020		
			NDCG@10	P@1	MRR@10	NDCG@10	P@1	MRR@10
X	X	X	0.686	0.767	0.855	0.614	0.648	0.749
✓	✓	✓						
✓	X	X						
X	✓	X						
X	X	✓						

- ▶ TERM: IR-Axiome LB1 und M-TDC
- ▶ QUER: IR-Axiome REG und ANTI-REG
- ▶ PROX: IR-Axiome PROX4 und PROX5

Experimente

Ergebnisse Ranking

- ▶ Beste Ranking-Ergebnisse aus 5 Epochen
- ▶ Trainingsdaten: 100.000 Ranking, 25.000 je Axiom

T E R M	Q U E R	P R O X	TREC Deep Learning 2019			TREC Deep Learning 2020		
			NDCG@10	P@1	MRR@10	NDCG@10	P@1	MRR@10
X	X	X	0.686	0.767	0.855	0.614	0.648	0.749
✓	✓	✓	0.684	0.767	0.855			
✓	X	X	0.674	0.698	0.794			
X	✓	X	0.683	0.767	0.843			
X	X	✓	0.658	0.651	0.747			

- ▶ TERM: IR-Axiome LB1 und M-TDC
- ▶ QUER: IR-Axiome REG und ANTI-REG
- ▶ PROX: IR-Axiome PROX4 und PROX5

Experimente

Ergebnisse Ranking

- ▶ Beste Ranking-Ergebnisse aus 5 Epochen
- ▶ Trainingsdaten: 100.000 Ranking, 25.000 je Axiom

T E R M	Q U E R	P R O X	TREC Deep Learning 2019			TREC Deep Learning 2020		
			NDCG@10	P@1	MRR@10	NDCG@10	P@1	MRR@10
X	X	X	0.686	0.767	0.855	0.614	0.648	0.749
✓	✓	✓	0.684	0.767	0.855	0.655	0.648	0.769
✓	X	X	0.674	0.698	0.794	0.626	0.630	0.761
X	✓	X	0.683	0.767	0.843	0.656	0.667	0.790
X	X	✓	0.658	0.651	0.747	0.643	0.667	0.767

- ▶ TERM: IR-Axiome LB1 und M-TDC
- ▶ QUER: IR-Axiome REG und ANTI-REG
- ▶ PROX: IR-Axiome PROX4 und PROX5

Zusammenfassung

- ▶ IR-Axiome ermöglichen Rankinganpassungen
- ▶ IR-Axiome unterschiedlich gut erlernbar
- ▶ Multi-Task-Learning und BERT kombinierbar
- ▶ IR-Axiome und Ranking können voneinander profitieren

Zusammenfassung

- ▶ IR-Axiome ermöglichen Rankinganpassungen
- ▶ IR-Axiome unterschiedlich gut erlernbar
- ▶ Multi-Task-Learning und BERT kombinierbar
- ▶ IR-Axiome und Ranking können voneinander profitieren

Future Work:

- ▶ Ranking- und Axiom-Architektur verbessern
- ▶ Skalierbarkeit und Übertragbarkeit des Ansatzes untersuchen
- ▶ Zusammenhänge zwischen IR-Axiomen ergründen

Zusammenfassung

- ▶ IR-Axiome ermöglichen Rankinganpassungen
- ▶ IR-Axiome unterschiedlich gut erlernbar
- ▶ Multi-Task-Learning und BERT kombinierbar
- ▶ IR-Axiome und Ranking können voneinander profitieren

Future Work:

- ▶ Ranking- und Axiom-Architektur verbessern
- ▶ Skalierbarkeit und Übertragbarkeit des Ansatzes untersuchen
- ▶ Zusammenhänge zwischen IR-Axiomen ergründen

Danke für Ihre Aufmerksamkeit!

Diskussion

Statistik Axiom-Präferenzen

IR Axiom	Vorbedingung=Wahr			Vorbedingung=Falsch		
	-1	0	1	-1	0	1
TFC1	216.744	165.080	211.852	1.205.526	427.602	1.193.196
TFC3	4.544	585.053	4.079	22.172	2.828.823	20.329
M-TDC	10.616	51.161	10.164	307.592	2.787.440	298.027
LB1	610.758	1.111.645	604.088	452.123	278.279	407.107
LNC1	104.010	14.647	99.261	1.596.916	61.800	1.588.366
TF-LNC	73.851	3.317.960	73.189	0	0	0
REG	951.658	1.562.557	950.785	0	0	0
ANTI-REG	1.089.194	1.300.155	1.075.651	0	0	0
DIV	1.695.911	116.491	1.652.598	0	0	0
STMC1	1.616.600	283.596	1.564.804	0	0	0
STMC2	59.472	3.342.367	63.161	0	0	0
PROX1	491.495	145.271	485.691	573.601	1.207.005	561.937
PROX2	538.603	52.329	521.525	1.018.009	309.503	1.015.031
PROX3	16.968	1.089.948	15.541	10.600	2.326.608	5.335
PROX4	1.606.588	261.922	1.594.190	0	0	0
PROX5	1.544.705	360.953	1.557.042	0	0	0
RS-TF	1.718.949	43.852	1.702.199	0	0	0
RS-TF-IDF	1.738.875	22.628	1.703.497	0	0	0
RS-BM25	1.774.654	19.477	1.670.869	0	0	0
RS-QL	1.772.231	19.473	1.673.296	0	0	0

Diskussion

BERT-Base-Uncased

- ▶ Cased hat größeres Vokabular
- ▶ Komplexere Zusammenhänge auf die vortrainiert
- ▶ Fine-Tuning schwieriger/komplexer
- ▶ Anfragen und Dokumente zu hohe Fehlerquote
Groß-/Kleinschreibung vermutet

Diskussion

Warum IR-Axiome?

- ▶ Überschneidungen in Eingabe, Zielstellung, „Sprache“
- ▶ 1 Axiom zu 1 Relevanz-Facette
- ▶ „Unbeschränkte“ Menge Trainings-/Testdaten
- ▶ Intuition + Literatur

Diskussion

Architektur und Eingabeformat

- ▶ [CLS] q [SEP] d1 [SEP] (Ranking, aktuell)
- ▶ [CLS] q [SEP] d1 + „ „ + d2 [SEP] (Axiom, aktuell)
- ▶ Problem: Modell Trennung d1 und d2 unwahrscheinlich
- ▶ [CLS] q [SEP] d1 [SEP] d2 [SEP] (Ranking, besser)
- ▶ [CLS] q [SEP] d1 [SEP] d2 [SEP] (Axiom, besser)
- ▶ Pairwise ist „besser“ als Pointwise
- ▶ Erlaubt Trennung von d1 und d2
- ▶ Problem: MT-DNN und BERT-Modelle

Quellenverzeichnis I

- [1] S. Arora and A. Yates. Investigating retrieval method selection with axiomatic features. *arXiv preprint arXiv:1904.05737*, 2019.
- [2] M. Hagen, M. Völske, S. Göring, and B. Stein. Axiomatic result re-ranking. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, pages 721–730, 2016.
- [3] X. Liu, K. Duh, and J. Gao. Stochastic answer networks for natural language inference. *arXiv preprint arXiv:1804.07888*, 2018.
- [4] X. Liu, P. He, W. Chen, and J. Gao. Multi-task deep neural networks for natural language understanding. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4487–4496, Florence, Italy, July 2019. Association for Computational Linguistics. URL <https://www.aclweb.org/anthology/P19-1441>.
- [5] D. Rennings, F. Moraes, and C. Hauff. An axiomatic approach to diagnosing neural ir models. In *European Conference on Information Retrieval*, pages 489–503. Springer, 2019.

Quellenverzeichnis II

- [6] C. Rosset, B. Mitra, C. Xiong, N. Craswell, X. Song, and S. Tiwary. An axiomatic approach to regularizing neural ranking models. In *Proceedings of the 42nd international ACM SIGIR conference on research and development in information retrieval*, pages 981–984, 2019.
- [7] M. Völske, A. Bondarenko, M. Fröbe, B. Stein, J. Singh, M. Hagen, and A. Anand. Towards axiomatic explanations for neural ranking models. In *Proceedings of the 2021 ACM SIGIR International Conference on Theory of Information Retrieval*, pages 13–22, 2021.