Alexander Bondarenko   Maik Fröbe   Marcel Gohsen   Sebastian Günther   Johannes Kiesel
Jakob Schwerter   Shahbaz Syed   Michael Völske   Martin Potthast   Benno Stein   Matthias Hagen

Webis at TREC 2021: Deep Learning, Health Misinformation, and Podcasts Track

Deep Learning Track

Overview

- Three runs with LambdaMART
- Focus: Anchor text features

50 features

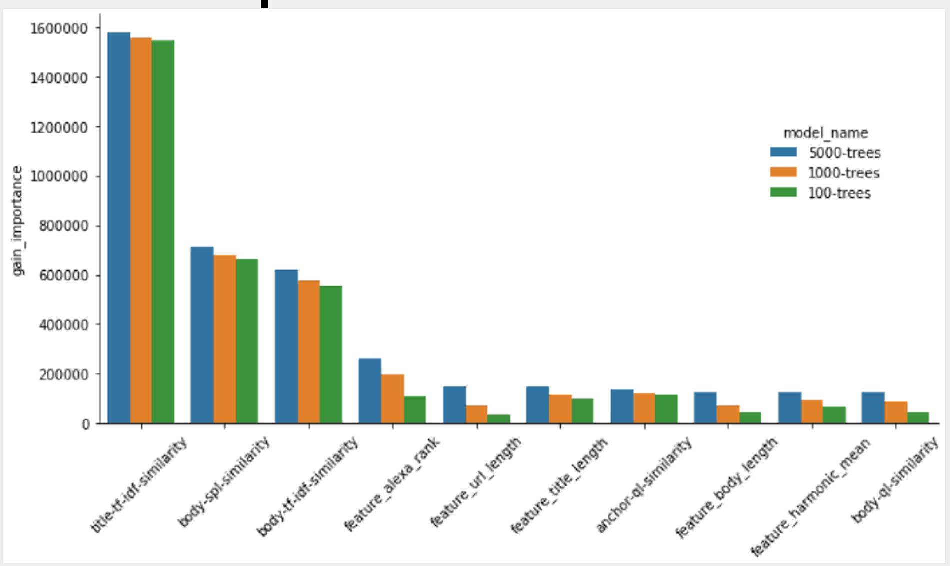
- 36 Query-Document features
  - Anchor Text, Title, Body, URL
- 8 Document features
- 6 Query features

Run	Trees	Features
webis-dl-1	5 k	all 50
webis-dl-2	5 k	41 (no anchors)
webis-dl-3	1 k	all 50

Features

Query-Document		Document/Domain		Query	
Description	Count	Description	Count	Description	Count
Term frequency	4	URL length	1	5W1H	1
TF · IDF	4	Slashes in URL	1	Length in tokens	1
BM25 score	4	Dots in Host	1	GPE Entities	1
F2 exp score	4	Body length	1	ORG Entities	1
F2 log score	4	Title length	1	Person Entities	1
QL score	4	Pagerank	1	Comparative	1
QLJM score	4	Harmonic Mean	1		
PL2 score	4	Alexa Rank	1		
SPL score	4				
Total					50

Feature Importance



Results

	nDCG@3	nDCG@10	P@1	P@3	P@10	MRR@10
webis-dl-1	0.6267	0.5831	0.9123	0.8480	0.7368	0.9356
webis-dl-2	0.6099	0.5747	0.9123	0.8421	0.7298	0.9396
webis-dl-3	**0.6323**	**0.5918**	**0.9298**	**0.8596**	**0.7456**	**0.9488**
baseline	0.5369	0.5116	0.7544	0.7427	0.6684	0.8367

Health Misinformation Track

RQ: Does argumentative re-ranking axioms improve the "helpfulness" while reducing the "harmfulness" of rankings for so called argumentative queries / questions?

Retrieval and Re-ranking

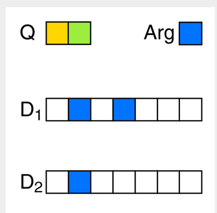
- Two baseline runs: BM25 and MonoT5
- Four re-rankings with 3 argumentative axioms

Runs

- **-ax1:** at least 1 axiom decides to swap document positions
- **-ax3:** all 3 axioms decide to swap document positions

The top-20 initial results are re-ranked with the weighted preferences (swap document positions or not) of the 3 axioms.

Ex.: **ArgUC (Argumentative Units Count)**: Favor documents which contain more argumentative units.



Given:

- Query Q
- Documents D_1, D_2 with $|D_1| \approx_{10\%} |D_2|$
- Arg_D: set of argumentative units of a document D

IF $count(Arg_{D_1}) > count(Arg_{D_2})$ **THEN** $rank(D_1, Q) > rank(D_2, Q)$

QTArg (Query Term Occurrence in Argumentative Units): Favor documents with the query terms close to argumentative units.

QTPArg (Query Term Position in Argumentative Units): Favor documents where the first appearance of a query term in an argumentative unit is closer to the beginning of the document.

Argumentative units (premises and claims) are identified with the BiLSTM-CNN-CRF argument tagging tool TARGER (Chernodub et al., 2019).

Run	Compatibility		nDCG (binary)			P@10 (binary)	
	Help	Harm	U/Co	U/Cr	U/Co/Cr	U/Co	Incor.
webis-bm25 (initial)	0.1292	0.1454	0.4275	0.4856	0.3796	0.3088	0.2906
webis-bm25-ax1	**0.1339**	0.1474	**0.4325**	**0.4877**	**0.3880**	0.3088	0.2844
webis-bm25-ax3	0.1318	0.1445	0.4285	0.4859	0.3802	0.3088	0.2844
webis-t5 (initial)	0.1314	0.1447	0.2383	0.2618	0.1912	0.3235	0.2969
webis-t5-ax1	0.1297	0.1449	0.2362	0.2645	0.1896	**0.3471**	**0.3344**
webis-t5-ax3	0.1327	**0.1438**	0.2392	0.2632	0.1907	0.3412	**0.3344**

U: Usefulness, Co: correctness, Cr: Credibility

Podcasts Track

Retrieval Task

- Four runs for podcast retrieval, all with BM25
- Classification for re-ranking
  - SVM trained with own annotations on Entertaining, Subjective, Discussion
  - Multiplying confidence with BM25 score

Runs

- webis pc bs: no re-ranking
- webis pc cola: COLA audio embeddings

Criterion	Run	nDCG@30	nDCG@1000	P@10
Entertaining	Webis pc bs	**0.1182**	**0.2330**	**0.0975**
	Webis pc cola	0.0522	0.1748	0.0450
	Webis pc rob	0.0351	0.1584	0.0275
	Webis pc co rob	0.0332	0.1620	0.0275
Subjective	Webis pc bs	**0.1725**	**0.3435**	**0.2000**
	Webis pc cola	0.0591	0.2443	0.0600
	Webis pc rob	0.0371	0.2250	0.0350
	Webis pc co rob	0.0430	0.2320	0.0550
Discussion	Webis pc bs	**0.1619**	**0.3208**	**0.1600**
	Webis pc cola	0.0598	0.2289	0.0625

Summarization Task

- Two runs: abstractive and extractive

Runs

- webis pc abstr: DistilBART abstractive summarization
  Input: 5 most entertaining sentences + their 5 previous and following ones
- webis pc extr: TextRank extractive summarization
  Output: 10 sentences with highest entertainment-biased TextRank

Run	EGFB score	E	G	F	B
Webis pc abstr	0.2332	0	6	33	154