Establishing Strong Baselines for TripClick Health Retrieval

Sebastian Hofstätter Sophia Althammer Mete Sertkan Allan Hanbury

TU Wien, Austria

Introduction

- In the web domain neural approaches lead to large effectiveness gains
- Generalizability of neural approaches remains unclear
- Re-test effectiveness of neural approaches on TripClick, a large-scale click data collection

Contributions

- Re-create training data without non-clicked results as negatives
- Establish strong baselines for TripClick with common neural re-ranking models
- Dense retrieval outperforms BM25 for initial candidate retrieval of TripClick

TripClick Dataset

- TripClick contains 1.5 million biomedical documents, 680k click-based training queries
- 3,525 test queries grouped by interaction frequency (Head, Torso, Tail)

🔍 Twin pregnancy

Planned caesarean section for women with a twin pregnancy. Background: twin pregnancies are associated with increased perinatal mortality, mainly related to prematurity, but complications during birth may contribute to perinatal loss or morbidity. the option of planned caesarean section to avoid such complications must therefore be considered. on the other hand, randomised trials of other clinical interventions in the birth process to avoid problems related to labour and birth (planned caesarean section for breech , and continuous electronic fetal heart rate monitoring), have shown an unexpected discordance between short - term perinatal morbidity and long - term neurological outcome. the risks of caesarean section for the mother in the current and subsequent pregnancies must also be taken into account . objectives : to determine the short - and long - term effects on mothers and their babies , of planned caesarean section for twin pregnancy.

Neural Re-ranking

- Re-creating the **training data** strongly improves re-ranking performance of TK
- Effectiveness results follow previous observations of **neural re-ranker**
- Domain-specific language modelling with an **Ensemble** shows performance improvements on all 3 test groups

			Head (DCTR)		*Torso (Raw)*		*Tail (Raw)*	
	Model	*BERT instance*	*nDCG*	*MRR@10*	*nDCG*	*MRR@10*	*nDCG*	*MRR@10*
Original Baselines	BM25	-	.140	.276	.206	.283	.267	.258
	TK	-	.208	.434	.272	.381	.295	.280
Our Re-Ranking	TK	-	.232	.472	.300	.390	.345	.319
	ColBERT	PubMedBERT	.278	.557	.340	.431	.387	.361
		SciBERT	.294	.595	.360	.459	.408	.377
	$BERT_{CAT}$	PubMedBERT	.296	.587	.359	.456	.409	.380
		Ensemble	**.303**	**.601**	**.370**	**.472**	**.420**	**.392**

Dense Retrieval

- **Dense retrieval** outperforms BM25 considerably
- **Judgement coverage** for the top-10 results

			Head (DCTR)		
Model	*BERT instance*	*J@10*	*nDCG@10*	*MRR@10*	*R@100*
BM25	-	31%	.140	.276	.499
	DistilBERT	39%	.236	.512	.550
$BERT_{DOT}$	SciBERT	41%	**.243**	**.530**	.562
	PubMedBERT	40%	.235	.509	**.582**

s_hofstaetter s.hofstaetter@tuwien.ac.at
sophiaalthammer sophia.althammer@tuwien.ac.at

sebastian-hofstaetter/tripclick