Round 2 results — Run cord19.vespa.ai-gbdt submitted from cord19.vespa.ai

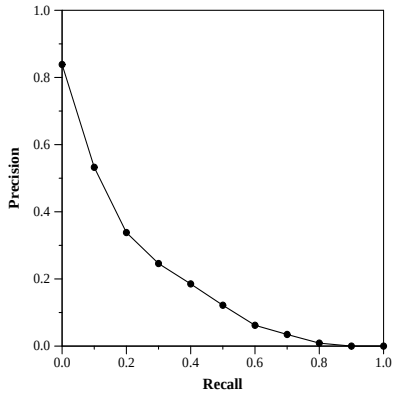
Run Description

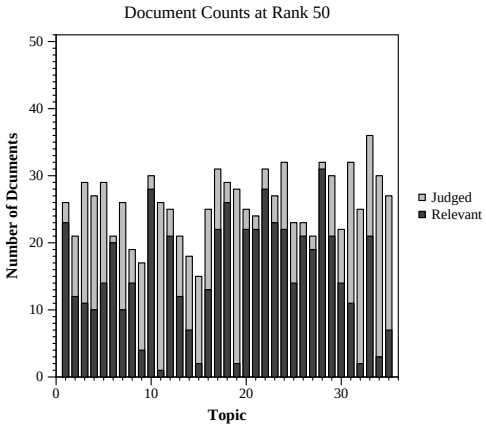
Query retrieval using bag of words from entity extraction using spacy en_core_sci_lg extracted from query, question and narrative. No query expansion or synonym expansion. nlp = spacy.load("en_core_sci_lg") ents = nlp(query + ' ' + question + ' ' + narrative).ents query = ' '.join([str(e) for e in ents]) BM25 retrieval and re-ranking 1K from first-phase (BM25) using a trained a GBDT model using lightGBM (https://docs.vespa.ai/documentation/lightgbm.html). The model uses 153 features from Vespa, ranging from matching features like bm25 in fields to semantic similarity (sent-scibert) over title and abstract using Vespa's nearest neighbor search operator. Parameters used with lightGBM: params = {'objective': 'lambdarank', 'metric': 'ndcg', 'ndcg_eval_at': '5,10', 'eta':0.01, 'num_leaves': 16, 'label_gain':"0,3,5"} More details on the run and how to reproduce https://github.com/vespa-engine/cord-19/trec-covid

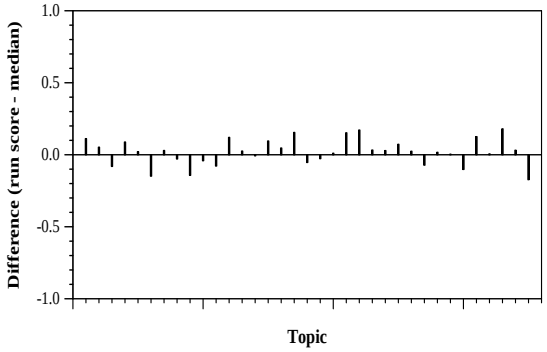
Summary Statistics	
Run ID	cord19.vespa.ai-gbdt
Topic type	feedback
Contributed to judgment sets?	yes

Overall measures	
Number of topics	35
Total number retrieved	35000
Total relevant	3002
Total relevant retrieved	1623
MAP	0.1884
Mean Bpref	0.3887
Mean NDCG@10	0.5422
Mean RBP(p=0.5)	0.6324 +0.0027

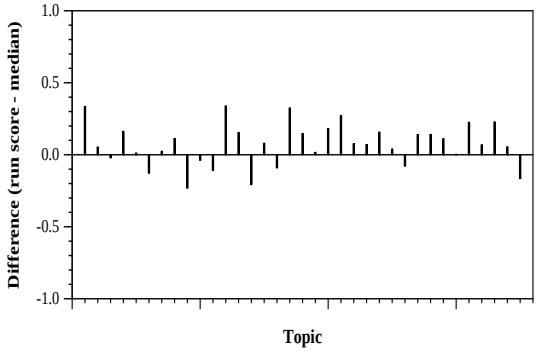
Document Level Averages	
	Precision
At 5 docs	0.6686
At 10 docs	0.5514
At 15 docs	0.5105
At 20 docs	0.4600
At 30 docs	0.3867
R-Precision	
Exact	0.2410





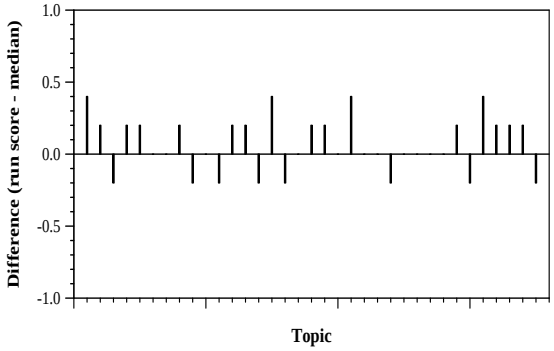


Per-topic difference from median bpref for all Round 2 runs

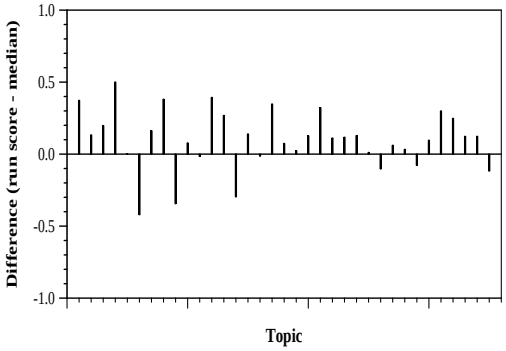


Per-topic difference from median NDCG@10 for all Round 2 runs

Round 2 results — Run cord19.vespa.ai-gbdt submitted from cord19.vespa.ai



Per-topic difference from median P@5 for all Round 2 runs



Per-topic difference from median RBP(p=0.5) for all Round 2 runs