

Round 2 results — Run udel_fang_L2R submitted from udel_fang

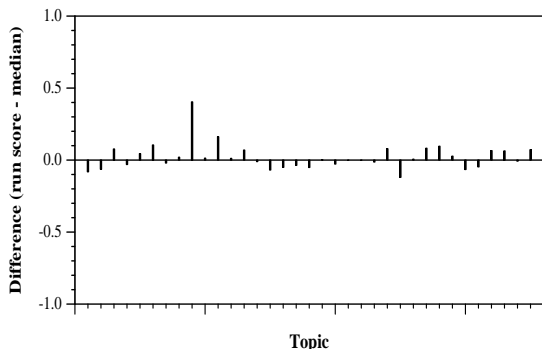
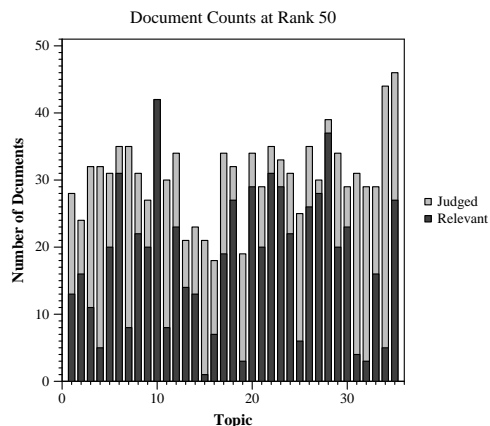
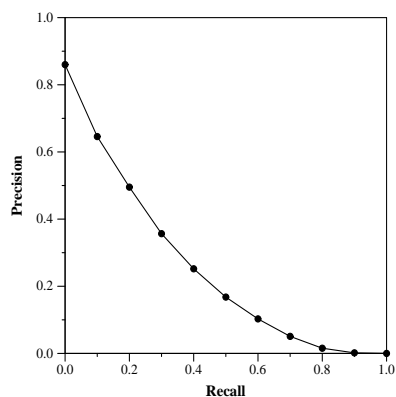
Run Description

We construct the keyword queries by using all non-stop words in the query field and all the entities tagged by SciSpacy in the question field. Then, we perform two baseline runs based on F2EXP, ranking title and abstract, and full text separately. After combining the features from the baseline runs, the keyword queries, and the metadata of all documents, we select the round one judgment file as the training data and use LambdaMART to re-rank the top 100 documents in the title and abstract baseline run. The best hyperparameters are chosen using cross-validation.

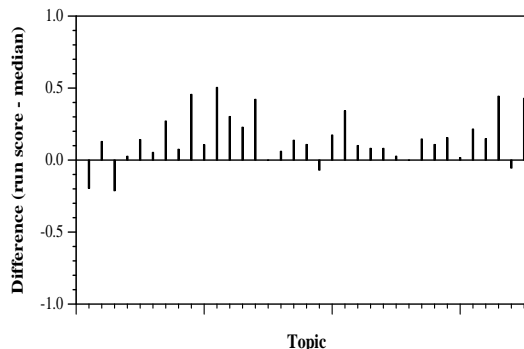
Summary Statistics	
Run ID	udel_fang_L2R
Topic type	feedback
Contributed to judgment sets?	no

Overall measures	
Number of topics	35
Total number retrieved	35000
Total relevant	3002
Total relevant retrieved	1713
MAP	0.2434
Mean Bpref	0.3904
Mean NDCG@10	0.6132
Mean RBP(p=0.5)	0.6426 +0.0193

Document Level Averages	
	Precision
At 5 docs	0.6857
At 10 docs	0.6571
At 15 docs	0.5886
At 20 docs	0.5457
At 30 docs	0.4657
R-Precision	
Exact	0.2892

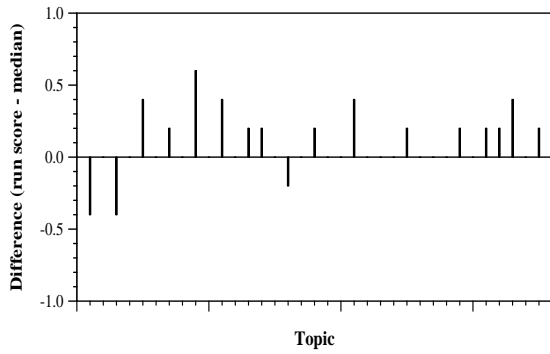


Per-topic difference from median bpref for all Round 2 runs

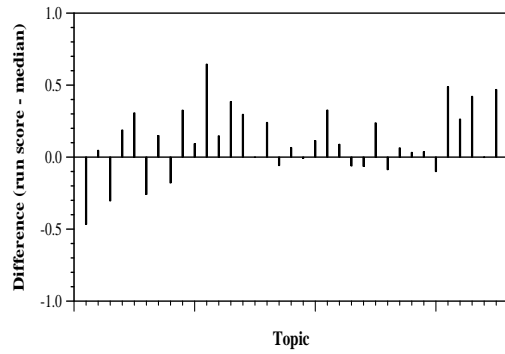


Per-topic difference from median NDCG@10 for all Round 2 runs

Round 2 results — Run udel_fang_L2R submitted from udel_fang



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