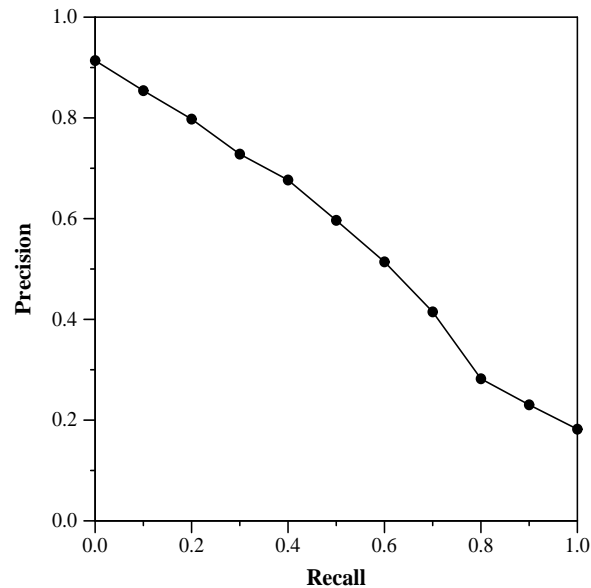
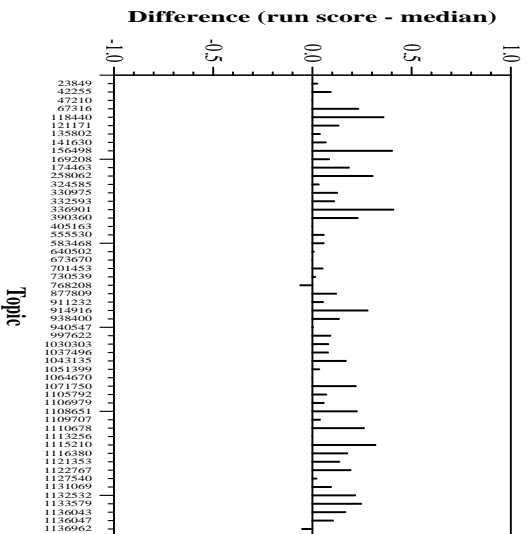


Deep Learning Track results — (PASH) PingAn Smart Health

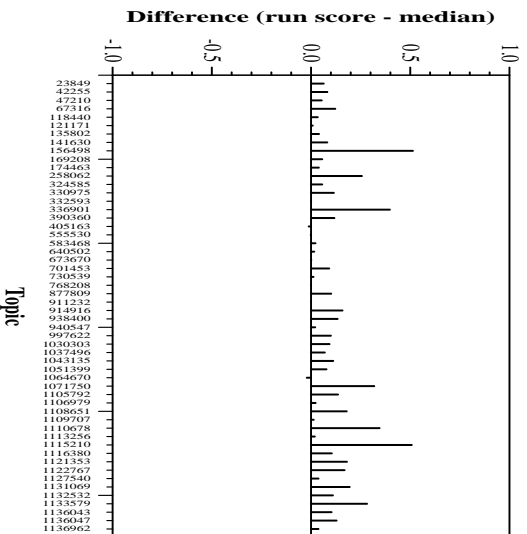
Run Description	
Run ID	pash_f3
Task	Passage Ranking
Subtask	Full ranking
Topic type	Automatic
Used deep nn model?	Yes
Type of training	Unsupervised data/model; Used passage task training labels
ORCAS data used for something other than training?	No
Training cost?	Pre-training stage takes about 16 days with 16 V100 GPUs, Multiple recall stage takes about many separate days. Fine-tuning stage takes about 30 hours with 4 V100 GPUs. Ensemble 10 models takes about 4 days.

Overall measures		Document Level Averages	
Number of topics	54	Precision	
Total number retrieved	54000	At 5 docs	0.7519
Total relevant	1666	At 10 docs	0.6481
Total relevant retrieved	988	At 15 docs	0.5284
MAP	0.5504	At 20 docs	0.4583
Mean NDCG@10	0.8005	At 30 docs	0.3617
Mean NDCG@1000	0.7120	R-Precision	
Mean Reciprocal Rank	0.8885	Exact	0.5416

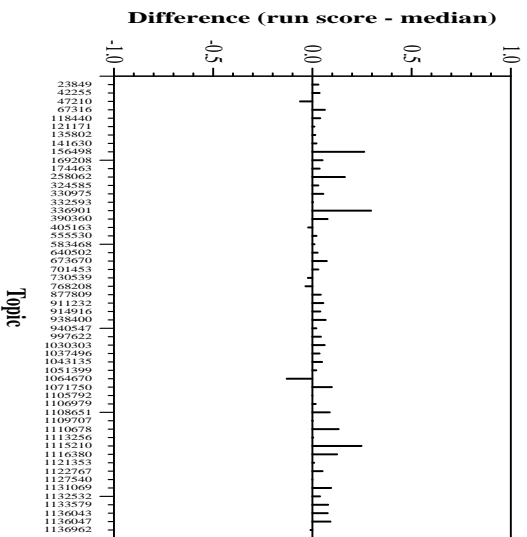




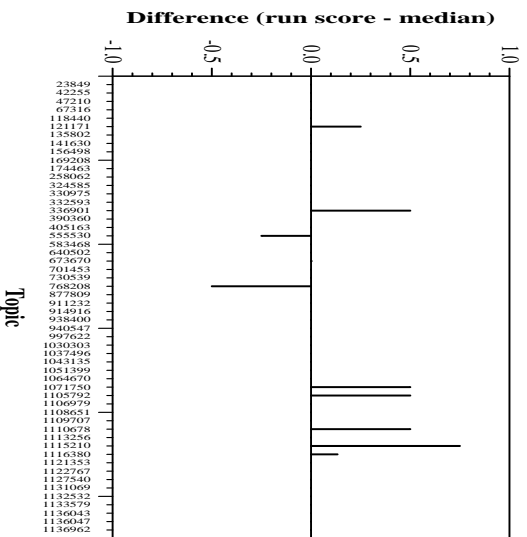
Per-topic difference from median for NDCG@10 for Passage Ranking runs



Per-topic difference from median for Average Precision for Passage Ranking runs



Per-topic difference from median for NDCG@1000 for Passage Ranking runs



Per-topic difference from median for Reciprocal Rank for Passage Ranking runs