Passage Reranking report

TrickEval, 12.11.2019

Base Model Used

We have used bert_large_cased as our base model and then fine tuned it for binary sentence classification

Methods

Experiments

1. We have varied our training data class distribution over a range of probabilities (1:100, 1:10)
2. We have experimented with varied sizes of training data (50,000 - 5,00,000 examples)
3. We have We have also tested the model by training for different number of epochs (1-4)
4. We also experimented with the final scoring metric based on the individual class scores
5. We also experimented with the batch sizes for training and since Google recommends a batch size of 32 or 64 (depending on memory) for training bert we have used 32.
6. For efficiency purposes the evaluation was split into 7 parts (since approx. 7000 queries) and carried out in parallel

Finalized model

1. We have observed that training for one epoch gives better result than for larger epochs.
2. We have also seen that larger number of training examples gives better results than repeating the same examples for multiple epochs.
3. For the final training set we have used a 1:0 ratio of 1:10 to overcome the skew of the data.
4. We have used exponential normalized scores over classes for scoring.
5. We have used a batch size of 32 for training and 1024 for evaluation.

6. Training one epoch on data takes about 40 mins on hpc.
7. Evaluating 1 of the 7 parts takes about 30-50 mins (highly varying in this range) on hpc.

Run instructions

- To build the model run ./build.sh from the submission home directory
- For reranking the script expects arguments in the format mentioned in the evaluation script. In addition to the two arguments given it also expects a third argument of top1000.eval.txt. Effectively, to the reranking script, the second argument should be query_file, the fourth argument should be the output file that is fed to trec and the sixth argument should be the top1000.eval.txt.