Testing the Effects of Pre-filtering on Learning to Rank Pipeline

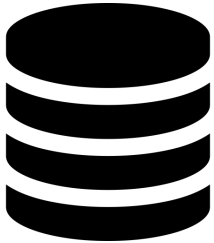
CSCI-539 Final Project

Jeffery Russell, Badah, Shin

Agenda

- Data Collection Used
- Methodology
- Experiment
- Preliminary Results
- Questions

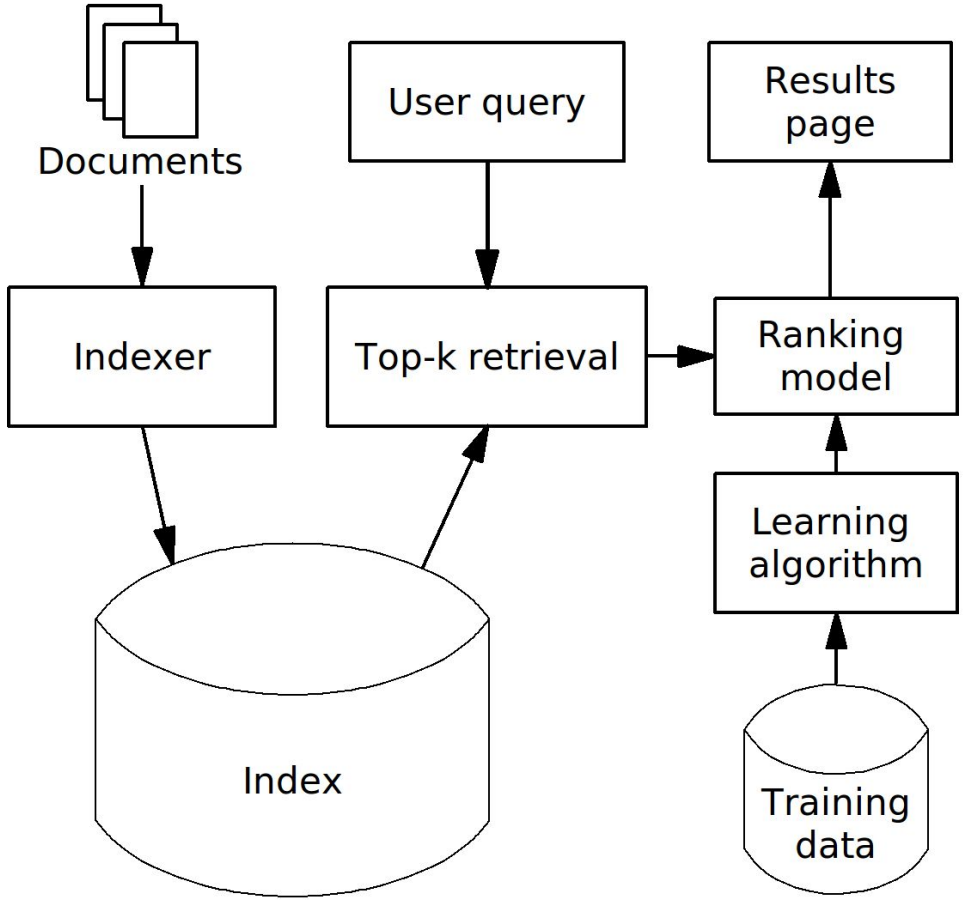
Dataset



- vaswani
  - Also known as the Vaswani and Cameron NPL test collection
  - Number of doc: 2083
  - Number of queries: 93
- trec-deep-learning-docs
  - a microsoft project from MS Marco from 2019
  - Number of doc: 16258
  - Number of queries: 200

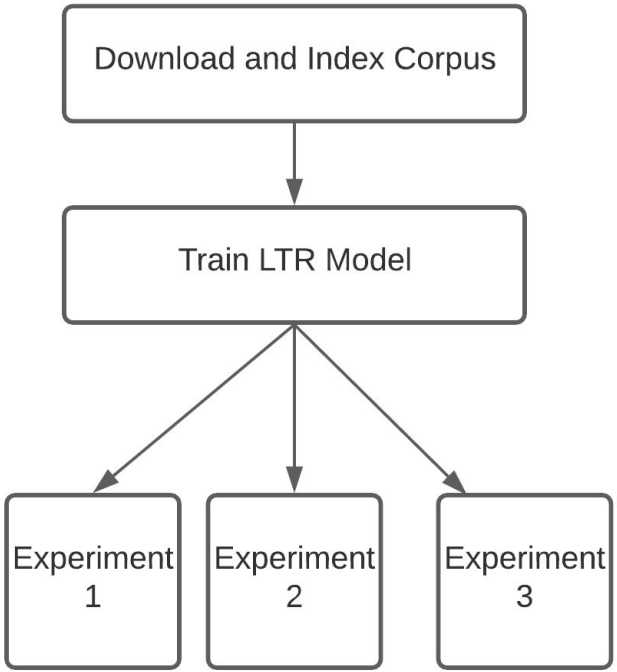
Learning to Rank Background

- A learning to rank pipeline has two ranking models:
  - Top-k retrieval: basic model to fetch potentially relevant documents
  - ML model that incorporates more features but is computationally expensive



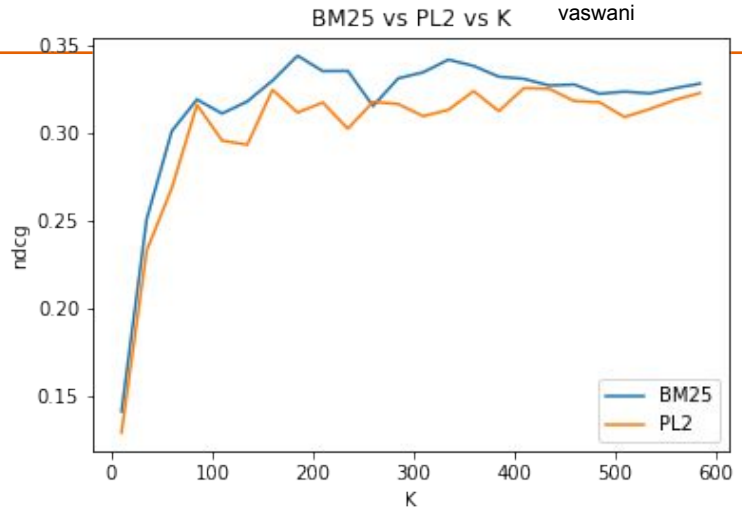
Methodology

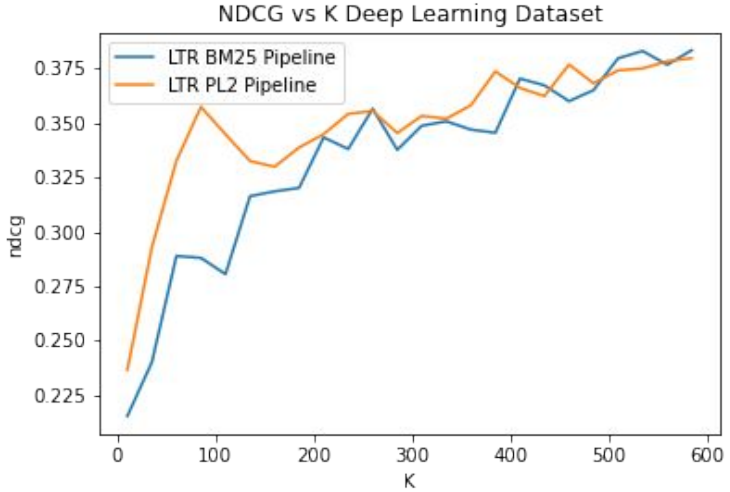
- Our objective of our experiment is to isolate and determine the effect that the top-k retrieval process has on end performance.
- Once a LTR model has been trained, it was used on all subsequent experiments.
  - The LTR model then gets paired with a top-k retrieval filter modified by the experiment.
- For purposes of this experiment, a Random Forest was trained using features from four different retrieval models



Effect of K on nDCG

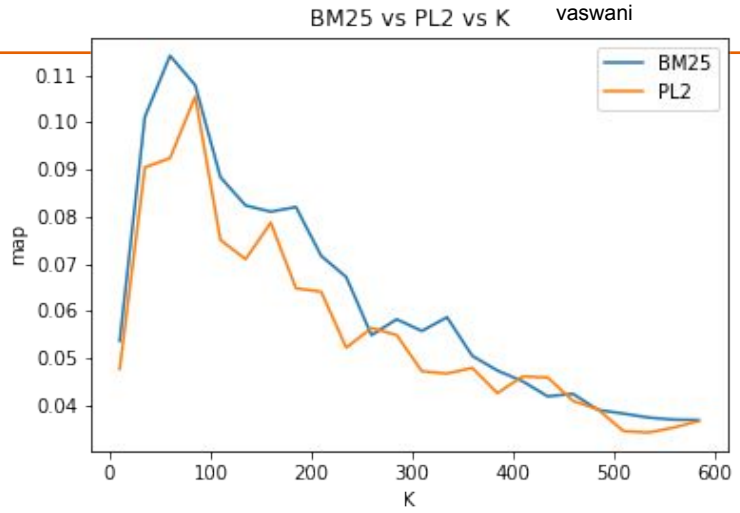
- In this experiment our trained model was used in a pipeline where we varied the number of documents returned by our top selection algorithm and we computed normalized Discounted Cumulative Gain (NDCG).
- For small k, the results varied until converting with larger k.

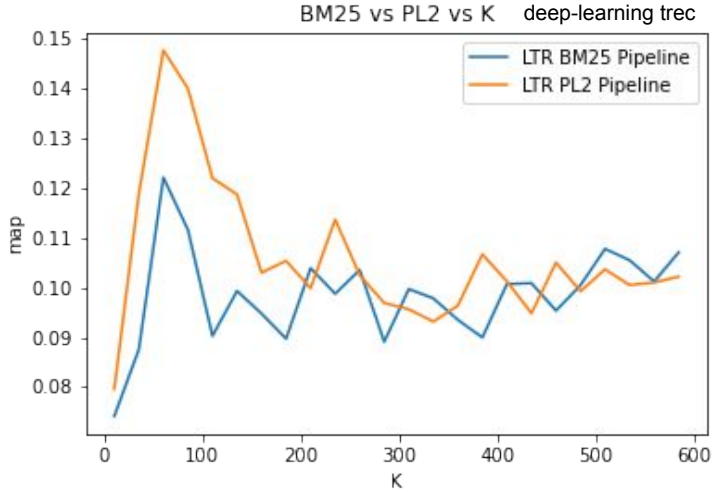




Effect of K on MAP

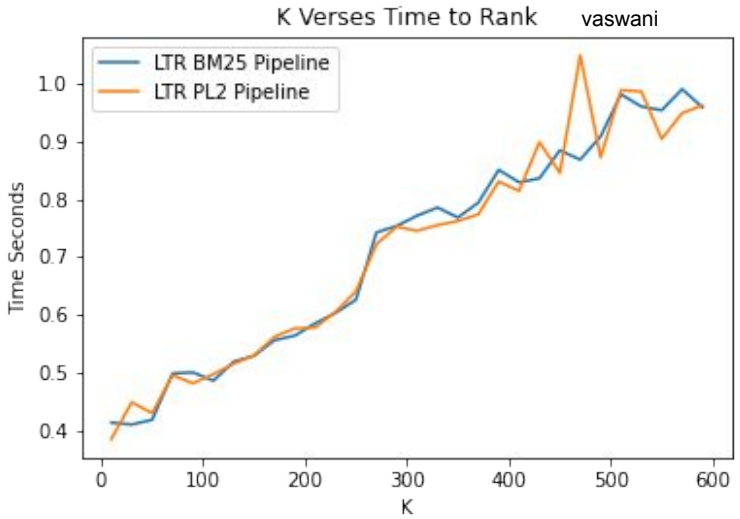
- In this experiment our trained model was used in a pipeline where we varied the number of documents returned by our top selection algorithm and we computed nDCG.
- Like last experiment, results converged with larger k.

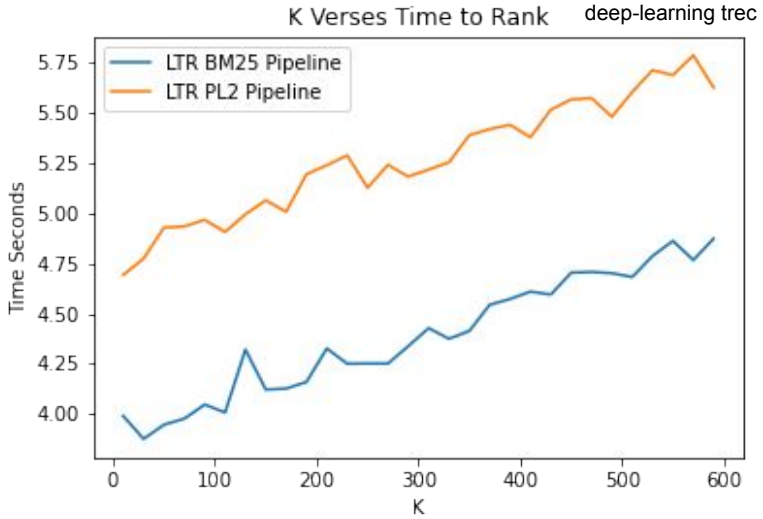




Effect of K on Execution Time

- In this experiment we time the time it takes to evaluate the dataset at varying levels of k.
- Intuitively makes sense that larger k's take longer since the ML model has to rank more documents.





Questions?