Paper Summaries

End-to-end neural ad-hoc ranking with kernel pooling

https://arxiv.org/pdf/1706.06613

Authors presented a kernel-based neural model for document ranking and named it K-NRM. The model includes a translation matrix that models word-level similarities via word embeddings, a kernel-pooling technique that extracts multi-level soft match features by using kernels, and combines the features into the final ranking score by using a learning-to-rank layer that uses pairwise ranking loss. The whole model is trained end-to-end. Query and document words are represented using distributed representations. Advantage of model is soft match achieved by the kernel-guided embedding learning. The dataset used is the search logs of Sogou.com, a Chinese commercial search engine.

Learning from Fact-checkers: Analysis and Generation of Fact-checking Language

https://dl.acm.org/doi/10.1145/3331184.3331248

The paper introduces an application to combat fake news by leveraging online fact-checkers; analyzing linguistic characteristics of fact-checking tweets; and using a deep learning framework to generate responses with fact-checking intention. The framework includes three components: shared word embedding layer; encoder to capture original tweet representation; decoder to generate a Fact Checking-tweet.

Learning to Rank with Selection Bias in Personal Search

https://dl.acm.org/doi/abs/10.1145/2911451.2911537

The paper focuses that various biases make it difficult to fully leverage click-through type of data and studies on how to leverage sparse click data in personal search. Bias estimation methods are proposed including their novel query-dependent method that captures queries with similar results and does not use a large quantity of click data for any given query-document pair and can successfully deal with sparse data. Also, the framework eliminates selection bias in personal search and provides an extensive empirical evaluation using large-scale live experiments.

From Royals to Vegans: Characterizing Question Trolling on a Community Question Answering Website

https://dl.acm.org/doi/10.1145/3209978.3210058

The paper focuses on a particular type of trolling: asking provocative questions on a community question-answering website i.e., questions aimed to inflame, upset, and draw attention from others on the community. Dataset was built from Yahoo Answers and then classification was performed on it using different classical machine learning algorithms.

Learning a Better Negative Sampling Policy with Deep Neural Networks for Search https://dl.acm.org/doi/10.1145/3341981.3344220

Sampling methods for documents selection for neural models deal with large class imbalances during training. Hence, careful selection of negative instances is necessary to avoid the risk of overfitting. The paper demonstrates that the method chosen for sampling negative documents is important for both the stability of training, as well as overall performance. Reinforcement learning approach is also used to optimize a policy over a set of sampling functions to improve performance over standard training practices with respect to IR metrics and is robust to hyperparameters and random seeds. Negative sampling problem has been formalized as a Markov Decision Process. Evaluation has been done by examining MAP of the top ranked 100 documents for Webscope L4 and the top 1000 for Robust04 datasets.