Paper Summaries

FAQ Retrieval using Query-Question Similarity and BERT-Based Query-Answer Relevance https://arxiv.org/pdf/1905.02851.pdf

The authors presented a Frequently Asked Questions retrieval system that considers similarity between a user's query and question; along with the relevance between the query and an answer. Similarity between the query and question is calculated using traditional unsupervised information retrieval system. QA pairs in a FAQ database are used to find the relevance between the query and answer by using BERT model. Evaluation is done on two datasets: localgovFAQ, a dataset in a Japanese administrative municipality domain; StackExchange dataset, public dataset in English. The proposed method outperforms baseline methods. MAP, MRR (Mean Reciprocal Rank), P@5, SR@k (Success Rate) and nDCG were used as the evaluation measures for the localgovFAQ dataset. MAP, MRR and P@5 were used for the StackExchange dataset.

Enquiring Minds: Early Detection of Rumors in Social Media from Enquiry Posts http://www-personal.umich.edu/~qmei/pub/www2015-zhao.pdf

Authors have proposed a technique to detect rumors or disputed factual claim: searching or identifying for the signal tweets(skeptical enquiries) using a set of regular expressions; clustering the signal tweets; detecting statement from each cluster to define the common text of the cluster; capturing all non-signal tweets matching any cluster's summary statement; rank the clusters by their likelihood of containing a rumor. Jaccard coefficient is used to measure similarity between tweets. Clusters are ranked by using classifiers: SVM and LIBSVM. Two datasets were used: BOSTON MARATHON BOMBING (BOSTON) and GARDENHOSE. Precision@N was used as the evaluation metric. In experimentation, about one-third of the tweets were real rumors, and about 70% of the top-ranked 10 clusters as rumors out of 50 candidate statements.

Learning similarity metrics for event identification in social media https://dl.acm.org/doi/10.1145/1718487.1718524

Authors evaluated a variety of techniques for learning multi-feature similarity metrics for social media documents. The task is to predict when social media documents correspond to the same event. Dataset used is of event images from Flickr. The learning techniques used were: ensemble-based, classification-based similarity with online clustering. NMI and B-cubed metrics were used for clustering. Baseline approaches were compared against the clustering approaches using four different similarity metric learning techniques: Ensemble-based approach combining partitions; Ensemble-based approach combining similarity scores; Similarity classifier using

Support VectorMachines; Similarity classifier using Logistic Regression. The classification based techniques resulted in significant improvement over text-based similarity, traditional approach.

Towards Conversational Recommender Systems

https://dl.acm.org/doi/10.1145/2939672.2939746

A framework is proposed to identify the questions to ask a new user to quickly learn their preferences, like dining experiences. The framework uses a probabilistic latent factor model and experiments with synthetic and real-world data to compare different feedback and question selection strategies. The framework using online user feedback improves recommendations over a static model by 25% after asking only two questions. A bandit-based approach for online recommendation is applied to restaurant recommendation. Generalized Thompson Sampling is used to sample the questions to ask a user and to incorporate observed feedback and user-item embeddings are learnt through interactions. Absolute and pairwise models are used to estimate the latent variables.

Spam Filter Evaluation with Imprecise Ground Truth

https://plg.uwaterloo.ca/~gvcormac/cormacksigir09-spam.pdf

The labels acquired from user feedback exhibit higher error rates than the best spam filters. Authors presented automatic and semi-automatic methods to reduce the influence of labelling errors on evaluation. The automatic method requires no access to the messages, while the semi-automatic method requires adjudication of a small fraction. Authors hypothesize that results after applying several filters to the same training data on fusion to form a set of labels has a lower error rate than the natural labels when used as ground truth for evaluation. The hypothesis is tested in several ways: direct comparison between pseudo-gold and gold standard labels; adjudication of differences between pseudo-gold and natural labels, using an independent adjudicator; comparison of performance measures (AUC and LAM); comparison of filter performance rankings; measuring the statistical power to discriminate between pairs of filters. Email messages from two separate corpora: the TREC 2005 Public Spam Corpus and the CEAS 2008 Live Challenge private corpus were used. Filters that return both hard result(ham or spam) and the soft result ("spaminess" score) is used. Authors through experiments evaluated the hypothesis to be true. Noise-tolerant spam filters perform not quite as well as the best filters with clean data.

Investigating Passage-level Relevance and Its Role in Document-level Relevance Judgment

http://www.thuir.cn/group/~YQLiu/publications/SIGIR2019Wu.pdf

The authors investigate the role of passage-level relevance in the document-level relevance judgment. An ad-hoc retrieval dataset with both passage-level and document-level relevance labels is constructed. Based on the relationship between passage-level and document-level relevance, authors also show that utilizing passage-level relevance signals can improve existing document ranking models. Corpus selected is THUCNews 3, a Chinese news dataset and queries are select queries from a Chinese search engine. BM25 score for each query-document pair is calculated and the documents are ranked. Two settings are tested in passage-level relevance annotation: Context-aware passage-level relevance annotation(CRA) which shows the whole document to the assessor to make relevance judgments for all passages within it; Independent passage-level relevance annotation (IRA): only one passage is showed to make a relevance judgment for it. The analysis reflects that (1) There is a strong correlation between the document-level relevance and the fractions of irrelevant passages to highly relevant passages; (2) Document-level relevance can be determined by position, length and query similarity of passages; (3) Sequential passage-level relevance within a document is a potential indicator for the document-level relevance.