

ECNU at 2018 eHealth Task 2: Technologically Assisted Reviews in Empirical Medicine

Huaying Wu¹, Tingting Wang¹, Jiayi Chen¹, Su Chen¹, Qinmin Hu², and Liang He¹

¹ Department of Computer Science & Technology, East China Normal University, Shanghai, 200062, China

² Department of Computer Science, Ryerson University, Toronto, Canada
{hywu,ttwang,jychen,schen}@ica.stc.sh.cn, vivian@ryerson.ca,
lhe@cs.ecnu.edu.cn

Abstract. The 2018 CLEF eHealth Task 2 has two sub-tasks in order to write a systematic review of evidence-based medicine. Researchers are required to retrieve relevant documents given by medical database for each query (sub-task 1) and re-rank the documents with the results of the Boolean search as the starting point(sub-task 2). We adopt BM25 with query expansion to acquire basic relationship and utilize a customized Paragraph2Vector to represent queries / documents trained by the training set of Boolean search. To compute the relevant score of given query-document pair, cosine similarity and logistic regression are taken in our experiments. Finally, we find that the combination has a better performance.

Keywords: Query expansion, Paragraph2Vector, Health Information Retrieval

1 Introduction

The ECNUica participates in CLEF 2018 eHealth Task 2 : TAR in Empirical Medicine, which proposes to do a sorting problem based on query-documents similarity in Systematic Reviews. There are multiple stages contained in Systematic Reviews: Boolean Search in each query, Screening queries title and Abstract, and Document Checking. The task focus on the first and second stages of the process.

In the Boolean Search stage, Participants need to do a basic binary classification for each document based on every query. Boolean query with relevant information constituted, which submits to a medical database containing details of medical studies built by experts, need to be classified into relevant or irrelevant. The database returns a set of potential relevant studies. In the following steps, Participants decide which ones are indeed relevant by screening titles, abstracts and full documents.

There are two sub-tasks for this task. One is to find documents in high relevance for each given query. The other one is to re-rank the documents retrieved in the first step given by experts. According to our work in past two years, we try to manage the data with learning-to-rank [6], word2vec [1, 7], relevance based relation between a query and documents [3], which perform well in other runs [4, 5, 7].

2 Methods

In sub-task 1, we choose BM25 algorithm to acquire a baseline of Boolean search. Furthermore, query expansion based on MeSH and pseudo relevance feedback (PRF) is taken to get a better result. In sub-task 2, we employ Paragraph2Vector to represent query and documents for similarity calculation.

2.1 Boolean search

Query Expansion In this stage, we do query expansion to improve retrieval precision. For better performance of experiments, we compare the expansion with PRF, MeSH and RPF + MeSH.

- The PRF returns top-10 relative features for each query.
- The MeSH database is applied to extract medical terms from titles.

We choose DescriptorName part from raw data as keywords of document, which describes theme of document with a series of words, and the words form MeSH as expansion. We do not use any part of protocol in both tasks. Thus, the query we use in both tasks contains: title and objective from original query, expansion from DescriptorName or MeSH. The results show that both PRF and MeSH can improve performance.

Model Training In the model selection stage, we compare the result of BM25, DRF_BM25 and PL2. For each algorithm, experiment is based on method (B-M25, DEF_BM25, PL2) only, method with PRF, method with MeSH, method with both PRF and MeSH. One-hot is used to represent every query for relative score calculation.

2.2 Ranking

Paragraph2Vector T. Mikolov proposed paragraph vector [2], which presented an unsupervised algorithm that learns fixed-length pieces of texts. With this method, we use Paragraph2Vector Model to represent all selected documents from words to fixed-length vector.

Under this framework, we should know how to learn vector representation of words first [1]. The objective of the word vector model is to maximize the average log probability.

After training word vectors, we use softmax function as activate function to learn the softmax weights and paragraph vectors on documents.

Logistic Regression With all CLEF 2017 eHealth training and testing queries and CLEF 2018 eHealth training queries as training dataset, we train a logistic regression model as a classifier. For each document given query, calculation about the relationship with the LR classifier is taken care. The text is the input of model while return a score of relevance.

3 Experiments

3.1 Dataset

For sub-task 1, we are provided with a test set consisting of 20 topics of Diagnostic Test Accuracy (DTA) reviews as follows.

- Topic-ID.
- The title of the review, written by Cochrane experts.
- A part of the protocol.
- The entire PubMed database

For sub-task 2, we are provided with different data in the same reviews as follows.

- The Boolean query manually constructed by Cochrane experts
- The set of PubMed Document Identifiers (PID's) returned by running the query in MEDLINE.

For training, we choose the CLEF eHealth 2017 queries and documents with all training part and testing part, CLEF eHealth 2018 queries and documents with training part.

3.2 Runs

We submit three runs for each sub-task whose descriptions are as follows.

In Sub-task 1:

- ECNU_TASK1_RUN1_BM25: The result retrieved on entire PubMed dataset by terrier platform with BM25 model and pseudo relevance feedback.
- ECNU_TASK1_RUN2_LR: Rerank all documents by a Logistic Regression classifier and Paragraph Vector.
- ECNU_TASK1_RUN3_COMBINE: A combination of previous two runs.

In Sub-task 2:

- ECNU_TASK2_RUN1_TFIDF: Rerank the pids by vector space model. Each document is represented as a vocabulary-size vector. Each dimension is the tf-idf score of a certain word. We use cosine similarity to rerank the document.
- ECNU_TASK2_RUN2_LR: a Logistic Regression classifier is used to rerank documents based on Paragraph Vector.
- ECNU_TASK2_RUN3_COMBINE: A combination of previous two runs.

Method	Target	Score
Task1 BM25	ap	0.072
	recall@1000	0.561
	rels_found	426
	num_rels	759
Task1 LR	ap	0.041
	recall@1000	0.408
	rels_found	310
	num_rels	759
Task1 Combine	ap	0.072
	recall@1000	0.561
	rels_found	426
	num_rels	759
Task2 TFIDF	ap	0.142
	recall@100%	0.994
	num_rels	3964
Task2 LR	ap	0.081
	recall@100%	0.779
	num_rels	3964
Task2 Combine	ap	0.146
	recall@100%	0.992
	num_rels	3964

Table 1. Evaluations in Task1 and Task2

Summary of Runs The run-3 of each sub-task shows better performance in training. We list some results in Table 1, which shows that BM25+PRF perform best compared to other methods. Both MeSH and PRF are employed for query expansion. But during experiments, performance declines when we take them simultaneously.

Experiments choose ap, recall@100, rels_found, num_rels as evaluation metrics, where ap presents average precision in documents, recall@100 shows the recall score at top-100 documents, num_rels reveals the number of total recalled documents and rels_found in sub-task1 shows the number of documents we find in the experiments.

4 Conclusions and Future Work

In the CLEF eHealth 2018 Task 2 TAR, ECNUica team take advantages of the Paragraph2Vector model. Combining with Statistical method, logistic regression with TF-IDF shows better performance, compared to LR method only or TF-IDF only. Although the representation of queries and documents can be taken to compute similarities by cosine distance, there are many aspects of our method which need improvement. In the future work, we will focus on more features from text and better out methods.

5 Acknowledgement

We thank reviewers for their review comments on this paper.

References

1. Mikolov T, Sutskever I, Chen K, et al. Distributed Representations of Words and Phrases and their Compositionality[J]. *Advances in Neural Information Processing Systems*, 2013, 26:3111-3119.
2. Le Q V, Mikolov T. Distributed Representations of Sentences and Documents[J]. 2014, 4:II-1188.
3. Lavrenko V, Croft W B. Relevance based language models[C]// *International ACM SIGIR Conference*. 2001:120-127.
4. Yang S., He Y., Hu Q.M., He L., Haacke E.M.: ECNU at 2015 eHealth Task 2: User-centred Health Information Retrieval. *Proceedings of the ShARe/CLEF eHealth Evaluation Lab (2015)*.
5. Yang S., He Y., Hu Q.M., He L.: ECNU at 2015 CDS Track: Two Re-ranking Methods in Medical Information Retrieval. *Proceedings of the 2015 Text Retrieval Conference*. (2015).
6. Yang S., He Y., Hu Q.M., He L., Liu H.Y., Wang Y.Y., Luo G.h.:ECNU at 2016 eHealth Task 3: Patient-centred Information Retrieval.(2016)
7. Chen j.y., Chen S., Song Y., , Liu H.Y., Wang Y.Y., Hu Q.M., He L., and Yang Y.: ECNU at 2017 eHealth Task 2: Technologically Assisted Reviews in Empirical Medicine. (2017)
8. Suominen, Hanna and Kelly, Liadh and Goeriot, Lorraine and Kanoulas, Evangelos and Azzopardi, Leif and Spijker, Rene and Li, Dan and Nvol, Aurlie and Ramadier, Lionel and Robert, Aude and Zuccon, Guido and Palotti, Joao. Overview of the CLEF eHealth Evaluation Lab 2018. *CLEF 2018 - 8th Conference and Labs of the Evaluation Forum, Lecture Notes in Computer Science (LNCS)*, Springer, September, 2018.
9. Kanoulas, Evangelos and Spijker, Rene and Li, Dan and Azzopardi, Leif. CLEF 2018 Technology Assisted Reviews in Empirical Medicine Overview. *CLEF 2018 Evaluation Labs and Workshop: Online Working Notes, CEUR-WS*, September, 2018.