FDUMedSearch at TREC 2015 Clinical Decision Support Track

Ronghui You[1,2], Yuanjie Zhou[3], Shengwen Peng[1,2], Shanfeng Zhu[1,2,*]

{15210240027,zhusf}@fudan.edu.cn

1 School of Computer Science, Fudan University, Shanghai 200433, P. R. China,
2 Shanghai Key Lab of Intelligent Information Processing, Fudan University, Shanghai 20433, P. R. China,
3 Shanghai Children's Medical Center, Medical School Shanghai Jiaotong University, 1678 Dongfang Road, Shanghai, 200127, China,

Abstract. This paper describes the participation of the FDUMedSearch team at TREC 2015 Clinical Decision Support track (CDS2015). Given the medical cases, the main purpose of CDS2015 is to develop effective information retrieval techniques in finding relevant documents for patient care. We used Indri as the retrieval engine, which implemented query likelihood method as the baseline. In addition, query expansion using Medical Subject Headings (MeSH), pseudo relevance feedback and classification were used to enhance the retrieval performance. We also tried to extract keywords in two different ways, automatically and manually. Experimental results show that our method achieved significant improvement over baseline methods.

Keywords. Biomedical Information Retrieval; Query Expansion; Pseudo Relevance Feedback

1 Introduction

TREC Clinical Decision Support track 2015 (CDS2015) focuses on linking PubMed Central (PMC) articles to the medical cases for patient care. There are 30 topics with both summary and description. These topics belong to three categories: Diagnosis, Test, and Treatment (10 topics per category). CDS2015 consists of two rounds of evaluation, Task A and Task B. Different from Task A, Task B provides a diagnosis field to the participants in Test and Treatment topics. In each task, we can upload at most three submissions. In each submission, only the summary or description of the topics can be used. The query can be constructed automatically or manually.

2 Methods

Here we summarize the information retrieval (IR) models and techniques used in our system.

[*]Corresponding author

2.1 Query Likelihood Model

We used Indri[1] as the retrieval engine and unigram query likelihood model [1] to get the relevant articles as the baseline. We adjusted the smoothing parameter λ to fit the long text. Each topic has both description and summary. The description has more information then summary, but it may contain many useless terms. So we used summary to construct the query in the baseline.

2.2 Keyword Extraction

To formulate the query automatically, we used a biomedical concept annotation tool[2] to extract the concepts of each topic as the keywords. Since the auto method may miss some important information, we further asked a doctor to help us to extract important keywords in the description of each topic in manual setting.

2.3 MeSH Terms Query Expansion

MeSH has been widely used in improving biomedical information retrieval [2-5]. We used the query to obtain the relevant citations in MEDLINE. The MeSH terms which appear in top retrieved citations are used in query expansion. For each topic, we used 30 MeSH Terms. To explore the effect of Major MeSH terms, we also try the setting of using Major MeSH terms only in query expansion.

2.4 Pseudo Relevance Feedback

Pseudo relevance feedback is a widely used technique in information retrieval. We used Top-K documents to carry out pseudo relevance feedback in our system. In general, based on the experimental results on CDS2014 dataset, k was set to 3 or 8 in our system.

2.5 Classifier

Previous study in CDS2014 found that classifying retrieved articles into diagnosis and treatment category could improve the searching performance [6]. Similarly, we train a text classifier using TF-IDF word features based on Clinical Hedges database [7]. The Clinical Hedges database consists of around 49000 documents, which were labeled by 8 categories, such as Therapy, Diagnosis, Prognosis, Reviews, Clinical Prediction Guide, Qualitative, Causation (etiology) and Economics. We focus on the treatment category. We used the classifier to score the retrieved documents, and re-ranked the documents based on searching and classifying scores.

---

[1] http://www.lemurproject.org/indri.php
[2] http://bioportal.bioontology.org/annotator

3 Experimental Settings and Results

3.1 The IR Techniques Used in Different Submissions

In each task, we uploaded three submissions with different configurations of IR techniques described in Section 2. As shown in Table 1, we used different configurations for different topic types. For example, the setting of FDUManual2 submission for Test topic in Task B is "Manual keywords; Major MeSH; Feedback; Manual diagnosis;". That is to say, the searching keywords were first constructed by the doctor. We used major MeSH terms in query expansion, as well as pseudo relevance feedback. Finally, manual diagnosis was also added to formulate the query.

Table 1, A summary of information retrieval techniques used in all 6 submissions by FDUMedSearch in the Task A and B.

Task	Submission	Diagnosis	Test	Treatment
Task A	FDUAuto1 Auto Summary	Auto keywords Major MeSH Feedback	Auto keywords Major MeSH Feedback	Auto keywords All MeSH Feedback Classifier
Task A	FDUAuto2 Auto Summary	Auto keywords Major MesH Feedback	Auto keywords All MesH Feedback	Autokeywords All MesH Feedback Classifier
Task A	FDUManual Manual Description	Manual keywords Major MeSH Feedback	Manual keywords Major MeSH Feedback	Manual keywords All MeSH Feedback Classifier
Task B	FDUAuto Auto Summary	Auto keywords Major MeSH Feedback	Autokeywords All MeSH Feedback Given diagnosis	Auto keywords All MeSH Feedback Classifier Given diagnosis
Task B	FDUManual1 Manual Description	Manual keywords Major MeSH Feedback	Manual keywords All MeSH Feedback Given diagnosis	Manual keywords All MeSH Feedback Classifier Given diagnosis
Task B	FDUManual2 Manual Description	Manual keywords Major MeSH Feedback Manual diagnosis	Manual keywords Major MeSH Feedback Manual diagnosis	Manual keywords All MeSH Feedback Classifier Manual diagnosis

3.2 Results

As shown in the Table 2, we present the overall performance of different submissions and baseline methods in terms of infNDCG, infAP, P@10 and R-prec. From the experimental result, we can see that all submissions outperform the baseline method significantly in both task A and B, which demonstrate the effectiveness of using IR techniques. In Task A, the best performed submission is FDUManual1 with an infNDCG of 0.2689, while the baseline method achieved an infNDCG of 0.2147. On the other hand, the best performed submission in Task B is FDUManual2 with an infNDCG of 0.3809, while the baseline method achieved an infNDCG of 0.3222.

As illustrated in Table 3, we further checked the performance of different submissions on each type of topic in terms of infNDCG. Overall FDUManual and FDUManual2 achieved good performance in every topic type, respectively. However, a notable exception is the Diagnosis type of Task A. FDUManual achieved the lowest infNDCG of 0.1901, which is even lower than the baseline method (0.2296). This suggests that the keywords extracted by the doctor work very poorly in the Diagnosis type in our submission.

Table 2, The overall performance of different submissions and baseline methods in both Task A and B

Task	Submission	infNDCG	infAP	P@10	R-prec
Task A	Baseline	0.2147	0.0438	0.3578	0.1811
Task A	FDUAuto1	0.2469	0.0599	0.3900	0.1847
Task A	FDUAuto2	0.2539	0.0600	**0.3933**	0.1889
Task A	FDUManual	**0.2689**	**0.0611**	0.3900	**0.1916**
Task B	Baseline	0.3102	0.0752	0.4689	0.2447
Task B	FDUAuto	0.3222	0.0766	0.4967	0.2246
Task B	FDUManual1	0.3288	0.0820	0.5100	0.2476
Task B	FDUManual2	**0.3809**	**0.1008**	**0.5600**	**0.2768**

Table 3, TheinfNDCG performance of different submissions and baseline methods in both Task A and B by topic types.

Task	Submission	Diagnosis	Test	Treatment	All
Task A	Baseline	0.2296	0.1694	0.2450	0.2147
Task A	FDUAuto1	**0.2756**	0.1769	0.2880	0.2469
Task A	FDUAuto2	0.2468	0.2179	0.2969	0.2539
Task A	FDUManual	0.1901	**0.2825**	**0.3340**	**0.2689**
Task B	Baseline	0.2296	0.3238	0.3772	0.3102
Task B	FDUAuto	0.2468	0.3394	0.3803	0.3222
Task B	FDUManual1	0.1901	0.3844	**0.4118**	0.3288
TaskB	FDUManual2	**0.3450**	**0.3860**	**0.4118**	**0.3809**

4 Discussion and Conclusion

From experimental result we can see that IR techniques are very helpful in improving the performance of medical information retrieval. In addition, manual keywords and diagnosis suggested by the domain expert are usually very helpful in boosting the searching performance. Nevertheless, we also find that unsuitable manual keyword would deteriorate the performance greatly. The strategies we used in CDS2015 were learnt from the CDS2014. Due to the small size of available topics, it is not surprisingly some strategies do not work very well in CDS2015. In the future, we will continue explore the optimal strategy for medical information retrieval.

5 Acknowledgement

This work has been partially supported by National Natural Science Foundation of China (61572139), and Scientific Research Starting Foundation for Returned Overseas Chinese Scholars, Ministry of Education, China.We would like to thank Prof. Brian Haynes for providing us Clinical Hedges database.

6 Reference

[1] Zhai C. Statistical Language Models for Information Retrieval: A Critical Review. Foundations and Trends in Information Retrieval **2**(3): 137-213 (2008)

[2] Liu K, Peng S, Wu J, Zhai C, Mamitsuka H, Zhu S. MeSHLabeler: improving the accuracy of large-scale MeSH indexing by integrating diverse evidence. Bioinformatics 31(12): 339-347 (2015)

[3] Zhu S, Zeng J, Mamitsuka H. Enhancing MEDLINE document clustering by incorporating MeSH semantic similarity. Bioinformatics 25(15): 1944-1951 (2009)

[4] Zhu S, Takigawa I, Jia Zeng, Mamitsuka H. Field independent probabilistic model for clustering multi-field documents. Inf. Process. Manage. 45(5): 555-570 (2009)

[5] Gu J, Feng W, Zeng J, Mamitsuka H, Zhu S. Efficient Semisupervised MEDLINE Document Clustering With MeSH-Semantic and Global-Content Constraints. IEEE T. Cybernetics 43(4): 1265-1276 (2013)

[6] Choi S, Choi J:SNUMedinfo at TREC CDS track 2014: Medical case-based retrieval task. TREC 2014

[7] Wilczynski NL, Morgan D, Haynes RB and the Hedges Team. An overview of the design andmethods for retrieving high-quality studies for clinical care. BMC Med Inform Decision Making. 5:20 (2005)