PKUICST at TREC 2015 Clinical Decision Support Track

Ziwei Zheng[a,b], Xiaojun Wan[a]
[a]Institute of Computer Science and Technology, Peking University
[b]School of EECS, Peking University
Haidian District, Beijing 100871, China
micheal_zzw@pku.edu.cn, wanxiaojun@pku.edu.cn

Abstract

The tasks at TREC 2015 Clinical Decision Support Track involved retrieval and ranking of medical journal articles for medical records of diagnosis, test or treatment. The PKUICST team participated in Task A and experimented with different strategies. In this report, we briefly introduce our techniques and results in the 2015 CDS task.

1 Introduction

The TREC 2015 Clinical Decision Support (CDS) track was designed to assess the ability of search engines to retrieve biomedical journal articles relevant for answering generic clinical questions about medical records [1]. Each topic in this track consists of a sentence-long summary and a paragraph-long description of a patient case, along with one of three types of clinical information: diagnosis, test or treatment, which is likely to support physicians and other health professionals with clinical decision-making tasks by medical records such as a list of symptom, a treatment plan and a particular test. The CDS task aims to retrieve articles to particular case based on the summary, description, or both.

The document collection for the TREC CDS 2015 track is a snapshot of the PubMed Central (PMC) Open Access Subset on January 21, 2014, which is a free archive of biomedical and life sciences journal literature at the U.S. National Institutes of Health's National Library of Medicine (NIH/NLM). The corpus contains the abstracts, full texts and other metadata of 733138 articles in the biomedical domain and the articles were furnished in NXML format. CDS participants were invited to submit up to 3 sets of ranked documents deemed relevant to 30 topics, 10 for each of the three types of information need.

The PKUICST team used a paragraph-based retrieval method for the task. A biomedical article is usually very long and it seldom focuses on a medical record, but one or several paragraphs may be focused on the medical record. An article with one or more paragraphs related to the medical record might be useful for the physicians. In most full-text retrieval models, the features of the whole article, like TF-IDF, always reduce the influence of such useful paragraphs. In order to emphasize the relevance of such paragraphs, we used a text-segmentation strategy to segment each article and evaluate the relevance of an article based on the relevance of its paragraphs, which will be introduced in Section 2.2.

A medical record describes a list of symptom(diagnosis), a treatment plan(treatment) or a particular test(test), but the major concept concerned by the physicians is the pathema which seldom appears in the record. In order to address this problem we enriched topics with Medical Subject Headings (MeSH) terms which have been used by some group [2] in TREC CDS 2014. MeSH is the National Library of Medicine's (NLM) controlled vocabulary thesaurus. It consists of sets of terms naming descriptors in a hierarchical structure that permits searching at various levels of specificity. An example will be showed in Section 2.1.

We officially submitted three runs. All retrieval lists were generated with Terrier [3] which is an open source, comprehensive, flexible and transparent platform for research and experimentation in text retrieval. Research can easily be carried out on standard TREC and CLEF test collections by Terrier.

2 Our Strategy

2.1 Topic Enrichment

As an example topic below, we created two fields <MeSHSummary> and <MeSHDescription> by extracting MeSH words which appear in the <summary> and <description> fields of each topic.

```
<topic>
<type>diagnosis</type>
<description>
    A 44-year-old male is brought to the emergency room after multiple bouts of vomiting that has a
"coffee ground"; appearance. His heart rate is 135 bpm and blood pressure is 70/40 mmHg. Physical
exam findings include decreased mental status and cool extremities. He receives a rapid infusion of
crystalloid solution followed by packed red blood cell transfusion and is admitted to the ICU for further
care.
 </description>
 <MeshDescription>
    Vomiting ; Heart ; Heart Rate ; Blood ; Blood Pressure ; Pressure ; Extremities ; Blood ;
  </MeshDescription>
  <summary>
    A 44-year-old man with coffee-ground emesis, tachycardia, hypoxia,  hypotension and cool,
clammy extremities.
  </summary>
  <MeshSummary>Tachycardia ; Hypotension ; Extremities ; </MeshSummary>
  <kwd>Vomiting ; Blood Pressure ; Tachycardia ; Hypotension ;</kwd>
  </topic>
```

In Mesh 2015, MeSH words are organized into sixteen major categories [4] which might be related to different types (diagnosis,treatment and test) of topics. We selected seven categories which correspond to the tree types of topics, divided them into three groups manually (showed in Table 1) and further created a <kwd> field by extracting the topic words of the categories corresponding to the given topic. For example, in the above topic, 'Vomiting', 'Tachycardia' and 'Hypotension' belong to category E and 'Blood Pressure' belongs to category C. We used all of these fields for retrieval.

Diagnosis	Treatment	Test
Diseases [C] ; Analytical, Diagnostic and Therapeutic Techniques and Equipment [E]	Chemicals and Drugs [D] ; Health Care [N]	Anatomy [A] ; Organisms [B] ; Phenomena and Processes [G]

Table 1 Classification of MeSH 2015

2.2 Text Segmentation and Evaluation

We considered to segment the article into a sequence of paragraphs, then evaluated the relevance of every paragraph. As a simple text segmentation strategy, we used the <title> and </title> tags to segment each article according to the article structure in PMC and calculate the relevance score of each paragraph. It means that the text between <title> and </title> and the text between </title> and <title> are treated as paragraphs. For example, the text in the following section example can be segmented into three paragraphs which are indicated by ▸ ◂. Thus, an article *T* can be segmented into a set of paragraphs $\{p_1,p_2\ldots,p_n\}$. $R(p_i)$ represents the relevance score of the paragraph p_i and $R(T)$ represents the relevance score of article T.

```
<title> ▸2. Materials and Methods ◂ </title>
<sec>
        <title> ▸Google Earth&#x02122; ◂</title>
        <p>
        ▸Google Earth&#x02122;, was freely downloaded from <ext-link ext-link-type="uri"
xlink:href="http://earth.google.com/">
        http://earth.google.com/</ext-link>
        and created the foundation for this project. Georeferenced data is read by Google
Earth&#x02122; and similar applications using the Keyhole Markup Language (KML) format which
can be written and edited using standard text editors. Thus, georeferenced data used in this project
was compiled in KML format. ◂
        </p>
</sec>
```

At first we simply used $R(T)$ = max $R(p_i)$ as an article's relevance score. This function nevertheless leads to an extreme case: such as some subtitles and some paragraphs might be too short to be relevant to the article. Like the example section above, the subtitle "Google Earth" segmented by <title> and </title> tags is irrelevant to the article. As a simple solution, using <p> tag or limiting a paragraph to a proper length might lose some important subtitles. So we considered to recombine the article with relevant paragraphs.

For article T, we define $T' = \{p_i \mid R(p_i)>\sigma\}$ and assume $R(T) = R(T')$, where σ is a threshold value to distinguish "relevant" and "irrelevant". The above strategy recombines the article with the relevant parts in the original article. In other words, it means to eliminate the irrelevant portions in the original article.

3 Results

In order to measure the effectiveness of our strategies, we used infNDCG as the major metric and reported our results in TREC CDS 2014 and TREC CDS 2015.

During our experiments, we compared different strategies of topic enrichment and text evaluation in CDS 2014. Table 2 compares different combinations of five fields in topics with the Max Function for text evaluation. In_exp and DLH13 represent two retrieval models provided by Terrier. On the basis of these results, we used all fields in our experiments.

infNDCG	<summary>	<description>	<summary>+ <MeSHSummary>	<description> +<MeSHDescription>	All fields without <kwd>	All five fields
In_exp	0.1704	0.1612	0.1699	0.138	0.1854	0.1872
DLH13	0.1498	0.1328	0.1631	0.1349	0.1693	0.1712

Table 2 Topic Enrichment in TREC CDS 2014

In Table 3, we compared the Max Function strategy and the Text-recombination strategy in TREC CDS 2014. The baseline strategy directly retrieved articles without any special evaluation strategy. We can see that the Text-recombination strategy always outperforms the Max Function strategy on the basis of different retrieval models. PL2 is another retrieval model provided by Terrier.

infNDCG	Baseline	Max Function Strategy	Text-recombination Strategy
In_exp	0.1734	0.1872	0.1903
DLH13	0.1675	0.1712	0.1964
PL2	0.1645	0.1613	0.2160

Table 3 Two strategies in TREC CDS 2014

Table 4 shows the results of our best retrieval model (PL2) and the Text-recombination strategy in TREC CDS 2015. The medians of results of all participants are also presented and our achieved performance values are much better than the medians, especially over the metrics of infNDCG, R_prec and Prec@10. Overall, the results of our strategies are very promising.

	PL2	Median
infAP	0.0442	0.0414
infNDCG	0.2232	0.2038
R_prec	0.1762	0.1615
Prec@10	0.3800	0.3433

Table 4 PKUICST results in TREC CDS 2015

4 Conclusion

For the TREC CDS 2015 task, we explored and experimented different strategies of topic enrichment and text evaluation. The use of MeSH achieved some slight effect and the use of Text-recombination led to more positive effect.

Reference

[1] Matthew S. Simpson, Ellen M. Voorhees, William Hersh. Overview of the TREC 2014 Clinical Decision Support Track. In Proceedings of TREC 2014 on Text Retrieval Conference Publications, November 2014.

[2] J. Gobeill, A. Gaudinat, E. Pasche, P. Ruch. Full-texts representation with Medical Subject Headings, and co-citations network reranking strategies forTREC 2014 Clinical Decision Support Track. In Proceedings of TREC 2014 on Text REtrieval Conference Publications, November 2014.

[3] I. Ounis, G. Amati, V. Plachouras, B. He, C. Macdonald, et al. Terrier: A High Performance and Scalable Information Retrieval Platform. In Proceedings of ACM SIGIR'06 Workshop on Open Source Information Retrieval (OSIR 2006), August 2006.

[4] Medical Subject Headings(1999,September 1). U.S. National Library of Medicine, 8600 Rockville Pike, Bethesda, MD 20894. Retrieved December 9, 2015, from the World Wide Web: https://www.nlm.nih.gov/mesh/MBrowser.html