Literature based Clinical Decision Support: Searching Articles According to Medical Needs

Baoli Li[1], Jikang Xiao[2], Shigen Jiang[2], Zhen Wang[2], Haoqi Ding[2], Yaojun Niu[2]

[1] Department of Computer Science
Henan University of Technology
Zhengzhou, P.R. CHINA
libaoli@gmail.com

[2] Cognitive Computing Group
Lilianinfo.Com
Shanghai, P.R. China

Abstract. Physicians often seek helps from scientific medical literature for better diagnoses and treatment, especially when dealing with intractable cases. The task is becoming more and more difficult as thousands of new articles are published every month. It's quite challenging and helpful to assist physicians locating highly relevant articles in limited time. In this study, targeting at fulfilling the task of the TREC-2016 Clinical Decision Support track, we explore effective methods to search medical articles according to physicians' information needs. The best solution we obtained takes the following strategies: using Medical Text Indexer for query expansion, concentrating on articles with abstracts, tagging articles according to different information needs, and combining results of multiple models.

1 Introduction

There are strong demands for personalized and accurate healthcare for getting better life. Ordinary patients need help to choose a suitable treatment that they can afford, while physicians often ask for better decisions on diagnoses and treatment for intractable cases. Intelligent clinical decision support, thus, is expected to benefit both patients and physicians. Different from end-to-end medical assistant, literature based clinical decision support, which aims at helping physicians quickly and accurately locating those highly relevant articles, is the first and critical step to achieve the final goal for better life. The Text REtrieval Conference (TREC), therefore, has set up such an evaluation task since 2014 [1-2]. Similar to the 2014 and 2015 track, the 2016 Clinical Decision Support Track focuses on the retrieval of biomedical articles relevant for answering generic clinical questions about medical records. Unlike previous years, actual electronic health record (EHR) patient records are used instead of synthetic cases[1]. The task is quite challenging, as the EHRs usually contain a significant number of abbreviations as well as other linguistic jargon and style.

We have two goals to attend the CDS track of this year: one is to understand the in-depth problems of a preliminary clinical decision support system, and the other is to train and build our capacity to implement such a system. Therefore, the basic strategy that we follow is: adopting the successful practices proved in the past evaluations, and trying to innovate in some key parts.

We explore effective methods to find medical articles based on physicians' different information needs. Our successful practices include: using Medical Text Indexer for query expansion, concentrating on articles with abstracts, tagging articles according to different information needs, and combining results of multiple models. We have designed other measures to further improve the results, such as using learning to rank and medical concepts, but, due to the limited time, these measures were not actually employed in producing the submitted runs.

2 The TREC-2016 CDS Track Task

The TREC-2016 CDS task is basically an information retrieval one, which aims at finding for each patient described by a topic up to 1,000 articles from a medical collection containing 1.25 million articles.

There are totally 30 topics, which are evenly distributed over three different information need types: diagnosis, test, and treatment. A sample topic (no. 27) is given in figure 1.

<topic number="27" type="treatment">
<note>
96F found unresponsive on ground at nursing home. Pt was
in dining room and found by staff. Unresponsive for 1 min after
found. Pt cannot recollect events preceding fall but with some
*c/o HA and some neck/shoulder discomfort. Taken to [**Hospital1 1218**]*
where NCHCT preformed at 18:32 showed ~9mm L parietal SDH.
C-spine negative.

Family / Social history:
dementia, HTN, afib, CAD
SURGICAL Hx: unknown
.
SOCIAL Hx: Daughter serves as HCP; Pt currently DNR/DNI except for
*elective procedure (****SEE CLARIFICATIOIN BELOW****).*
.
ALLERGIES: NKDA

Physical Examination
General Appearance: No acute distress, Thin

[1] http://www.trec-cds.org/2016.html

Eyes / Conjunctiva: PERRL, Conjunctiva pale
Head, Ears, Nose, Throat: Normocephalic, Poor dentition
Lymphatic: Cervical WNL, Supraclavicular WNL
Cardiovascular: (S1: Normal), (S2: Normal)
Peripheral Vascular: (Right radial pulse: Present), (Left radial pulse: Present), (Right DP pulse: Diminished), (Left DP pulse: Diminished)
Respiratory / Chest: (Expansion: Symmetric), (Breath Sounds: Clear : bialterally)
Abdominal: Soft, Non-tender, Bowel sounds present
Extremities: Right: Absent, Left: Absent
Skin: Warm
Neurologic: Attentive, Follows simple commands, Responds to: Verbal stimuli, Oriented (to): A+O x 2, Movement: Not assessed, Tone: Not assessed

Imaging: CT head w/o contrast Acute left subdural hematoma measuring 1.5 cm maximal dimensions with leftward subfalcine herniation of 8 mm, downward transtentorial herniation with obliteration of the left suprasellar cistern, and uncal herniation. No fx, destructive infiltrative lesion involving the skull base
</note>
<description>A 96 y/o female found unresponsive on ground at nursing home. Pt was in dining room and found by staff. Unresponsive for 1 min after found. Pt cannot recollect events preceding fall but with some c/o HA and some neck/shoulder discomfort. NCHCT showed ~9mm L parietal SDH. C-spine negative. Imaging: CT head w/o contrast Acute left subdural hematoma measuring 1.5 cm maximal dimensions with leftward subfalcine herniation of 8 mm, downward transtentorial herniation with obliteration of the left suprasellar cistern, and uncal herniation. No fx, destructive infiltrative lesion involving the skull base.</description>
<summary>A 96 y/o female found unresponsive on ground at nursing home pressents with headache, herniation, and some neck/shoulder discomfort. CT head shows acute left subdural hematoma.</summary>
</topic>

Figure 1. A sample topic of TREC-2016 CDS track.

3 Key designs of our TREC-2016 CDS system

We build our TREC-2016 CDS system based on Elastic Search[2].

3.1 Text analysis

To better match topics and articles, measures are designed from two directions to process the given topics and the articles.

2 https://www.elastic.co/products/elasticsearch

For each patient / topic, we try to obtain a better representation by analyzing the medical record with MTI[3] and Metamap[4]. For each article in the collection, we planned to analyze it with the same strategy. However, it took us much time to process the whole collection of more than 1 million articles. We had to take some compromise measures finally. For example, we annotate each article with Metamap only with title and keywords. The results are not good during validation stage. So, the final submitted runs were derived without Metamap analysis. As the MTI service can only be accessed via www, it took much longer time than the locally installed Metamap service to process the whole collection. We finally decided to just use MTI to process the topics and combine the topics and MTI recognized terms to form the queries sending to the elasticsearch engine.

Besides analyzing text with MTI and Metamap, we design the following strategies to process the whole medical article collection: filtering out articles without abstracts, simplifying an article to its title, abstract, and keywords, and classifying each article into one of the three categories: diagnosis, test, and treatment. Intuitively, articles without abstract may not be scientific ones, and title, abstract, and keywords are usually good compact representation of an article. Other two reasons not using full article are: to speed up the processing time and to make the system more practical as full text is not always accessible. After removing articles without abstract, we obtain the final dataset containing about 1.1 million articles.

For classifying articles, we use a SVM classifier [3], which is trained by a dataset derived from the topic file, the relevance judgement file, and the article collection of the two past 2014 and 2015 evaluations.

3.2 Traditional Information Retrieval Measures

Besides analyzing both topics and articles in the collection, we combine the results of multiple retrieval models. Four different IR models are implemented in elasticsearch: TFIDF, Okipi BM25, DFR (Divergence from Randomness), and IB (Information based). To produce the final results, we give these four models equal weights. In addition, we also planned to use learning to rank for further improving the results. Unfortunately, we finally ran out of time, and could not rerank the final submitted result with some learning to rank algorithms.

4 Results and Discussion

We totally submitted three runs. They take the same strategies but use different parts of topics: note, description, or summary. Table 1 shows the overall results over the 30 topics of our runs. To better understand how good our runs are among all the submitted systems, we calculate the average median values of all the submitted runs using note, description, and summary respectively. Table 2 gives the results. The values are averaged over the 30 topics.

[3] https://ii.nlm.nih.gov/MTI/index.shtml
[4] https://metamap.nlm.nih.gov/

We get the best results when using summary, but the system performs much poor with note. Description helps to get results between summary and note. It is consistent with our expectation, as summaries produced by human are more compact and accurate. On the other hand, we can say that our system does not perform well in understanding the medical records. In the future, we need to concentrate on how to better understand medical text.

Table 1. The results of our three submitted runs (over 30 topics).

Run	infAP	infNDCG	R-prec	P@10
1(note)	0.0053	0.0654	0.0302	0.1300
2(description)	0.0080	0.0747	0.0338	0.1467
3(summary)	0.0257	0.1748	0.0752	0.2667

Table 2. The average median results of all submitted runs (over 30 topics).

Run	infAP	infNDCG	R-prec	P@10
1(note)	0.0099	0.1228	0.0792	0.1833
2(description)	0.0065	0.1043	0.0648	0.1533
3(summary)	0.0196	0.1859	0.1220	0.2633

From tables 1 and 2, we can see that when using note as patient representation our system achieves poorer results than the average median. The gaps become narrower when using description. When using summary, our system performs better than the average median on two metrics: infAP and P@10.

5 Conclusions and Future Work

In this report, we give the design of our system for the TREC-2016 Clinical Decision Support track. Our successful practices include: using Medical Text Indexer for query expansion, concentrating on articles with abstracts, tagging articles according to different information needs, and combining results of multiple models.

This is the first time that we participate in this evaluation, and we plan to re-design the system based on the experience and lessons gained in this activity. In the future, we would like to focus on how to accurately understand the medical records with existing resources and more natural language processing techniques.

Acknowledgments. This work was partially supported by the Henan Provincial Research Program on Fundamental and Cutting-Edge Technologies (No. 112300410007), and the High-level Talent Foundation of Henan University of Technology (No. 2012BS027).

References

1. M. S. Simpson, E. M. Voorhees, and W. Hersh. Overview of the TREC 2014 Clinical Decision Support Track. In Proceedings of the TREC 2014.
2. K. Roberts, M. S. Simpson, E. M. Voorhees, and W. Hersh. Overview of the TREC 2015 Clinical Decision Support Track. In Proceedings of the TREC 2015.
3. C.-C. Chang and C.-J. Lin. LIBSVM: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology, 2:27:1--27:27, 2011.