Information Retrieval to Address Patients' Questions when Reading Clinical Reports

Introduction

The Task 3 of 2013 ShARe/CLEF eHealth Evaluation Lab aimed to provide valuable and relevant documents to patients, so as to satisfy their health-related information needs.[1] The use case of Task 3 was described as follow: when a patient is reading his discharge summary from a recent hospital stay, questions may come up that would prompt a web search to learn more about his health conditions. The hypothesis of Task 3 is that by utilizing collateral information from discharge summaries, an information retrieval (IR) system in healthcare domain would have better performance. In this review I provided an overview of the evaluation lab as it represents a classic IR question, and described the strategy used by the winner team.[2]

Overview of 2013 ShARe/CLEF eHealth Evaluation Lab Task 3

The document data set is comprised of a collection of 1 million documents of web pages crawled from medical web sites targeting both public and healthcare professionals. 65 medical disorders are chosen and a discharge summary from each disorder is generated. Queries were generated by healthcare experts, mimicking what laypeople would ask after reading each discharge summary to find out more about the diseases (e.g., “pneumonia steroids duration”). For each query, an IR system that implements a standard Okapi BM25 scheme is used to generate a benchmark model, which did not incorporate any domain-specific adaption from discharge summaries. There were 5 training queries and 50 testing queries. Relevance assessment was performed by healthcare experts using the Relevation system.[3]

For this task, each team were allowed to submit up to seven runs with one mandatory run without the use of discharge summary. Total 13 teams participated in this task. Results evaluated including P@10 and NDCG@10, with the expectation that the winner system should beat the performance of baseline BM25 retrieval model (P@10 0.4860, NDCG@10 0.4328). BM25 proved to be a strong baseline that is hard to beat, and only 4 teams delivered best run above the baseline score. The winner team is from Mayo Clinic, USA with P@10 of 0.5180 and NDCG@10 of 0.4665 in their best run.

Summary of the IR system used by teamMayo

TeamMayo used a two-step ranking strategy. In step one, text retrieval occurred in text space by applying query likelihood language model, Markov random field (MRF) model, relevance model (MRM), and Medical Subject Headings (MeSH) based query expansion. In step two, output from Step 1 were re-ranked in the concept space where everything was represented by the medical concepts. The key philosophy from teamMayo is to use an NLP approach for concept identification that incorporate medical domain knowledge from discharge summary to improve pure text-based retrieval.

Text-based retrieval

Indri[4] was used to build the index for the target collection in the text space, after html cleaning and stopping words removal. First, a standard query likelihood language model with Dirichlet smoothing was used as baseline. Next, several advanced models were tried in different combinations. MRF model has been shown to be superior than basic query likelihood model in medical records collections by introducing term proximity information.[5] MRM model was used to apply “external expansion” of query term by incorporating external corpora based on language modeling.[6] The purpose of MRM is to reduce vocabulary gap between query language and document language. Of note, teamMayo took advantage of utilizing a large medical record database from Mayo Clinic between 2001– 2010 to construct MRM model. Lastly, additional query expansion was performed with MeSH, which were designed to categorize scientific and medical literature. As an example, after above steps an initial query of "shortness breath swelling" may be expanded to include relevant medical terms (e.g., "dyspnea paroxysmal", “pulmonary edema”) at different weights.

Concept-based Retrieval for Re-ranking

It was hypothesized that discharge summary would contain ‘hidden’ concepts that did not appear in the query but were related to query concepts. To identify such ‘hidden’ information, medical concepts were created from documents and discharge summaries by using a clinical NLP annotation tool called MedTagger. MedTagger extracts and annotates concepts from clinical text by leveraging large corpus statistics, machine learning, knowledge bases, and syntactic parsing.[7] MedTagger identifies Concept Unique Identifiers (CUL) by calculating query likelihood-based relevance score, and then weights each CUI by its term frequency. Finally, teamMayo re-ranked the output from text-based retrieval by weighting the results from concept-based retrieval.

Results from teamMayo

The best results coming from run 2 (P@10 of 0.5180 and NDCG@10 of 0.4665). In this run, initial ranking in text-space was obtained by a linearly combined MRF and MRM model. Then the top 1000 retrieved documents were re-ranked in the concept-space by incorporating CUL information from discharge summaries. Of interest, the results from run 5 is comparable to run 2 (P@10 of 0.5040 and NDCG@10 of 0.4618). Run 5 skipped the re-ranking process (text-based retrieval only with MRF, MRM, and MeSH-based query expansion), suggesting a pure text-based retrieval model alone is quite strong.

Conclusion

It appears simple BM25 model produces a very strong baseline IR system in the use case of Task 3, which has proved hard to beat. The best results come from a two-step ranking strategy incorporating multiple state-of-the-art models (text-space ranking) and leveraged information in patients’ discharge summaries (concept-space ranking). However, a pure text-based retrieval model by teamMayo also achieved comparable performance. Task 3 highlighted challenges in using domain knowledge to improve IR for healthcare needs.

Reference

1. Goeuriot L, Jones GJF, Kelly L, et al. ShARe/CLEF eHealth Evaluation Lab 2013, Task 3: Information retrieval to address patients' questions when reading clinical reports. CLEF 2013 online working notes 2013;8138.
2. Zhu D, Wu ST, Masanz JJ, Carterette B, Liu H. Using Discharge Summaries to Improve Information Retrieval in Clinical Domain.
3. Koopman B, Zuccon G. Relevation! An open source system for information retrieval relevance assessment.1243-1244.
4. Dang V. The lemur project-wiki-ranklib. Lemur Project 2013.
5. Metzler D, Croft WB. A markov random field model for term dependencies.472-479.
6. Diaz F, Metzler D. Improving the estimation of relevance models using large external corpora.154-161.
7. Liu H, Bielinski SJ, Sohn S, et al. An information extraction framework for cohort identification using electronic health records. AMIA Summits on Translational Science Proceedings 2013;2013:149.