Clinical Decision Support Track

Alexander Cook Nithin Sivakumar Peihao Yuan

October 26, 2017

1 Task

1.1 Definition

The core task is to find the most relevant literatures from the massive biomedical literatures according to the patient's condition and the needs of doctors. The TREC Clinical Decision Support Track 2016 (CDS2016) requires selected articles from the PubMed Central (PMC) Biomedical Literature to contain the most relevant articles for the corresponding symptom and clinical purpose for 30 Topics with summary, description and note.

1.2 Inputs

The inputs are the query types. There are three query types or the clinical purposes of the search, namely Diagnosis, Test and Treatment. We will be using admission notes from MIMIC-III. An admission note describes a patient's chief complaint, relevant medical history, and any other information obtained during the first few hours of a patient's hospital stay, such as lab work.

1.3 Outputs

Outputs are PMCID for the patient admission details. The run files are of the standard trec-eval format.

TOPIC-NO Q0 PMCID RANK SCORE RUN-NAME

2 Motivation

2.1 Domain

Java will be our main domain to implement the project.

2.2 Importance

By solving this task, medical data of the patients can be tracked and reported.

2.3 Usage

This will be really helpful for physician to find the right treatment, diagnosis or test for the patient based on the patients health record.

3 Method

3.1 How

Our work involves three aspects: the expansion of the query, medical literature preprocessing and scoring model selection.

3.2 Algorithms

We have planned to use all the algorithms we have learnt so far and come up with best one.

3.2.1 Pseudo relevance feedback

Pseudo relevance feedback is the most commonly used query expansion. The principle is that after a regular search, assuming that the results of the first few documents are more accurate. We select a number of representative vocabularies from the first few selected documents as an extension to add to the original query. Then a second round retrieve can be performed to obtain more relevant literatures. There are two parameters that we need to consider in the search process:

- TopDocuments
- TopTerms

4 Evaluation

4.1 Data

We are going to use data.qrels from the CDS2015 set for evaluation.

4.2 Ground Truth

Correctness is - If the query returns the only true positive results.

4.3 Evaluation measure

We are planning to Precision@R, NDCG, MAP.

5 Expection

Our expectation is that optimization of query expansion and weighting model of retrieve should provide better results compare to other models.