IIIT-H at CLEF eHealth 2017 Task 2: Technologically Assisted Reviews in Empirical Medicine

Jaspreet Singh[1,2] and Lini Thomas[1,2]

[1] DSAC, IIIT Hyderabad,
Hyderabad, India
{jaspreet.singh,lini.thomas}@research.iiit.ac.in
[2] KCIS, IIIT Hyderabad,
Hyderabad, India

Abstract. Observational evidence in clinical practice is critical in healthcare and policy making. Researchers spend a lot of time searching for relevant published articles to write a systematic review of a topic. In this paper, we present our participation as the team of IIIT Hyderabad at Task2 Technologically Assisted Reviews in Empirical Medicine as an effort to automate this task and deliver relevant information in medical literature. We base our approach on query expansion according to relevance feedback. Query expansion is a standard technique in information retrieval tasks with growing use in medical literature [1, 2]. Articles returned from pubmed query performed during a systematic review are first indexed using lucene's inverted index. The query is porcessed for term boosting, fuzzy search and used for scoring documents according to TF-IDF similarity. Relevance feedback is used to update the query and become more pragmatic.

Keywords: medical information retrieval, relevance feedback, query expansion

1 Introduction

Diagnostic tests are critical to healthcare. Well designed reviews of results from Diagnostic test accuracy(DTA) studies will help in decision making in medical domain [3]. But there are enormous amount of articles published every year. Information retrieval in medicine has caught attention due to significant implications of evidence-based medicine and rapidly expanding medical libraries. Automatic screening of medical literature will help evolve retrieval techniques applicable in other domains as well. CLEF eHealth Task2 [4, 5] is an effort towards this purpose.

We participate in Task 2: Technologically Assisted Reviews in Empirical Medicine, evaluating information retrieval of medical documents. The task focuses on ranking and thresholding methods for DTA reviews. We proposed a system which is based on query expansion using fuzzy logic and relevance feedback to get relevant documents. Relevance feedback is used earlier in various information retrieval systems[6–8]. Fuzzy search make query flexible and helps improve recall. Relevance feedback helps reconstruct the query to deal with any ambiguous information need [9]. Thus, we use both techniques in our system.

Each query is initially converted into a fuzzy query. The documents pertaining to each topic are indexed using lucene[3]. These indexed documents are searched using query provided by Cochrane experts. The query is updated to include more terms from relevant documents from an initial set of ranked documents provided by lucene and remove terms from irrelevant ones. Since the initial ranking of a few documents gives high average precision, the idea is to let unique terms be picked from them to better represent a query. The updated query is used to rank remaining documents.

2 Methodology

In this section we explain our methodology in detail. For simple evaluation runs, we try to optimize recall by ranking approximately half of the documents. However, for cost effective measures, we stop when we don't find any query updates or average precision in the last set of ranked results falls below a threshold (0.1 in most cases). A summary of the runs submitted to the task is shown in Table 1 .

2.1 Indexing

We let lucene index each topic's documents. Lucene breaks each document into words to create an inverted index. This index consists of terms with set of documents that contain it. Later, it is utilized for efficient search. To reduce noise and false positives, we remove stop words from the documents at the time of indexing. Lucene separates document information into fields. We create fields for title, abstract etc. from pubmed documents, as the queries specifies terms along with fields to search them from.

2.2 Query Reformulation

The query provided by the Cochrane experts vary in length and have complex boolean logic. We use a fuzzy search system to expand it. The system allows terms close to the base term to be included in the expanded query. For example, search terms like "dysplasia" also include terms like "dysplastic" and "dysplasias". Although the OVID medline search syntax includes some amount of regex present in the query, we make every term go through fuzzy search system before adding it to expanded query.

[3] `https://lucene.apache.org/core/`

Run	Evaluation	Relevance feedback
1	simple	query + relevant terms
2	simple	query - irrelevant terms
3	simple	query + relevant terms - irrelevant terms
4	simple	query + relevant terms - irrelevant terms + term boosting
5	cost-based	query + relevant terms
6	cost-based	query - irrelevant terms
7	cost-based	query + relevant terms - irrelevant terms
8	cost-based	query + relevant terms - irrelevant terms + term boosting

Table 1. Summary of the runs submitted for evaluation.

2.3 Relevance Feedback and Term Boosting

After building the document index and query reformulation, we make use of TF-IDF scoring model. Vector space models lets reweigh search terms quickly and uses cosine for calculating similarity between document and query. Four similarity measures are incorporated - tf, idf, coord and length Norm. Where coord is number of terms in the query that were found in the document and length Norm is measure of the importance of a term according to the total number of terms in the field.

Initially, we request a small set of ranked and scored documents from lucene. This initial set is inspected for relevance. We found from our experiments on the training data that about half of this set is relevant. Let (r_d) be the set of relevant document and (nr_d) be the set of not relevant documents in the initial ranking. The search query is appended with boolean OR with top occuring terms from r_d and boolean NOT from top occuring terms from nr_d given that they don't already occur in the query. To prevent overpopulating terms in the query and drifting away from desired result, we restrict the count of new terms five percent of average article size. Once updated, the new query is used to rank remaining documents.

We boost a term for scoring if it occurs in r_d for multiple iterations. These terms get n times as much weight of any other term if they occur again in the n^{th} iteration. Incorporating this, we found that though we are providing a binary relevance feedback, our system has the advantages of a graded feedback. Relevance feedback system is applied to queries containing more than 1500 documents. Apart from the submitted runs, we found that this technique was effective on queries having less documents.

2.4 Results

We submitted eight runs for this task. Four of which are for simple evaluation and four for cost-based evaluation.

References

[1] Zhenyu Liu and Wesley W. Chu. Knowledge-based query expansion to support scenario-specific retrieval of medical free text. *Information Retrieval*, 10(2):173–202, 2007.

[2] M.C. Díaz-Galiano, M.T Martín-Valdivia, and L.A. Ureña-López. Query expansion with a medical ontology to improve a multimodal information retrieval system. *Computers in Biology and Medicine*, 39(4):396 – 403, 2009.

[3] Gah Juan Ho, Su May Liew, Chirk Jenn Ng, Ranita Hisham Shunmugam, and Paul Glasziou. Development of a search strategy for an evidence based retrieval service. *PLOS ONE*, 11(12):1–14, 12 2016.

[4] Hanna Suominen, Liadh Kelly, Lorraine Goeuriot, Evangelos Kanoulas, Rene Spijker, Aurélie Névéol, Guido Zuccon, and João R. M. Palotti. Overview of the CLEF ehealth evaluation lab 2017. In *Experimental IR Meets Multilinguality, Multimodality, and Interaction - 8th International Conference of the CLEF Association, CLEF 2017, Dublin, Ireland, September 11-14, 2017, Proceedings*, Lecture Notes in Computer Science. Springer, 2017.

[5] Evangelos Kanoulas, Dan Li, Leif Azzopardi, and Rene Spijker. Overview of the CLEF technologically assisted reviews in empirical medicine. In *Working Notes of CLEF 2017 - Conference and Labs of the Evaluation forum, Dublin, Ireland, September 11-14, 2017.*, CEUR Workshop Proceedings. CEUR-WS.org, 2017.

[6] Pragati Bhatnagar and Narendra Pareek. Improving pseudo relevance feedback based query expansion using genetic fuzzy approach and semantic similarity notion. *Journal of Information Science*, 40(4):523–537, 2014.

[7] Jagendra Singh and Aditi Sharan. Relevance feedback based query expansion model using borda count and semantic similarity approach. *Intell. Neuroscience*, 2015:96:96–96:96, jan 2015.

[8] Paul Alexandru Chirita, Claudiu S. Firan, and Wolfgang Nejdl. Personalized query expansion for the web. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '07, pages 7–14, New York, NY, USA, 2007. ACM.

[9] Ruthven, Ian, Lalmas, and Mounia. A survey on the use of relevance feedback for information access systems. *Knowl. Eng. Rev.*, 18(2):95–145, jun 2003.