Investigation of Concept-based Proximity Matching - GRIUM@Clinical Decision Support Track 2015 Task 1a

Xiao Jie Liu, Jian-Yun Nie

RALI, DIRO, University of Montreal
C.P.6128, succursale Centre-ville, Montreal, Quebec Canada H3C 3J7

{xiaojiex, nie}@iro.umontreal.ca

Abstract.
Concepts are often used in Medical Information Retrieval. Previous studies showed that exact concept matching (using either the exact concept expression or concept IDs) is not effective, while using concept expressions for proximity matching is more effective. In our participation in Clinical Decision Support Track 2015 task 1a, we investigated the utilization of proximity matching based on concepts. Our results suggest that this matching strategy can be helpful. In this report, we describe the methods tested as well as their results.

Keywords: Information Retrieval, MetaMap, Proximity Matching, Concept

1 Introduction

The Clinical Decision Support Track 2015 task 1a investigates techniques for linking medical cases to information relevant for patient care. In making clinical decisions, physicians often look for information about the best practice for their patients. Information relevant to a physician can be related to a variety of clinical tasks such as determining a patient's most likely diagnosis given a list of symptoms, deciding on the most effective treatment plan for a patient having a known condition, and determining if a particular test is indicated for a given situation.

In addition to traditional bag-of-words approaches, IR in medical area benefits from the rich resources available, such as the UMLS Metathesaurus [1], as well as the concept mapping tools (e.g. MetaMap [2]) specifically designed for medical literature. It may seem that using the concepts identified and using such resources and tools could help medical IR. However, several previous studies ([3][4][5]) showed that a naive utilization of the mapping results is not effective. On the other hand, when the identified concepts are used as phrases, which are matched at proximity, one can obtain better results. In our participation in Clinical Decision Support Track 2015 task 1a, we will use the same strategy and we aim to test its effectiveness in this task, which is slightly different from the CLEF eHealth task 3a.

This report is organized as follows. In section 2, we will describe the proximity matching strategy. In section 3, we describe the retrieval methods used in our participation. In section 4, we report the experimental results. Preliminary conclusions will be drawn in section 5.

2 Proximity Matching Strategy

Given a phrase, usually formed of several words (e.g. heart attack), proximity matching aims to allow the phrase to match different variants of the expression. If the words of the phrase occur within a text window, we consider that there is a proximity matching. This proximity matching is combined with two additional matchings: exact matching which matches the exact expression of a concept, and bag-of-words matching which uses the set of words involved in all the expressions of the concept.

Let us consider the example of “heart attack”. This expression is identified by MetaMap as denoting the concept C0027051. This concept can be expressed by the following expressions :{“heart attack”,”myocardial infarction”,”coronary attack”}.

The traditional bag-of-words matching score is determined using the words (or stems) “heart” and “attack”. Therefore, the exact matching component tries to determine documents containing any of these expressions. In Indri, this corresponds to #1(heart attack), #1(myocardial infarction) and #1(coronary attack).

The proximity matching component identifies documents in which one of the expressions appears within a text window of the size |expression|+1, i.e. one word larger than the size of the expression. This strategy turned out to work well in our previous study [3]. In Indri, this corresponds to #uw3(heart attack), #uw3(myocardial infarction) and #uw3(coronary attack).

Finally, the bag-of-words concept matching component uses all the words from the above concept expressions. In Indri, this corresponds to #combine(heart attack myocardial infarction coronary attack).

The final score of a document is determined by the following linear combination:

S=λ_1(#1(heart attack) #1(myocardial infarction) #1(coronary attack))+λ_2 (#uw3(heart attack) #uw3(myocardial infarction) #uw3(coronary attack))+λ_3 (#combine(heart attack myocardial infarction coronary attack)). λ_1, λ_2, λ_3 are the weighting parameters.

Below is the table of matching strategy and expression in Indri.

Matching Strategy	**Expression in Indri**
Exact-Phrase matching	#1()
Prox-Phrase matching	#uwN()
BOW-Concepts matching	#combine()

Table 1. Matching Strategy and Expression in Indri

3 Retrieval Methods Tested

In this section, we describe the methods we tested in our participation.

3.1 Baselines

As a baseline, we use a traditional approach based on language modeling, with Dirichlet smoothing. In this method the score of a document D given a query Q is determined as follows:

$$S(Q, D) = \frac{1}{n}\sum_{i=1}^{n} \log P(q_i|D) \tag{1}$$

where Q is the query, D is the document, n is the length of query and $P(q_i|D)$ is the probability of document language model to create query term q_i, which is adjusted by Dirichlet smoothing below:

$$P(q_i|D) = \frac{tf_{q_i,D} + \mu\frac{tf_{q_i,C}}{|C|}}{|D| + \mu} \tag{2}$$

where tf is the term frequency of the query term q_i in document D, C is the whole collection, and μ is the smoothing parameter which is set to 2000.

Another baseline is BM25, in which the relevance score of a document d is computed as follows:

$$RSV_d = \sum_{t\in q} \log\left[\frac{N}{\mathrm{df}_t}\right] \cdot \frac{(k_1 + 1)\mathrm{tf}_{td}}{k_1((1-b) + b \times (L_d/L_{\mathrm{ave}})) + \mathrm{tf}_{td}} \tag{3}$$

where N is the total number of documents in the collection, df_t is the document frequency of term t, tf_{td} is the term frequency of t in the document d, k_1 and b are constants, L_d and L_{ave} are the document length and the average document length in the collection.

3.2 Query Expansion

In addition to the baselines, we tested the combined method described in Section 2. This combined method led to the best experimental result in CLEF eHealth task 3a.

$$S(Q|D) = \lambda_1 S_{BOW} + \lambda_2 S_{Exact\text{-}Phrase} + \lambda_3 S_{Prox\text{-}Phrase} + \lambda_4 S_{BOW\text{-}Concepts} \tag{4}$$

where S_{BOW} is the score from BOW method (with language model) with the original query, $S_{Exact\text{-}Phrase}$ is the score from exact concept phrase matching, $S_{Prox\text{-}Phrase}$ is the score from proximity concept phrase matching, $S_{BOW\text{-}concepts}$ is the score from the bag-of-words of concept expressions matching. λ_1, λ_2, λ_3 and λ_4 are the parameters and $\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 = 1$.

4 Experiments

The target document collection is the Open Access Subset of PubMed Central (PMC). PMC is an online digital database of freely available full-text biomedical literature. Each document contains JATS tags (http://jats.nlm.nih.gov/publishing/). We extract content inside <BODY> tag and delete other tags to be the document. Finally dataset contains a total of 733,138 documents.

There are in total 30 queries in this year's task. Below is an example.

Clinical Decision Support Track 2015 Query Example:

<topic number="2" type="diagnosis">
<description>
A 62 yo male presents with four days of non-productive cough and one day of fever. He is on immunosuppressive medications, including prednisone. He is admitted to the hospital, and his work-up includes bronchoscopy with bronchoalveolar lavage (BAL). BAL fluid examination reveals owl's eye inclusion bodies in the nuclei of infection cells.
</description>
<summary>
A 62-year-old immunosuppressed male with fever, cough and intranuclear inclusion bodies in bronchoalveolar lavage
</summary>
</topic>

We use Indri/Lemur as the basic experimental platform for all the methods. We use <summary> to be the query. Usual pre-processings are used: Porter stemming and removal of stopwords (Lemur stoplist). All the methods are listed in Table2 below. Because in previous experiments, language model is worse than BM25, we just list it here (GRIUM_EN_Run0) to compare, but not submit it. We submit GRIUM_EN_Run1 and GRIUM_EN_Run2 at last.

Run	Experiment Method	Submission
1	Baseline (language model with Dirichlet smoothing, μ=2000)	GRIUM_EN_Run0
2	Baseline(BM25)	GRIUM_EN_Run1
3	Using UMLS concept-based proximity matching λ_1=0.8, λ_2=0.15, λ_3=0.05, λ_4=0	GRIUM_EN_Run2

Table 2. Experiments setup for our methods

For evaluation, 4 measurements are infAP, infNDCG, R-prec (precision at R where R is the number of known relevant documents), and P@10. Note that the P@10 values are exact since more than the top 10 documents retrieved were judged for each run. The table below summarizes the results of our methods.

Submission	infAP	infNDCG	R-prec	P@10
GRIUM_EN_Run0	0.0398	0.1950	0.1536	0.3467
GRIUM_EN_Run1	0.0417	**0.2119**	0.1581	0.3533
GRIUM_EN_Run2	**0.0434**	0.2107	**0.1621**	**0.3567**

Table 3. Results of our 3 methods

In GRIUM_EN_Run1 and GRIUM_EN_Run2, the performance for most of queries among the 30 queries is better than the median. So the final average performance is between median and the best.

Among all the 3 methods, GRIUM_EN_Run2 produced the highest scores on three measurements: infAP, R-prec and P@10. GRIUM_EN_Run1 produced highest scores on measurement infNDCG. For this task the language model (GRIUM_EN_Run0) is not better than BM25 (GRIUM_EN_Run1). However, once query expansion and proximity matching strategy are used, the performance of language model (GRIUM_EN_Run2) becomes better than BM25 (GRIUM_EN_Run1).

For GRIUM_EN_Run2, we did not tune parameters λ_1, λ_2, λ_3, $\lambda_{4.}$ They have been set according to previous experience. It is possible that with different settings, we could obtain better results. This is what we will test in the future.

5 Conclusion

In this year Clinical Decision Support Track, we focused on the method with proximity concept matching strategy. We use MetaMap to extract concepts from queries and all the expressions of the concepts are used as phrases. We submit 2 runs – a baseline run (BM25) and a run using concept proximity matching. Our runs are generally better than the median. In particular, when using proximity matching strategy, we observe slight improvements over the baseline (BM25). This may indicate that such a method could be useful to professional dataset.

Reference:

1. Bodenreider Olivier. The unified medical language system (UMLS): integrating biomedical terminology. Nucleic acids research, vol. 32, no suppl 1, p. D267-D270. 2004.
2. Bedrick Steven, G. Sheikhshabbafghi. "Lucene, MetaMap, and language modeling: OHSU at CLEF eHealth 2013." Proceedings of the ShARe/CLEF eHealth Evaluation Lab. 2013.
3. Wei Shen, Jian-Yun Nie, Xiao Hua Liu, Xiao Jie Liu. An Investigation of the Effectiveness of Concept-based approach in Medical Information Retrieval GRIUM@CLEF2014HealthTask3. 2014.
4. Aronson A. R., Rindflesch T. C. Query expansion using the UMLS Metathesaurus. In Proceedings of the AMIA Annual Fall Symposium. American Medical Informatics Association, p.485. 1997.
5. Koopman Bevan, Zuccon Guido.2012 Exploiting SNOMED CT Concepts & Relationships for Clinical Information Retrieval. TREC 2012 Track. 2012.
6. Jimmy Lin, Demner-Fushman Dina. The Role of Knowledge in Conceptual Retrieval: A Study in the Domain of Clinical Medicine. SIGIR'06. 2006.
7. M Zhong, X.J. Huang. Concept-based biomedical text retrieval. SIGIR'06. 723-4. 2006.
8. Srinivasan P. Query expansion and MEDLINE. Information Processing and Management,32(4): 431-443. 1996.
9. Jun Miao, Jimmy Xiangji Huang, Zheng Ye. Proximity-based Rocchio's Model for Pseudo Relevance Feedback. SIGIR'12. 2012.