

ECNU at 2015 eHealth Task 2: User-centred Health Information Retrieval

Yang Song^{1,2}, Yun He^{1,2}, Qinmin Hu^{1,2,3}, Liang He^{1,2}, and E. Mark Haacke³

¹ Shanghai Key Laboratory of Multidimensional Information Processing

² Department of Computer Science & Technology, East China Normal University, Shanghai, 200241, China

{ysong,yhe}@ica.stc.sh.cn, {qmhu,lhe}@cs.ecnu.edu.cn

³ MR Research Facility, Department of Radiology, Wayne State University, Detroit, MI 48201, USA
nmrimaging@aol.com

Abstract. This paper presents our work on the 2015 CLEF eHealth Task 2. In particular, we propose a Web-based query expansion model and a learning-to-rank algorithm to better understand and satisfy the task.

Keywords: Web-based Query Expansion, learning-to-rank

1 Introduction

The goal of the ShARe/CLEF(Cross-Language Evaluation Forum) eHealth Evaluation Lab is to evaluate systems that support people in searching for and understanding their health information. The 2013 and 2014 eHealth tasks focus on the investigation of the effect of using additional information such as a related discharge summary and external resources such as medical ontologies on the effectiveness of information retrieval systems[1][2].

The 2015 CLEF eHealth Task 2 [3][4] is a continuation of the previous CLEF eHealth Task 3 that ran in 2013 and 2014. In this year’s task, we are asked to research mimic queries that are confronted with a sign, symptom or condition and attempt to find out more about the health conditions the users have. For example, when confronted with signs of jaundice, non-experts may use queries like “white part of eye turned green” to search for information that allows them to diagnose themselves or better understand their health conditions. These queries are often circumlocutory in nature, where the long and ambiguous wording is used in place of the actual name to refer to a condition or disease.

Our experiments on Task 2 aim to investigate in effectiveness of our Web-based query expansion model and the customized learning-to-rank algorithm for medical information retrieval. Figure 1 presents our framework of integration of the Web-based query expansion model and the learning-to-rank algorithm. Particularly, we take advantage of the Web search engine to obtain better expansion terms. At the same time, we learn the features from the previous 2013 and 2014

data as the training data and then test them on the 2015 task through the customized learning-to-rank algorithm. Furthermore, we adopt multiple classic information retrieval (IR) models, such as BM25, language model (LM) to get the expansion candidates, in order to get rid of the influence of single model. Finally, our submit ten runs for evaluation.

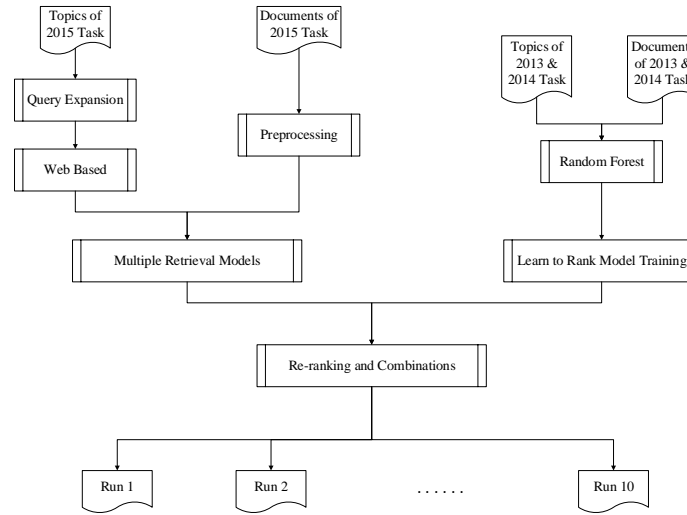


Fig. 1. Framework

2 Methods

2.1 Web-based Query Expansion

The task of this year is user-centered health information retrieving. Hence, our work concentrates on exploring circumlocutory queries that users may pose when they are faced with signs and symptoms of a medical condition. In other words, the queries in 2015 do not contain the medical terms. This results in that the traditional IR models are difficult to match the queries by the medical documents.

We are motivated to solve this problem by adopting the Web search engines to translate the general description into the medical terminologies. After that, we adopt the related medical technical terms as the query expansion terms into the query. Then, the new query is matched by the IR models again and gains the ranking documents as the output for evaluation.

Here is a common scenario. A user is searching the signs and symptoms of a disease in Google search engine. Google returns the top documents which mostly contain the possible diseases. Then, the user adopts the diseases and re-searches the names directly in Google. Generally, her/his question can be partially answered. In our observations, the names of the possible diseases given by Google are mostly MeSH terms.

Therefore, a Web-based query expansion model is proposed as follows. Note that we applied the similar model in the 2014 TREC Microblog track [5] which achieved better results than most of the runs.

- Query is searched by Google and the top 10 concurrent Web titles and snippets (if existed) are crawled from the Web page.
- By applying the MeSH database, the medical terms are extracted from both the titles and the snippets.
- The frequency of each stemmed medical term is calculated. Only the terms appearing more than n times are kept for expanding, which can be denoted as Q_{web} .
- The final query is formulated as $Q = Q_0 \cup Q_{web}$, where Q_0 represents the initial query.

In addition, since the queries are to find out “what is the patient’s diagnosis?”, “what tests should the patient receive?” and “how should the patient be treated?”, we manually add the keywords ‘diagnose’, ‘test’ and ‘treatment’ as the regular expansion terms to all queries.

2.2 The Customized Learning-to-rank Algorithm

The results of Wei Shen’s approach [6] have shown that the bag-of-concepts is less effective than bag-of-words approach. Then, we treat the concepts as the features that we need to learn in the queries. Based on the MeSH dictionary, we customize a learning-to-rank algorithm to learn the features from the previous 2013 and 2014 tasks. Then, we test the features in the 2015 task.

Feature Extraction: We extract the weighting score and the rank of each document-query pair from a retrieval model as the features[7]. The weighting score is the result of the first retrieval round, which represents the relevance assessed by the retrieval model. The ranking is given by the weighting scores as a non-weighting based feature. To utilize the advantage of different retrieval models, we obtain the scores and the corresponding ranks from the BM25, PL2 and BB2 model. Hence, in our learning-to-rank platform, the dimension of the feature vector is six.

Random forest and Model Training: Random forest is composed of multiple decision trees which have no relationship with each other. Each decision tree will classify the samples in the testing data set, the final result of classification depends on the vote of all the trees[8]. We apply random forest to classify the

document-query pairs into two groups as relevant and irrelevant. Here we apply 20 decision trees in our forest. The training data is transformed from the results in the 2013 and 2014 task. The weighting scores and ranks of the BM25, PL2 and BB2 model are extracted to represent the document-query pairs in the previous results.

Model Application: Firstly, we apply the BM25, PL2 and BB2 model to get three initial results. Then a new result is obtained by the combination of scaled scores of three retrieval model. Secondly, random forest model is utilized to classify this new result. The document-query pairs which are classified as relevant will award extra relevance scores. Finally, the results are re-ranked by their new scores.

2.3 Combination

We apply equation 1 to normalize the scores of each candidate. Then, we add up all the normalized score of each document among these candidates, followed by the documents which are ranked by the total normalized score. Finally, the top 1000 documents for each query are extracted as final results for evaluation.

$$score_normalized^i = \frac{score_{max} - score^i}{score_{max} - score_{min}} \quad (1)$$

3 Experiments and Evaluation

The whole corpus of the 2015 eHealth Task 2 is the same as the 2014 one. Since the Web pages with the HTML format contain many tags where most of them are useless in retrieval, we eliminate all css, JavaScript part and HTML tags at the document preprocessing stage. Only texts which are visible in the web pages are indexed. Furthermore, we observe that the pages which link to each other are highly similar to each other. Therefore, we make an assumption that the linked pages are not relevant, if a page is not retrieved as relevant to the query. This results in that the URLs in the document are removed.

We adopt Terrier to conduct our experiments on the given data sets. Here we submit ten runs where the description for each run is as follows.

- ECNU_EN_Run.1: the baseline with the TF-IDF model.
- ECNU_EN_Run.2: we utilize the learning-to-rank model to re-rank the results, based on the random forest training method, where we set the trees with the number of 100.
- ECNU_EN_Run.3: we utilize the Google search engine to translate the queries into the medical terms, where there terms of 'diagnose', 'test' and 'treatment' are added by mandatory into queries.
- ECNU_EN_Run.4: we utilize the same method as ECNU_EN_Run.2, while selecting 20 trees to do classification.
- ECNU_EN_Run.5: we utilize the learn-to-rank model to re-rank the results, based on the random forest training method, where the queries are expanded by the Google search engines.

- ECNU_EN_Run.6: we utilize the Google search engine to obtain the expanded medical terms with the BM25 model.
- ECNU_EN_Run.7: we utilize the Google search engine to obtain the expanded medical terms with the traditional pseudo-relevant feedback method.
- ECNU_EN_Run.8: we combine the above seven runs with manually parameter settings.
- ECNU_EN_Run.9: we use BM25 with the pseudo relevant feedback.
- ECNU_EN_Run.10: we combine the runs which are retrieved by BM25, PL2 and TF-IDF.

The primary evaluation measure of this year is the precision at the first 10, and the secondary measure is Normalized Discounted Cumulative Gain at top 10(NDCG@10). Evaluation of our submissions is summarized in Table 1.

Table 1. Evaluation of our submissions

Run	P@10	NDCG@10	MAP
ECNU EN Run.1	0.3470	0.3144	0.2056
ECNU EN Run.2	0.3606	0.3220	0.1682
ECNU EN Run.3	0.5394	0.5086	0.3052
ECNU EN Run.4	0.3168	0.2065	0.3652
ECNU EN Run.5	0.3152	0.3006	0.1515
ECNU EN Run.6	0.4227	0.3978	0.2443
ECNU EN Run.7	0.3227	0.3004	0.1887
ECNU EN Run.8	0.4530	0.4226	0.2738
ECNU EN Run.9	0.3606	0.3203	0.2280
ECNU EN Run.10	0.4667	0.4525	0.2754

4 Conclusions and Future Work

This year we mainly focus on the 2015 CLEF eHealth task 2. We propose a Web-based query expansion model and a customized learning-to-rank algorithm to achieve the better performance for medical information retrieval. Our best submission obtains 0.3052 in terms of MAP, 0.5394 in terms of p@10 and 0.5086 in terms of NDCG@10. In the future, we will continue on the Web-based query expansion method for better understand the queries.

5 Acknowledgement

This research is sponsored by the postdoc fellowship (PDF) from the Natural Science & Engineering Research Council (NSERC) of Canada. We also thank anonymous reviewers for their review comments on this paper.

References

1. Goeriot, L., Jones, G.J.F., Kelly, L., Leveling, J., Hanbury, A., Müller, H., Salantera, S., Suominen, H., Zuccon, G.: ShARe/CLEF eHealth Evaluation Lab 2013, Task 3: Information retrieval to address patients' questions when reading clinical reports. In: CLEF 2013 Online Working Notes. (2013)
2. Goeriot, L., Kelly, L., Li, W., Palotti, J., Pecina, P., Zuccon, G., Hanbury, A., Jones, G.J.F., Mueller, H.: Share/clef ehealth evaluation lab 2014, task 3: User-centred health information retrieval. CEUR-WS(2014)
3. Palotti, J., Zuccon, G., Goeriot, L., Kelly, L., Hanbury, A., Jones, G.J.F., Lupu, M., Pecina, P.: CLEF eHealth Evaluation Lab 2015, task 2: Retrieving Information about Medical Symptoms. In: CLEF 2015 Online Working Notes. (2015)
4. Lorraine G., Liadh K., Hanna S., Leif H., Aurelie N., Cyril G., Joao P., Guido Z.: Overview of the CLEF eHealth Evaluation Lab 2015. In: CLEF 2015 - 6th Conference and Labs of the Evaluation Forum. Lecture Notes in Computer Science (LNCS), Springer (2015)
5. Chen, Q., Hu, Q.M., Pei, Y.J., Yang, Y., He, L.:ECNU at TREC 2014: Microblog Track. (2014)
6. Shen, W., Nie, J.Y., Liu, X.H., Liui, X.: An investigation of the effectiveness of concept-based approach in medical information retrieval GRIUM@CLEF2014eHealthTask 3. In: Proceedings of the ShARe/CLEF eHealth Evaluation Lab. (2014)
7. Choi, S., Choi, J.: Exploring effective information retrieval technique for the medical web documents: SNUMedinfo at CLEFeHealth2014 Task 3. In: Proceedings of the ShARe/CLEF eHealth Evaluation Lab. (2014)
8. Breiman, L.:Random Forests. Machine Learning, 45(1):5-32, (2001)