DUTIR at the Session track in TREC 2011

Wenfei Liu, Hongfei Lin , Yunlong Ma, Tianshu Chang

School of Computer Science and Technology,
Dalian University of Technology, Dalian 116024
{wenfei.liu, kevinma, tianshuchang}@mail.dlut.edu.cn hflin@dlut.edu.cn

Abstract. This paper presents the DUTIR submission to TREC 2011 Session Track, the task is to test whether systems can improve their performance for a given query by using previous queries and user interactions with the retrieval system. We use language model as the basic retrieval model. Query Expansion is applied to reformulate the original query. Some other technologies like Pseudo Relevance Feedback (PRF) are also applied. The results show that our methods are effective, but the results of Query Expansion method using previous queries and user interactions are unanticipated.

1 Introduction

Users often begin an interaction with a search engine with a sufficiently under-specified query that they will need to reformulate their query and/or information need several times before they find either the thing or everything they are looking for. A session is a sequence of reformulations (along with any user interaction with the retrieval system) in service of satisfying an information need.

The goal of Session Track is to test whether systems can improve their performance for a given query by using previous queries and user interactions with the retrieval system (including clicks on ranked results, dwell times, etc). There is a single task which contains four subtasks this year (RL1, RL2, RL3 and RL4), you can see the TREC 2011 Session Track Guidelines[1] for detail.

This paper describes the word done by members of DUTIR. Our goal of participating in the TREC 2011 Session Track is to get the ability to processing of large-scale corpus and do some research on query sessions. As the limited of the size of our Hadoop[2] cluster, we conduct our experiments on the smaller “Category B” collection of 50 million documents while the track encourages participants to use the whole ClueWeb09 collection. We complete the task by using Language Model, PRF, Query Expansion and so on.

The remainder of this paper is organized as follows. In Section 2, we describe the methods used for Session Track in detail. In Section 3, we demonstrate the results of our experiment and give some corresponding analysis. Finally, we give a brief conclusion and a plan for future work in Section4.

2 Our Methods for Session Track

2.1 Preprocessing of corpus

In order to improve the efficiency of indexing and retrieval, we use a Hadoop cluster which consists of 10 PC to preprocess the Clueweb corpus. There are some other processes like

extracting text from web page, stemmer, removing stop words and so on. We borrow the methods used in Ivory Platform[3] which is a a Hadoop toolkit for web-scale information retrieval research.

2.2 Retrieval Model and PRF

As we know, Language Model can get a better performance in some track tasks in general. So we use Language Model[4] as the basic retrieval model, and dirichlet smooth[5] is used in the Language Model. The equation of computing document score is denoted as follows:

$$LogP(Q \mid D) = \sum_{i=1}^{n} \log \frac{f_{qi,D} + u \frac{c_{qi}}{|C|}}{|D| + u} \tag{1}$$

Where u is set to 1000 in the experiment.

The Pseudo Relevance Feedback model used in our experiments is RM[6] , in this model, we consider the top 50 documents returned by the first retrieval as pseudo relevant documents, and select the top 30 terms as expansion terms, then construct a new query for retrieving again.

2.3 Query Expansion Method

Query expansion is a technology to match additional documents by expanding the original search query. People can use pseudo relevant documents or some other resources(such as Wikipedia, social tags). In the Session track task, we can rely on the information contained in previous queries and user interactions, we use Query Expansion method in the last three subtasks.

In RL2 subtask, we select expansion terms from the query session prior to the current query, the terms occurs in the query which is more close to (measured by time of occurrence) the current query will get a higher weight, the equation of computing term weight is denoted as follows:

$$Weight(q' \mid Q') = \alpha * Weight(q' \mid Q) + \beta * \frac{\frac{1}{N} \sum_{i=1}^{N} \frac{1}{\sqrt{i}} Score(q')}{MAXSCORE} \tag{2}$$

Where Q' is the expansion query; q' is the expansion query terms; N is the number of prior queries; $Score(q')$ is the term weight in one prior query; $Weight(q'|Q)$ is the term weight in original query. In the experiment, we set α to 0.8, β to 0.2.

In RL3 subtask, we consider a prior query and its corresponding ranked list as a big document. The term weighting function is the same as equation 2.

In RL4 subtask, we make a special handling for clicked urls, if one term q' occurs in the corresponding title and snippet, the weight of term q' selected from this prior query will be negative. In equation 2, the corresponding Score(q') will be negative. We think that if the user have clicked the url displayed in the prior query, he should be not satisfied with this result. The weight of term in this result should be lower. Otherwise, we haven't consider the content of the corresponding web pages. According to the observation, we find that there is much more noisy information in the web pages.

3 Experiment Results

We have just submitted one group result for four subtasks. Two approaches are given to evaluate the track results:

(a) They computed evaluation scores by considering relevant those documents that are relevant to any subtopic or the general topic; if a document is relevant to more than one subtopic then the maximum grade is considered as the relevance grade of the document; the runs evaluated in this way are called “allsubtopics”.

(b) They computed evaluation scores by considering relevant those documents that are relevant to the subtopic(s) that the <current query> corresponds to; as before if a query corresponds to more than one subtopic and a document is relevant to more than one of these subtopics the maximum grade is considered as the relevance grade of the document; the runs evaluated this way are called “lastquerysubtopics”.

Table 1. Result of Allsubtopics(nDCG@10)

Result	RL1	RL2	RL3	RL4
Middle	0.3056	0.3130	0.3084	0.3263
DUTIR	0.3498	0.3592	0.3492	0.3511
Max	0.3663	0.4061	0.4086	0.4320

Table 1 shows the result of “allsubtopics”, the Middle result presents the middle value of all results submitted, DUTIR presents our results and the Max result presents the max value of all results. We can get that each of the four subtasks is better than the middle value, our result of RL1 is close to the max value. However, the results of Query Expansion method using previous queries and user interactions are unanticipated, they are just a little higher than the result of RL1. The Middle results have the similar phenomenon, while the Max results meet the expectations, the results using previous queries and user interactions are better than the result of RL1 obviously.

Table 2. Result of Lastquerysubtopics (nDCG@10)

Result	RL1	RL2	RL3	RL4
Middle	0.2208	0.1923	0.1859	0.1972
DUTIR	0.2636	0.2653	0.2563	0.2575
Max	0.2758	0.3034	0.3062	0.3051

Table 2 shows the result of “lastquerysubtopics”, they are similar to the results displayed in Table 1. A small difference is that the results using previous queries and user interactions are worse than RL1 results either in our results or in the Middle results. We think that our Query Expansion method has brought in some noise, we will research on how to clean the noise in the future.

4 Conclusions

In this paper, we present our submission to the TREC 2011 Session Track. We use some technologies like Language Model, PRF, Query Expansion to retrieval on the Clueweb B corpus. The results show that our methods can get a good performance. Because of the noise in query sessions and we haven’t do some work on understanding the query intent of users, the results of Query Expansion method using previous queries and user interactions are unanticipated. Our future work will focus on how to clean the noise in query sessions and understand the query intent.

References

[1] TREC 2011 Session Track Guidelines. http://ir.cis.udel.edu/sessions/guidelines.html.

[2] Apache Hadoop. http://hadoop.apache.org.

[3] Ivory Platform. http://lintool.github.com/Ivory.

[4] J. M. Ponte and W. B. Croft. A language modeling approach to information retrieval. In Proceedings of SIGIR1998, pages: 275-281, 1998.

[5] C. Zhai and J. Lafferty. A study of smoothing methods for language models applied to information retrieval. ACM Transactions on Information Systems, 22(2), pages: 179-214.

[6] Lavrenko V, Croft W B. Relevance based language models[C]. Proceedings of the 24th Annual International Conference on Research and Development in Information retrieval, New Orleans, Louisiana, USA, 2001: 403–410.

[7] Yuan Lin, Hongfei Lin, Song Jin, Zheng Ye. Social annotation in query expansion: a machine learning approach[C]. Proceedings of the 34th Annual International Conference on Research and Development in Information Retrieval, Beijing, China, 2011: 405-414.