International Journal of Latest Trends in Engineering and Technology
Special Issue SACAIM 2017, pp. 162-164
e-ISSN:2278-621X

A STUDY ON QUERY SUGGESTION AND EXPANSION IN INFORMATION RETRIEVAL

Mr.Ruban[1], Swathi A[2] & Misrina M Y[3]

Abstract: Information Retrieval is the activity of obtaining information whose content matches with a user query from a large collection of information sources. As creating well-designed queries is difficult for most users, it is mandatory to use query expansion to obtain relevant information. Query Expansion techniques are widely applied for improving the efficiency of the textual information retrieval systems. In information retrieval, a query doesn't uniquely identify a single object, instead several objects. In order to prevent these problems query expansion comes into picture. In this paper few techniques of query suggestion and query expansion are discussed.
Keywords: Information Retrieval, Query Expansion, Information extraction, Text analysis.

1. INTRODUCTION

Internet has become one of the most important part of our routine. One of the most common uses of the internet is gathering information. Almost any kind of information on any topic can be found on the internet. So the task of information retrieval systems is to retrieve any information from the internet. Information retrieval is the process of obtaining particular information from a large collection of information resource according to the users query. Since most of the web users are naive, it becomes difficult to pose queries which are always correct. So a need for a method to effectively transform the queries will be helpful to retrieve data. Query expansion is one of the modern approaches in information retrieval for improving the effectiveness of information retrieval system. Query Expansion methods helps to add terms to the query, with the goal of improving the search query thus helping to retrieve more efficient results. Many methods are available to do query expansion.

Pseudo-relevance feedback assumes that top-ranked documents returned for the query are relevant, and uses the terms extracted from those documents for expansion. However, PRF methods are not usually applicable in practice, as they add significant overhead to the retrieval system. An alternative that has been already proposed is using external data for expansion. In particular, the methods are using query log and snippets for query expansion. Using search engine as external evidence to expand the query has many benefits: (1) the query logs of search engines reflect the interests of a large number of users. Different users may submit various queries to express the same information need. Therefore, the query can be expanded by using the wisdom of crowd.(2) Current commercial search engines use sophisticated ranking functions on billions of documents. The higher quality of feedback documents can have a direct impact on the performance of query reformulation methods. (3) The traditional PRF models need two retrieval processes per query. In the next sections we will be explaining the different query expansion techniques, query suggestion, previous work that is already done and finally the conclusion.

2. QUERY EXPANSION

Query expansion (or query reformulating) is the process of supplementing the original query with additional terms, and it can be considered as a method for improving retrieval performance. Current information retrieval systems are restricted by many factors reflecting the difficulty to satisfy user requirements expressed by short queries. Expanding these user queries is a common technique in information retrieval to cover the gap between the original user query and his need of information. The most used technique is where the original user query is expanded with new terms extracted from different sources. Efthimiadis has done a complete review on the classical techniques of query expansion. The main problem of query expansion is that in some cases the expansion process worsens the query performance. Improving the robustness of query expansion has been the goal of many researchers in the last few years. There are two service modes in information retrieval i.e one is based on the keyword and another is based on the hierarchical directories. In the first mode, users enter keywords to obtain necessary information but the length of the query provided by the users will be maximum three to four words where in there are high chances to obtain irrelevant information. At present, many researches show that query expansion, which extracts terms from a subset of the initial retrieval results for expansion, can enhance the retrieval performance of short queries effectively.

[1] Asst.professor,IT department, St.Aloysius College,AIMIT,Mangaluru,Karnataka,India.
[2] Student, St.Aloysius College,AIMIT,Mangaluru,Karnataka,India.
[3] Student, St.Aloysius College,AIMIT,Mangaluru,Karnataka,India.

3. QUERY SUGGESTION

Another approach to solve the inefficiency and inaccuracy in the information retrieval system is query suggestion. It is very common for a user to reformulate their query when they didn't receive desired result from their original query. The system can improve the user's searching effort by providing suggestions by guessing the user intention, according to users past search. A series of experiments on man-machine interaction of information retrieval system indicate that instead of automatic query expansion, users prefer to use query suggestion to improve the effectiveness of their original query. Providing effective and useful query suggestion is the most important motivation for query suggestion [.A].Click-through based query suggestion: Click-through based query suggestion focus on mining user's click pattern in a search log. Traces of the click-through for each query are recorded. The clicked URL can be used to exploit the relationship between different queries.

If the queries in the same cluster are classified as the same or similar topic, the queries within the same cluster will be used as the query suggestion. Leung, Ng and Lee proposed a method that provides personalization query suggestion based on a personalized concept based clustering technique. Instead of providing similar suggestion to every user, they clustered user click-through data to predict user intention and preference. A personalized query suggestion was given to every user based on their past behaviour.[B].Session based query suggestion :Session based query suggestion based on the assumption on every search query in the same session is related to each other in one way or another . A few assumptions can be made regarding session based query suggestion. (a) When a number of queries in the same session in a short time are usually submitted by the same user. (b) In a same session, the user often tried to change their query or try a new query to get a better result. (c) Query submitted by a user in the same session usually is about a single topic.

4. RELATED WORK

Relevance feedback is a well established approach for expanding query by choosing important terms or expression, attached to the documents retrieved from original query that had been identified as relevant by the users or the system assume the top ranked documents as relevant. In order to make a successful expansion, a few assumptions must be made. First of all, the user must have sufficient knowledge about the document they desire to compose the initial query. Misspelling, cross-language information retrieval, and also mismatch of searcher's vocabulary versus collection vocabulary cannot be solved just using relevance feedback. There are three different types of feedback, ad hoc or blind feedback, implicit feedback and explicit feedback.

In Literature Query expansion are studied in different ways for instance query expansion methods are classified into Automatic Query Expansion (AQE) methods and Interactive Query Expansion Methods (IQE) [1] Earlier studies reveal that many of the automatic query expansion methods rely on the Relevance Feedback techniques [2] proposed by salton and Buckley , in which the terms featuring prominently in documents marked relevant by the user are automatically added to the query.

Later Srinivasan came up with a Retrieval Feedback technique [3] that adds terms from the top relevant documents to the query. This technique has shown considerable improvement in many retrieval tasks. Query logs was used as a means of query expansion by Hangs et al [4]. Later Huang et al [5] proposed a query expansion algorithm of pseudo relevance feedback based on matrix-weighted association rule mining.

However in the year 2001 Aronson [6] proved that query refinement that is based on ontology is much more efficient than the other methods that were available. Using ontology for query expansion goes back to 1994 where Voorhees [7] attempted using the Domain independent ontology WordNet for query expansion. Since then there has been some works done in this area. The word sense information and the ontology was used for query expansion by Navigli and Velardi[8]. They succeeded in using ontology to extract the semantic domain of a word and then the query is expanded further using co-occurring words. Further Query refinement techniques based on domain and geographical ontology was studied by Fu,G et al [9]. The Domain ontology was modelled after tourism which consists of some non-spatial terms such as "near" whereas the geographical ontology consists of some spatial terms such as place names. A domain specific ontology based on Stockholm University Information systems (SUiS) was developed by Nilsson et al[10].

Wikipedia is the biggest encyclopedia available freely on the web. Though developed by people around the globe, Wikipedia content is well structured and correct. Being developed by people is an advantage, its growing rapidly and contains wide varieties of topics. These features makes Wikipedia a good knowledge source for query expansion.Recently, many approaches are being developed which use Wikipedia for query expansion. Li et al .[11] proposed query expansion using Wikipedia by using the category assignments of its articles. The base query is run against a Wikipedia collection and each category is assigned a weight proportional to the number topic ranked articles assigned to it. Articles are then re ranked based on the sum of weights of the categories to which each belongs. The method shows improvement over PRF in measures favouring weak queries. Few recent works are given in the following table.

Sl.No	Query Reformulation	Data Sources	Term Extraction Methodology	Term Representation
1.	Li et al.[11]	Wikipedia	Title & in/out links of Wikipedia pages	Individual terms and phrases
2.	Pal et al [12]	WordNet	Synonyms,holonyms and	Individual terms

			meronyms of the query terms	
3.	Paik et al.[13]	Top retrieved set of documents	All terms in feedback documents	Individual terms
4.	Yin et al.[14]	Query logs & Snippets	All terms in top retrieved snippets	Individual terms
5.	Xiong & Callan[15]	FreeBase	All terms in top retrieved documents	Individual terms

5. CONCLUSION

The Query Expansion techniques are being used in most of the commercial search applications. All the techniques proves that the improvements in search results happen primarily because of the new terms that are added. As new and new sources of finding these candidate terms are explored in the research, we firmly believe that the lexical resources and the web resources will be a powerful repository to do query expansion effectively.

6. REFERENCES

[1] Query Expansion, Efthimiadis, E.N(1996), Annual Review of information science and Technology.

[2] Salton G & Buckley C (1990). Improving Retrieval performance by relevance feedback, Journal of the American society for Information Science.

[3] Padmini Srinivasan, "Retrieval Feedback in MEDLINE", Journal of the American Medical Informatics Association, 3(2):157-167, 1996c, doi: 10.1136/jamia.1996.96236284

[4] C.Hang, W. Ji-Rong and N. Jian-Yun, "Probabilistic query expansion using query logs", proceedings of the eleventh international conference on World wide Web(2002)

[5] M.Huang, x.Yan and S.Zhang, "Query expansion of pseudo Relevance feedback based on matrix-weighted association rules mining", Journal of Software , 20(7):1854-1865 (2009)

[6] Padmini Srinivasan, "Retrieval Feedback in MEDLINE", Journal of the American Medical Informatics Association, 3(2):157-167, 1996c, doi: 10.1136/jamia.1996.96236284

[7] E.M Voorhees, "Query Expansion using lexical-semantic relations", proceedings of the 17th annual international ACM SIGIR conference on Research and development in information Retrieval(1994)

[8] R.Navigli and P.Velardi, "An analysis of ontology-based query expansion strategies", workshop on Adaptive Text Extraction and Mining(2003).

[9] Lin Fu,Dion Hoe-Lian Goh and Schubert shouo-Boon Foo, "Evaluating the effectiveness of a collaborative querying environment", proceedings of the 8th international conference on Asian digital libraries [2005].

[10] K.Nilsson, H.Hjelm and H.Oxhammar, "SuiS – cross-language ontology-driven information retrieval in a restricted domain", Proceedings of the 15th NODALIDA conference(2005)

[11] Li,Y,Luk, W.P.R.,Ho,K.S.E.,and Chung ,F.L.K.[2007].Improving weak ad-hoc queries using Wikipedia as external corpus.In proceedings of 30th annual international ACM SIGIR conference on research and development in information retrieval,SIGIR '07,pages 797-798,New York,NY,USA.ACM.

[12] Pal,D.,Mitra,:Improving Query expansion using WordNet .Journal of the association for information science and technology[2014]

[13] Paik,J.H.,Pal,D.,Parui,S.K:Incremental blind feedback : an effective approach to automatic query expansion .ACM transactions on Asian language information processing.

[14] Yin, Z., Shokouhi, M., Craswell, N.: Query expansion using external evidence. In: European Conference on Information Retrieval, pp. 362{374. Springer (2009)

[15] Xiong, C., Callan, J.: Query expansion with freebase. In: Proceedings of the 2015 International Conference on The Theory of Information Retrieval, pp. 111{120. ACM (2015).