Relevance Feedback in Information Retrieval Systems

Working Paper · October 2017

CITATIONS
2

READS
15,542

1 author:

Anmol Hamid
Bahria University
7 PUBLICATIONS **79** CITATIONS

SEE PROFILE

Relevance Feedback in Information Retrieval Systems

Anmol Hamid, MS(CS)
Department Of Computer Science
Bahria University
Islamabad, Pakistan
anmolhamid168@yahoo.com

***Abstract*—Relevance in Information Retrieval defines how much the retrieved information meets the user requirements. Relevance feedback is the feature that includes in many IR systems. It takes the results from the query and user gave feedback and then system checks whether this retrieved information is relevant enough to execute another new query. The basic architecture of Feedback is shown in Figure 1. There are three types that are discussed in Relevance Feedback are Explicit Feedback (EF), Implicit Feedback (IF), Pseudo/ Binary Feedback.**

I. Introduction

There are large amount of information objects stored electronically and the Information retrieval systems allows the user to access them. A user enters a query and IR system returns the results according to the need of the user. These objects may referred as images, piece of text, webpages, some segment of the video. This process of IR systems has uncertainty means that the person who is searching does not express the idea for retrieving the result and sometimes searchers may not be able to find the exact information they need but when the documents or set of documents the system retrieved are shown up in results then they indicate how much the information is relevant or not This led to the idea of Relevance Feedback (RF) [1].In the information retrieval systems relevance denotes how well the retrieved documents or the set of documents fulfill the users information need. For the retrieval accuracy relevance feedback is playing a very effective role. It is an iterative process that helps to improve the performance and accuracy of the retrieval systems. In the Relevance Feedback approach user submits the feedback and then present that information to the IR systems. System uses this information in two ways i.e. quantitative approach: retrieved the more relevant documents. And qualitative approach: retrieved the similar documents that are relevant to the documents. This process of RF is referred as a cycle of activities. In this process when user submits a query it performs processing and returns results according to the need of user and then takes feedback from the user whether the retrieved results shows relevancy or not according to his/ her query, after this the system reformulate the relevant results based on the users feedback and retrieved more relevant results This process is thus known as the iteration of Relevance Feedback in information Retrieval Systems. In the section II discussed the types of feedback. Section III contains literature review of relevance feedback it describes the approaches used by authors for the getting the effectiveness in the RF and the section IV concludes this work.

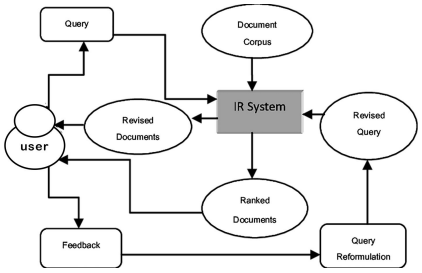

Fig. 1.

II. Types of Relevance Feedback

The types which are mainly discussed in Relevance Feedback are explicit feedback, implicit feedback, and pseudo feedback.

A. Explicit Feedback

In Explicit Feedback user indicate the relevance of the document that are retrieved by a query. There are two ways to indicate the relevance. Binary or Graded relevance system. The document is relevant or not is indicated by Binary relevance feedback. The relevancy of document is indicated by giving some kind of description, by using words or scaling through numbers (such as 1 shows that document is relevant, 2 shows not relevant and so on) this is known as Graded RF

B. Implicit Feedback

This kind of feedback is indicated by noting users behavior. Like what kind of document they view or not. The time spent on viewing a document, browsing actions these kinds of actions can be used to provide relevant information to the user. Thats why it is called Implicit Feedback because it implicitly taken feedback from the actions of user.

C. Pseudo Feedback

The other name of Pseudo Relevance Feedback is blind relevance feedback. The manual part of the feedback is automated in this Feedback process so user gets the improved results. The system finds the relevant documents and assumed that the top k documents that ranked are the relevant ones and then used methodology of the relevance feedback. The procedure of Relevance feedback is

1. Taking the results by initial query and assumed that they are top "k" results.
2. Select top 10-25 terms from these documents by using tf-idf weights
3. Use Query Expansion as method which adds these terms to query and match the returned documents for the new query and then finally returns the output which shows that most relevant documents are returned.

III. LITERATURE REVIEW

Relevance Retrieval in information technology denotes how much the retrieved information meets the user requirements. The relevance feedback is one of the feature of information retrieval system. This research aims to review previous research on Relevance Retrieval. Some of the researchers work, their proposed techniques how to get better results and which models they used are discussed below

Xu, J. a. These authors done a comparison on local document analysis also contains (local feedback (LF) and local context analysis (LCA)) and global techniques in automatic query expansion. They focused on two techniques i.e. the focus on the word relationships known as (global technique) and the documents retrieved by initial query known as the process of local Feedback .Local feedback can be treated as relevance feedback and the most relevant documents are ranked in this process. For the solution of word mismatch problem the method of query expansion is proposed in this paper. In query expansion the query is expanded by using the words and the phrases having similar meaning with those words that occurred in query. By this methods the number of matching words in relevant documents are increased .Their experimental results shows that local feedback when using with local context analysis is more effective and the local context analysis which uses some of the global techniques outperform the simple local feedback techniques in terms of the effectiveness and predictability [2].

Xu, Z. a. These authors surveyed several different Relevance Feedback algorithms for document retrieval and summarize their pros and cons that are 1. The Rocchio Algorithm 2. The SVM Relevance Feedback Algorithm (SVMRF) 3. SVM Active Learning for Relevance Feedback (Active SVMRF) and introduces a novel hybrid RF approach using a support vector machine (HRFSVM), which selects the inexact documents and the most relevant ones and takes feedback from users then ranked the documents for user browsing. They used n Reuters-21578 dataset for the experiments. Their Experiments that applied on Reuters-21578 dataset shows that their Hybrid Relevance Feedback using SVM performs better than other two Relevance Feedback algorithms [3].

Drucker, H. a. done a comparison of (SVM) with Rocchio algorithm, Ide regular and Ide dec-hi in IR using the relevance feedback method. They defined that there problem statement as in a database user wants to retrieve a set of documents Most of the articles in database are relevant to the users need and most of them are not relevant. The relevancy of the document can be obtained by understanding the perception of the user. There results shows that if the initial search obtains some poor results and the visibility of the topic is also low then SVMs performed much better than other techniques. If using TF-IDF approach marginally SVM performs better results than using Ide dec-hi. SVM also performed better when using TF or binary weighting [4].

Lv, Y., & Zhai, C presents that for the improvement of retrieval accuracy Relevance Feedback proved to be very effective. The most important problem in relevance feedback method is to get the optimal balance between the original query and the information from the feedback. In this paper author presents an approach that adaptively predict the optimal balance coefficient for each query and each information that is collected. They reviewed modeling framework and also discuss how to integrate their provided feedback prediction function into these frameworks. The Mixture Model Feedback Method, the KL-Divergence Retrieval Model. They also proposed three heuristics methods that includes 1.discrimination of query 2.discrimination of feedback documents, 3.divergence between query and feedback documents to characterize the balance between query and feedback information. For predicting the balance coefficient they also used a regression (logistic regression) approach. Their results show that relevance feedback coefficient is better than fixed coefficient feedback [5].

Another author Salton, G., & Buckley states that Relevance Feedback is designed in such a way that produced relevant results according to the users feedback. The evaluated data thus included in demonstration of the effectiveness of the various methods. Each operation in the text retrieving is performed iteratively using relevance feedback approach. The basic feedback procedures include Vector Processing Methods (VPM), Probabilistic Feedback Methods. The authors concluded that the relevance feedback is used for processing the queries that are based on previously retrieved relevant documents and none of the relevant documents. These RF methods then evaluated by using six document collections in different areas. A simple vector modification process that adds new query terms and modifies the weight of existing terms appears most useful in this connection [6].

Another research done by Zhou, X. S., & Huang, T. S based on relevance feedback is in the multimedia IR. They tried to put emphasis on the uniqueness of the problem, and done comparison on the assumptions, implementations, and merits of various solutions in the literature. They compile a list of critical issues to consider when designed a relevance feedback algorithm. With a comprehensive review as the main portion, this paper also offers some Issues to Consider when Designing

a Relevance Feedback Algorithm that includes Negative examples, Singularity issue in sample covariance matrix. Feature normalization, Pre-clustering and relevance feedback, Global vs. regional query, Complexity of the nearest neighbor search [7]

.

IV. Conclusions

For achieving accuracy and effectiveness in query modification process Relevance Feedback is an excellent technique. From the Literature review it should be noted that the techniques which authors proposed improved the accuracy in the Relevance feedback and their results are better than the previous results. When using SVM with Relevance feedback the Results are more efficient and accurate or by combining different algorithms with SVM outperformed the other techniques.

References

[1] I. Ruthven and M. Lalmas, "A survey on the use of relevance feedback for information access systems," *The Knowledge Engineering Review*, vol. 18, no. 2, pp. 95–145, 2003.

[2] J. Xu and W. B. Croft, "Quary expansion using local and global document analysis," in *ACM SIGIR Forum*, vol. 51, no. 2. ACM, 2017, pp. 168–175.

[3] Z. Xu, X. Xu, K. Yu, and V. Tresp, "A hybrid relevance-feedback approach to text retrieval," in *ECIR*. Springer, 2003, pp. 281–293.

[4] H. Drucker, B. Shahrary, and D. C. Gibbon, "Support vector machines: relevance feedback and information retrieval," *Information processing & management*, vol. 38, no. 3, pp. 305–323, 2002.

[5] Y. Lv and C. Zhai, "Adaptive relevance feedback in information retrieval," in *Proceedings of the 18th ACM conference on Information and knowledge management*. ACM, 2009, pp. 255–264.

[6] G. Salton and C. Buckley, "Improving retrieval performance by relevance feedback," *Readings in information retrieval*, vol. 24, no. 5, pp. 355–363, 1997.

[7] X. S. Zhou and T. S. Huang, "Relevance feedback in image retrieval: A comprehensive review," *Multimedia systems*, vol. 8, no. 6, pp. 536–544, 2003.