Chapter 2

Uncertain Confidence Network–Based Collaborative Information Retrieval Relevance Feedback Algorithm

Fatiha Naouar
MARS Research Laboratory, Tunisia

Lobna Hlaoua
MARS Research Laboratory, Tunisia

Mohamed Nazih Omri
https://orcid.org/0000-0001-7803-0179
MARS Research Laboratory, Tunisia

ABSTRACT

Collaborative retrieval allows increasing the amount of relevant information found and sharing history with others. The collaborative retrieval can reduce the retrieval time performed by the users of the same profile. This chapter proposes a new relevance feedback algorithm to collaborative information retrieval based on a confidence network, which performs propagation relevance between annotations terms. The main contribution in this work is the extraction of relevant terms to reformulate the initial user query considering the annotations as an information source. The proposed model introduces the concept of necessity that allows determining the terms that have strong association relationships estimated to the measure of a confidence. Since the user is overwhelmed by a variety of contradictory annotations, another contribution consists of determining the relevant annotations for a given evidence source. The experimental study gives very encouraging results.

DOI: 10.4018/978-1-7998-1021-6.ch002
INTRODUCTION

The optimization of retrieval time may be caused by the formulation of collaborative queries through dialogue and the mutual consultation of queries sent and the search results received by everyone. It also enables the sharing of search history by displaying the search results put in order by relevance. One of the most popular tools for sharing results and personal judgments is the annotations. Several problems exist with respect to Collaborative Retrieval (hereafter CR). There is, in particular, the problem of the relevance of information. Indeed, a user always finds problems in meeting his/her needs in relevant information.

As in classical Information Retrieval (hereafter IR), the Collaborative Information Retrieval (hereafter CIR) is designed to return and display a set of documents to a user according to his need. On the other hand, the users of a retrieval system, are not always specialists in this field (Lin & Wang, 2006), they can make a bad choice of terms to express its information needs. A reformulation of query has been necessary, since the initial user query can return unsatisfactory results. It’s a question then of amending the original user query and this happens by adding meaningful terms to improve the initial result returned. Three types of approaches use different techniques to select the terms to be added to a query. The difference between these algorithms is based on the choice of terms to be used to reformulate the query. In the first algorithm, the mechanism used allows the user to select the terms that are related to the original query terms. In the second algorithm, the re-injecting of the terms is to be selected from a terminological resource (like semantic network, thesaurus, ontology, ...). While in the third algorithm the terms are selected for a reformulation from the results of previous research (also called relevance feedback).

The first type of algorithm is based on the overall analysis of the collection of documents considered (Baeza-Yates & Ribeiro-Neto., 1999). This algorithm consists of analyzing the entire set of documents, in the collection, to extract relevant terms to add to the initial query. Two techniques are used: the similarity thesaurus and the statistical thesaurus. The most prevalent technique is based on statistical analysis of query logs (Cui, Wen, Nie & Ma, 2002). The objective is to automatically reformulate a query by adding terms of the documents that are correlated with the query terms. The correlation is based on a conditional probability. The second type of algorithm, in the literature, uses terminological resources such as the anthologies or the thesaurus containing the vocabulary to enrich queries (Abderrahim, 2013; Harb, Khaled & Nagdy, 2011). This type of algorithm uses anthologies with the equivalence relations and the subsumption (Navigli & Velardi, 2003) to find the similar terms to the original query. In (Boughammoura, Omri & Hlaoua, 2011, 2012, 2013) the authors propose a new rendering algorithm of deep Web forms which is easy to interpret by user and reflects the exact meaning of query. The final type of algorithm (Omri, 1994) is based on the principle of relevance feedback which also aims to reformulate an initial query to better reflect the contents of the collection. To do this, the user begins by submitting its initial application to the system that renders a first set of documents that the user must judge as relevant or irrelevant. The judgment of the relevance of the documents initially returned is used to select the terms to add to the original query (Lin & Wang, 2006).

In this context, the authors suggest to improve the research performance, using the relevance feedback to extend the initial query. This technique consists of extracting terms from documents deemed relevant and considered in a new extended query. This technique has been used in classical Information Retrieval (Rocchio, 1971). In their work, Singh and Sharan (2015) considered the retrieved documents, by a first search, the top ranked as relevance feedback. This is achieved by the combination of the two approaches namely, the co-occurrence of initial query terms with the terms constituting the corpus and a contextual window based approach is used to select the terms from the tops feedback documents. The