Information Processing and Management 48 (2012) 969–977

Contents lists available at SciVerse ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/infoproman

A split-list approach for relevance feedback in information retrieval

H.C. Wu [a,*], R.W.P. Luk [a], K.F. Wong [b], J.Y. Nie [c]

[a] *Department of Computing, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong*
[b] *Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong, Shatin, NT, Hong Kong*
[c] *Department of Computer Science and Operations Research (DIRO), The University of Montreal, Montreal, Canada*

ARTICLE INFO

Article history:
Received 23 September 2010
Received in revised form 28 March 2012
Accepted 28 March 2012
Available online 13 May 2012

Keywords:
Active feedback
Document-context
Experimentation

ABSTRACT

In this paper we present a new algorithm for relevance feedback (RF) in information retrieval. Unlike conventional RF algorithms which use the top ranked documents for feedback, our proposed algorithm is a kind of active feedback algorithm which actively chooses documents for the user to judge. The objectives are (a) to increase the number of judged relevant documents and (b) to increase the diversity of judged documents during the RF process. The algorithm uses document-contexts by splitting the retrieval list into sub-lists according to the query term patterns that exist in the top ranked documents. Query term patterns include a single query term, a pair of query terms that occur in a phrase and query terms that occur in proximity. The algorithm is an iterative algorithm which takes one document for feedback in each of the iterations. We experiment with the algorithm using the TREC-6, -7, -8, -2005 and GOV2 data collections and we simulate user feedback using the TREC relevance judgements. From the experimental results, we show that our proposed split-list algorithm is better than the conventional RF algorithm and that our algorithm is more reliable than a similar algorithm using maximal marginal relevance.

1. Introduction

Relevance feedback (RF) is known to be effective for improving retrieval effectiveness (Harman, 1992; Rocchio, 1971; Salton & Buckley, 1990). RF requires effort and time from users to judge whether a document is relevant to the user's information need. When a user judges a particular document to be non-relevant, in the point of view from users, some effort and time is wasted because the document provides no relevant information to the user. As a result, it is better to have more relevant documents to be judged by the user in the RF process. However, judging two very similar relevant documents also wastes effort and time from users because the information contained in the two documents is nearly the same. Thus, no additional relevant information is provided to the user. Therefore, two main factors would affect the user's satisfaction in the RF process:

(a) the number of relevant documents (the more the better), and
(b) the diversity of the documents (the more diverse the better).

In conventional RF, the user would judge documents from the top ranked ones in the result list by assuming that the top ranked documents contain more relevant information. In some cases, the top ranked documents are very similar to each

[*] Corresponding author. Tel.: +852 2766 5143; fax: +852 2774 0842.
E-mail addresses: cshcwu@comp.polyu.edu.hk (H.C. Wu), csrluk@comp.polyu.edu.hk (R.W.P. Luk), kfwong@se.cuhk.edu.hk (K.F. Wong), nie@IRO.UMontreal.CA (J.Y. Nie).

0306-4573/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.ipm.2012.03.007

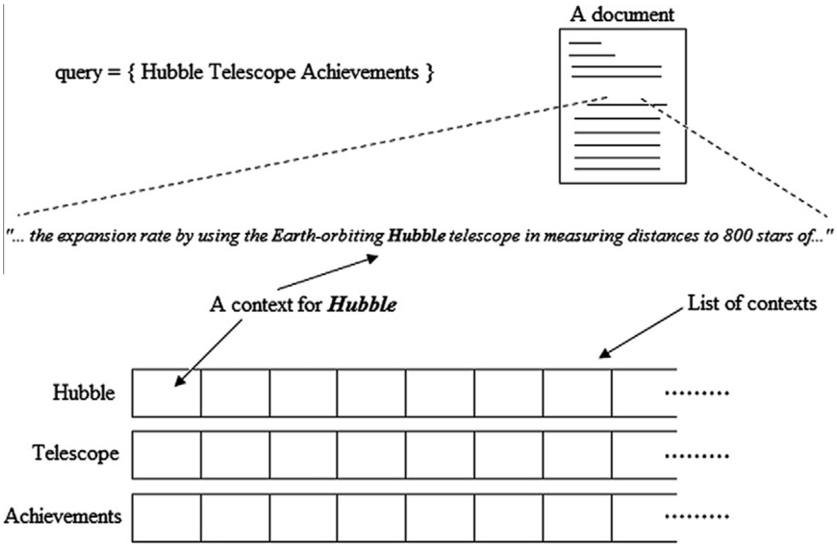
A document
query = { Hubble Telescope Achievements }
"... the expansion rate by using the Earth-orbiting Hubble telescope in measuring distances to 800 stars of..."
A context for Hubble
List of contexts
Hubble
Telescope
Achievements
....

Fig. 1. Illustration of the lists of document-contexts for the query "Hubble Telescope Achievements".

other or even identical. Judging the relevance of the nearly identical documents provides no additional useful information to both the user and the retrieval system. Therefore, the set of documents used for RF may not be the top ranked ones. This is called active feedback (Chen & Lu, 2010; Shen & Zhai, 2005; Sia, Zhu, Chi, Hino, & Tseng, 2007; Xu & Akella, 2008a, 2008b; Xu, Akella, & Zhang, 2007; Zhang & Zhang, 2010) in which the retrieval system actively chooses suitable documents for the user to judge for relevance.

Our proposed algorithm is a kind of active feedback algorithm which uses document-contexts by splitting the retrieval list into sub-lists according to the query term patterns that exist in the top ranked documents. Fig. 1 shows an example of the lists of document-contexts for the query "Hubble Telescope Achievements". Note that only lists for single query terms are shown. Document-contexts are used because we believe that relevant information is located around query terms (Wu, Luk, Wong, & Kwok, 2007, 2008). By splitting the retrieval list into sub-lists, we hope that the proportion of relevant documents in a particular sub-list will be higher than that of the others. Therefore, the scores of document-contexts in the particular sub-list will be higher. By that we can increase the number of relevant documents judged by the user in the RF process. Also, the set of documents being judged by the user in the split-list approach is different from the set of top ranked documents. Hence, it can increase the diversity of the documents being judged.

The rest of the paper is organized as follows. In Section 2, we describe some active feedback algorithms including the gapped method and cluster method in Shen and Zhai (2005) and the maximal marginal relevance method (Carbonell & Goldstein, 1998). Section 3 describes our split-list approach using document-contexts. We show the experiment results in Sections 4 and 5 concludes the paper.

2. Related work

In this section we review some active feedback algorithms which do not use the top N_{rf} ranked documents for feedback. In 2007, Xu et al. (2007) proposed a method to incorporate diversity and density measures in performing relevance feedback. They showed improvements in retrieval effectiveness over other active feedback methods using cross-validation. In our experiments, instead of using cross-validation, we calibrate the parameters using one collection and use the fixed parameter values for other collections.

By modeling active feedback using the risk minimization framework for retrieval (Lafferty & Zhai, 2001), Shen and Zhai (2005) formalized the active feedback problem as a decision making problem and experimented two active feedback algorithms. One is Gapped-Top-N_{rf} and the other one is N_{rf}-Cluster-Centroid. Maximal marginal relevance (MMR) (Carbonell & Goldstein, 1998) is also described in this section.

2.1. Gapped-Top-N_{rf}

In the Gapped-Top-N_{rf} algorithm, instead of judging the top N_{rf} ranked documents like the baseline RF algorithm, a gap of G documents is introduced between two judged documents. As a result, the ith judged document is ranked at $i + (i - 1)G$. For example, if $G = 2$, the set of judged documents will have rank numbers 1, 4, 7, ..., $N_{rf} + (N_{rf} - 1)2$ in the retrieval list. With $G = 0$, the Gapped-Top-N_{rf} is essentially the baseline RF algorithm using the top N_{rf} ranked documents. The Gapped-Top-N_{rf}