Information Processing and Management 45 (2009) 413–426

Contents lists available at ScienceDirect

Information Processing and Management

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

Assigning appropriate weights for the linear combination data fusion method in information retrieval

Shengli Wu [a,*], Yaxin Bi [a], Xiaoqin Zeng [b], Lixin Han [b]

[a] School of Computing and Mathematics, University of Ulster, Newtownabbey, UK
[b] Department of Computer Science, Hohai University, Nanjing, China

ARTICLE INFO

Article history:
Received 16 June 2008
Received in revised form 13 February 2009
Accepted 20 February 2009
Available online 21 March 2009

Keywords:
Data fusion
Information retrieval
The linear combination method
Weight assignment

ABSTRACT

In data fusion, the linear combination method is a very flexible method since different weights can be assigned to different systems. However, it remains an open question which weighting schema should be used. In some previous investigations and experiments, a simple weighting schema was used: for a system, its weight is assigned as its average performance over a group of training queries. However, it is not clear if this weighting schema is good or not. In some other investigations, different numerical optimisation methods were used to search for appropriate weights for the component systems. One major problem with those numerical optimisation methods is their low efficiency. It might not be feasible to use them in some situations, for example in some dynamic environments, system weights need to be updated from time to time for reasonably good performance. In this paper, we investigate the weighting issue by extensive experiments. The key point is to try to find the relation between performances of component systems and their corresponding weights which can lead to good fusion performance. We demonstrate that a series of power functions of average performance, which can be implemented as efficiently as the simple weighting schema, is more effective than the simple weighting schema for the linear data fusion method. Some other features of the power function weighting schema and the linear combination method are also investigated. The observations obtained from this study can be used directly in fusion applications of component retrieval results. The observations are also very useful for optimisation methods to choose better starting points and therefore to obtain more effective weights more quickly.

1. Introduction

Information retrieval as a core technology has been widely used for the WWW search services and digital libraries. In recent years, an increasing number of researchers have been working in this area and many different techniques have been investigated to improve the effectiveness of retrieval. Quite a large number of retrieval models have been proposed and experimented with test document collections. For example, in the book "Modern Information Retrieval" written by Baeza-Yates and Ribeiro-Neto (1999), 11 different retrieval models were discussed. In addition, some other aspects such as user relevance feedback, document representation, query representation, query expansion, phrase recognition, thesaurus, context analysis, structure analysis, link analysis, and so on, may have a considerable impact on retrieval results. It is very likely that no single information retrieval system is able to deal with all these aspects at the same time, let along deal with them in the same way as some other systems. Therefore, many information retrieval systems developed are close in effectiveness but

* Corresponding author. Tel.: +44 2890366585.
E-mail address: s.wu1@ulster.ac.uk (S. Wu).

0306-4573/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ipm.2009.02.003

different from each other in implementation. In such a situation, data fusion, which uses a group of information retrieval systems to search the same document collection, and then merges the results from these different systems, is an attractive option to improve retrieval effectiveness. Recently, meta-search becomes an application of the data fusion technique. If the document collections across different web services are more or less the same, then the data fusion methods can be applied directly; if the document collections are quite different, then some variations of the data fusion methods can be used for obtaining more effective results (Wu & McClean, 2007).

Quite a few data fusion methods such as CombSum (Fox, Koushik, Shaw, Modlin, & Rao, 1993; Fox & Shaw, 1994), CombMNZ (Fox et al., 1993; Fox & Shaw, 1994), the linear combination method (Bartell, Cottrell, & Belew, 1994; Vogt & Cottrell, 1998, 1999), Borda fusion (Aslam & Montague, 2001), the probabilistic fusion method (Lillis, Toolan, Collier, & Dunnion, 2006), the correlation method (Wu & McClean, 2005, 2006a), Markov chain-based methods (Dwork, Kumar, Naor, & Sivakumar, 2001; Renda & Straccia, 2003), Condorcet fusion (Montague & Aslam, 2002), and the multiple criteria approach (Farah & Vanderpooten, 2007) have been proposed, and extensive experiments using TREC data have been reported to evaluate these methods. Experimental results show that, in general, data fusion is an effective technique for improvement of effectiveness, and very often the fused results are better than the best component results involved.

The linear combination data fusion method is a very flexible method since different weights can be assigned to different systems. In some related experiments, (e.g., Aslam & Montague, 2001; Thompson, 1993; Wu & Crestani, 2002; Wu & McClean, 2006a), a simple weighting schema was used: for a system, its weight is set as its average performance over a group of training queries. This weighting schema is straightforward and can be calculated or updated easily, therefore it is especially suitable in a very dynamic environment. However, it has not been investigated how good this weighting schema is.

In some previous researches, different numerical optimisation methods such as golden section search (Vogt & Cottrell, 1998, 1999) and conjugate gradient (Bartell et al., 1994) were used to search suitable weights for component systems. One major drawback of using these optimisation methods is their low efficiency. In some situations, such as the WWW, and digital libraries, where documents are updated frequently, each component system's performance may vary considerably from time to time. The weights for the systems should be updated accordingly. In such a situation, it may not be feasible to use those very time-consuming weighting methods to update weights frequently.

In this paper, we investigate the weighting issue through extensive experiments. The key point is to try to find the relation between performances of component systems and their corresponding weights which can lead to good fusion performance. As we shall see later in this paper, a power function weighting schema with a power of between 2 and 8, is more effective than the simple weighting schema for data fusion. On the other hand, those power function weights can be obtained as efficiently as simple weights. In fact, the simple weighting schema can be regarded as a special case of a power function weighting schema with a power of 1.

The rest of this paper is organized as follows: in Section 2 we review some related work on data fusion, especially on the linear combination method. Section 3 describes the linear combination method and the weighting issue. Section 4 discusses an experiment to evaluate different performance weighting schemas. Further observations from the experiment are discussed in Section 5. Section 6 provides a few conclusive remarks.

2. Previous work on data fusion

Usually relevance-related scores are provided by information retrieval systems for all retrieved documents. Some algorithms, such as CombSum (Fox et al., 1993; Fox & Shaw, 1994), CombMNZ (Fox et al., 1993; Fox & Shaw, 1994), the linear combination method (Bartell et al., 1994; Thompson, 1993; Vogt & Cottrell, 1999; Wu & Crestani, 2002), the probabilistic fusion method (Lillis et al., 2006), and the correlation method (Wu & McClean, 2005, 2006a), make use of relevance-related scores assigned to documents in component retrieval results. Others, such as Borda fusion (Aslam & Montague, 2001), Markov chain-based methods (Dwork et al., 2001; Renda & Straccia, 2003), Condorcet fusion (Montague & Aslam, 2002), and the multiple criteria approach (Farah & Vanderpooten, 2007) make use of the rank that each document occupies in each component result, as the scores are not always available.

Relevance-related scores obtained from different information retrieval systems may be diverse. Usually it is impossible to compare them directly and some kind of score normalization is required. Linear score normalization methods have been discussed in Lee (1997), Montague and Aslam (2001) and Wu et al. (2006), and non-linear score normalization methods have been discussed in Manmatha, Rath, and Feng (2001) and Nottelmann and Fuhr (2003). Since different retrieval systems use different ways to score documents, different score normalization methods may be required for different retrieval systems to obtain better effectiveness. However, the linear score normalization method has been widely used before in data fusion experiments.

CombSum, CombMNZ and some other methods were investigated by Fox et al. (1993) and Fox and Shaw (1994). They found that CombSum and CombMNZ outperformed the others. CombSum sets the score of each document in the fused result to the sum of the scores obtained by the component result, while in CombMNZ the score of each document is obtained by multiplying this sum by the number of results which have non-zero scores.

Lee (1997) conducted an experiment with six submitted results to TREC 3. He found that CombMNZ was slightly better than CombSum in his experiment. However, later experiments, for example, in Montague and Aslam (2001), Lillis et al. (2006), Wu, Crestani, and Bi (2006) and Wu and McClean (2006b) and others, found that Lee's conclusion is not always true and the probability that CombSum and CombMNZ are better than each other is roughly the same.