Available online at www.sciencedirect.com

ScienceDirect

Information Processing and Management 44 (2008) 838–855

INFORMATION PROCESSING & MANAGEMENT

www.elsevier.com/locate/infoproman

Query-level loss functions for information retrieval ☆

Tao Qin [a,*], Xu-Dong Zhang [a], Ming-Feng Tsai [c], De-Sheng Wang [a], Tie-Yan Liu [b], Hang Li [b]

[a] *Department of Electronic Engineering, Tsinghua University, Beijing, 100084, PR China*

[b] *Microsoft Research Asia, No. 49 Zhichun Road, Haidian District, Beijing 100080, PR China*

[c] *Department of Computer Science and Information Engineering, National Taiwan University, Taiwan 106, ROC*

Received 9 March 2007; received in revised form 2 July 2007; accepted 2 July 2007
Available online 27 September 2007

Abstract

Many machine learning technologies such as support vector machines, boosting, and neural networks have been applied to the ranking problem in information retrieval. However, since originally the methods were not developed for this task, their loss functions do not directly link to the criteria used in the evaluation of ranking. Specifically, the loss functions are defined on the level of documents or document pairs, in contrast to the fact that the evaluation criteria are defined on the level of queries. Therefore, minimizing the loss functions does not necessarily imply enhancing ranking performances. To solve this problem, we propose using query-level loss functions in learning of ranking functions. We discuss the basic properties that a query-level loss function should have and propose a query-level loss function based on the cosine similarity between a ranking list and the corresponding ground truth. We further design a coordinate descent algorithm, referred to as RankCosine, which utilizes the proposed loss function to create a generalized additive ranking model. We also discuss whether the loss functions of existing ranking algorithms can be extended to query-level. Experimental results on the datasets of TREC web track, OHSUMED, and a commercial web search engine show that with the use of the proposed query-level loss function we can significantly improve ranking accuracies. Furthermore, we found that it is difficult to extend the document-level loss functions to query-level loss functions.

Keywords: Information retrieval; Learning to rank; Query-level loss function; RankCosine

1. Introduction

Web search engines are changing people's life, and continuously enhancing the accuracy (relevance) of search also becomes an endless endeavor for information retrieval (IR) researchers. The key issue in web search

☆ This work was performed when the first and the fourth authors were interns at Microsoft Research Asia. The work was supported in part by the Joint Key Lab on Media and Network Technology set up by Microsoft and Chinese Ministry of Education in Tsinghua University.

* Corresponding author. Tel.: +86 10 6278 9944.

E-mail addresses: tsintao@gmail.com (T. Qin), zhangxd@tsinghua.edu.cn (X.-D. Zhang), mftsai@nlg.csie.ntu.edu.tw (M.-F. Tsai), wangdsh_ee@tsinghua.edu.cn (D.-S. Wang), tyliu@microsoft.com (T.-Y. Liu), hangli@microsoft.com (H. Li).

0306-4573/$ - see front matter
doi:10.1016/j.ipm.2007.07.016

is to construct a ranking function such that given a query the ranking function can rank the retrieved web pages in a way that can maximally satisfy users' search needs. Traditional approaches (Baeza-Yates & Ribeiro-Neto, 1999) resort to empirical methods in ranking model construction. These include content based methods such as BM25 (Robertson, 1997) and link based methods such as PageRank (Page, 1998). As more and more information (e.g., query log data) useful for search becomes available, the limitation of empirical tuning also becomes clearer, that is, it becomes very difficult, if not impossible, to tune the models with hundreds or thousands of features. The approach of employing machine learning techniques to address the problem naturally emerges as an effective solution and several methods have been proposed along the direction. Typical methods include RankBoost (Freund, Iyer, Schapire, & Singer, 2003), ranking SVM (Herbrich, Graepel, & Obermayer, 2000; Joachims, 2002), and RankNet (Burges et al., 2005), which are based on boosting, support vector machines and neural networks, respectively. From the machine learning perspective, ranking, in which given a query and its associated documents we are to rank the documents as correctly as possible, also becomes a new branch of supervised learning, in addition to classification, regression, and density estimation (Vapnik, 1998).

However, it should be noted that the aforementioned machine learning methods were not proposed directly for IR, and therefore their loss functions are only associated to some extent with the evaluation criteria in IR, such as mean average precision (MAP) (Baeza-Yates & Ribeiro-Neto, 1999), mean precision at n (P@n) (Baeza-Yates & Ribeiro-Neto, 1999), and normalized discounted cumulative gain (NDCG) (Jarvelin & Kekalainen, 2000, 2002). All the IR criteria are on the query-level; specifically, given two queries, no matter how different the numbers of documents retrieved for the two queries are, they contribute equally to the final performance evaluation. In contrast, the loss functions of the learning algorithms are defined on the level of documents (Nallapati, 2004) or document pairs (Burges et al., 2005; Freund et al., 2003; Herbrich et al., 2000; Joachims, 2002). Therefore, minimizing the loss functions does not necessarily lead to enhancing the accuracy in terms of the evaluation measures.

In order to solve this problem, we propose employing query-level loss functions in learning of ranking functions for IR.

In this paper, we first discuss what kind of properties a good query-level loss function should have. Then we propose a query-level loss function, cosine loss, as an example, which is based on the cosine similarity between a ranking list and the corresponding ground truth with respect to a given query. With the new loss function, we further derive a learning algorithm, RankCosine, which learns a generalized additive model as ranking function.

Next, we discuss whether it is possible to extend the document or document-pair level loss functions of the existing methods (ranking SVM, RankBoost, and RankNet) to the query-level.

We used two public datasets and one web search dataset to evaluate the effectiveness of our method. Experimental results show that the proposed query-level loss function is very effective for information retrieval. Furthermore, we found that it is in general difficult to extend the loss functions in the existing methods to the query-level.

The rest of this paper is organized as follows. In Section 2, we give a brief review on related work. In Section 3, we justify the necessity of using query-level loss functions for IR and discuss the properties that a good query-level loss function should have. We then give an example of query-level loss function, cosine loss, and derive an efficient algorithm to minimize the loss function in Section 4. Experimental results are reported in Section 5. In Section 6, we discuss the possibility of extending the loss functions of the existing methods to the query-level. Conclusions and future work are given in Section 7.

2. Related work

In recent years many machine learning technologies (Burges et al., 2005; Crammer & Singer, 2002; Dekel, Manning, & Singer, 2004; Freund et al., 2003; Herbrich et al., 2000; Joachims, 2002; Nallapati, 2004) were applied to the problem of ranking for information retrieval. Some early work simply tackled this problem as a binary classification problem (Nallapati, 2004), in which the assumption is made that a document is either relevant or irrelevant to the query, and the goal of learning is to classify relevant documents from irrelevant documents. However, in real-world applications, the degree of relevance of a document to a query can be discretized to multiple levels. For example, we can consider the use of three categories: highly relevant, partially