International Journal of Innovative Computing, Information and Control
Volume **7**, Number **12**, December **2011**

ICIC International ©2011 ISSN 1349-4198
pp. 6739–6756

RANKOM: A QUERY-DEPENDENT RANKING SYSTEM FOR INFORMATION RETRIEVAL

Jung-Yi Jiang, Lian-Wang Lee and Shie-Jue Lee

Department of Electrical Engineering
National Sun Yat-Sen University
No. 70, Lienhai Rd., Kaohsiung 80424, Taiwan
{ jungyi; len }@water.ee.nsysu.edu.tw; leesj@mail.ee.nsysu.edu.tw

Received July 2010; revised November 2010

Abstract. *Ranking plays an important role in information retrieval, aiming to sort the documents retrieved for a given query in the descending order of relevance. Recently, many approaches based on the idea of "learning to rank" have been proposed for doing ranking. Most of them consider all the documents of the training queries to build a static, query-independent ranking model. In this paper, we propose an adaptive, query-dependent framework for learning to rank based on a distributional similarity measure for gauging the similarity between queries. For each training query, one individual ranking model is learned from its associated set of documents. When a new query is consulted, the individual trained models of those training queries most similar to the new query are obtained and combined into a joint model which is then used to rank the documents retrieved for the new query. Experimental results show that our proposed approach works very well compared with other methods.*

Keywords: Learning to rank, Query-dependent ranking model, Distributional similarity measure, Ranking SVM, NDCG performance measure

1. **Introduction.** Ranking plays an important role in almost all the activities related to information retrieval, such as collaborative filtering, expert finding, data mining, web mining and anti web spam, and has attracted a lot of attention [20]. Its major concern is to sort the documents retrieved for a given query in the descending order of relevance. Recently, "learning to rank" has been a popular endeavor for dealing with the ranking problem [4, 5, 21, 26, 27, 29, 31, 32, 34]. The idea of "learning to rank" is to learn how to do ranking from the given training data using machine learning techniques.

Many "learning to rank" approaches have been proposed. They can be basically divided into three categories. The first category is point-wise [2, 7, 9, 11, 21, 36], which transforms ranking to the problem of classification or regression on documents. Ailon et al. [2] describe an efficient reduction of the ranking problem to the binary classification problem. Cossock et al. [9] propose a regression approach for solving the subset ranking problem. Fan et al. [11] propose a genetic programming based method for the study of context-specific ranking function discovery with automatic term weighting. Wang et al. [30] propose an immune programming based ranking function discovery approach. The second category is pair-wise [4, 12, 17], in which the learning task is formalized as classifying document pairs into two disjoint sets. For instance, Freund et al. [12] perform the learning by a boosting method, referred to as Rankboost. Herbrich et al. [14] propose a learning algorithm for ranking based on support vector machines [6, 8, 18, 33, 37]. Joachims et al. [17, 24] apply the RankSVM machine to document retrieval, and derive document pairs for training from users's clickthrough data. Qin et al. [22] employ multiple hyperplanes to rank documents. Burges et al. [4] apply RankNet to large-scale web search. The