Neurocomputing 150 (2015) 99–105

Contents lists available at ScienceDirect

Neurocomputing

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

Group-enhanced ranking

Yuan Lin [a], Hongfei Lin [a], Kan Xu [a], Ajith Abraham [b,c], Hongbo Liu [a,b,c,*]

[a] School of Computer, Dalian University of Technology, Dalian 116023, China
[b] Machine Intelligence Research Labs, Scientific Network for Innovation and Research Excellence, Auburn, WA 98071, USA
[c] School of Information, Dalian Maritime University, Dalian 116026, China

ARTICLE INFO

Article history:
Received 11 October 2013
Received in revised form
30 December 2013
Accepted 24 March 2014
Available online 2 October 2014

Keywords:
Information retrieval
Learning to rank
Groups
Loss functions

ABSTRACT

An essential issue in document retrieval is ranking, which is used to rank documents by their relevancies to a given query. This paper presents a novel machine learning framework for ranking based on document groups. Multiple level labels represent the relevance of documents. The values of labels are used to quantify the relevance of the documents. According to a given query in the training set, the documents are divided into several groups based upon their relevance labels. The group with higher relevance labels is always ranked upon the ones with lower relevance labels. Further a preference strategy is introduced in the loss functions, which are sensitive to the group with higher relevance labels to enhance the group ranking method. Experimental results illustrate that the proposed approach is very effective, with a 14 percent improvement on TD2003 dataset evaluated by MAP.

1. Introduction

Ranking is the very important to design effective web search engines, since the ranking model directly influences the relevance of search results [1–3]. Many approaches were proposed to construct a model for ranking. There are the content-based approaches [4], such as the vector space model [5], BM25 [6] and the language model [7]. PageRank [8] and Hits [9] are famous link-based approaches. In the search environment, we usually have to confront a large amount of information. It becomes very difficult to tune the models with a great number of features [10,11]. Some new attempts are made by introducing machine learning methods to information retrieval to address the problem. Learning to rank [12–14] is an effective approach. However, its performance is dependent directly on the document samples and ranking loss function. In this paper, we present a novel group ranking framework, in which the loss is defined on the groups of documents with same relevant label. The documents with the same level label are categorized into one group, and the ranking task is reduced from ranking the multiple documents to ranking several groups. We further develop the loss functions by our preference strategy, which are sensitive to the group with higher relevance labels.

The rest of this paper is organized as follows. Related works are discussed in Section 2. In Section 3, we briefly discuss the loss functions for ranking. The proposed group ranking framework is presented in Section 4. Then we illustrate the experimental results and discussions in Section 5. Finally, we draw conclusions and point out the future works.

2. Related works

In learning to rank, there are mainly three methods: pointwise [15], pairwise [16] and listwise [17]. Pointwise samples single document using classification loss functions for ranking model [18]. Pairwise applies preference document pairs as training samples and also transform the ranking problem into classification [19]. Listwise defines its loss function to train the ranking model from dataset [20,21]. ListMLE and ListNet are two important kinds of listwise approaches. ListMLE [21] is a feature-based ranking algorithm that minimizes a probabilistic likelihood loss function. And its listwise samples are defined by the permutation probabilities in the Luce model [22]. ListNet [17] is a robust listwise approach based on cross entropy loss function. Listwise approach can achieve the better ranking accuracies than pairwise and pointwise approaches on most of datasets of Letor [23].

However, usually only top k positions of ranking play a key role in information retrieval [24]. Xia et al. [24] develop a top-k ranking framework through likelihood loss to improve the top-k ranking performance. The top-k ranking loss function is used to obtain the relevant documents on the top-k positions in the document list.

* Corresponding author at: Machine Intelligence Research Labs, Scientific Network for Innovation and Research Excellence, Auburn, WA 98071, USA.
E-mail addresses: yuanlin@dlut.edu.cn (Y. Lin), hflin@dlut.edu.cn (H. Lin), xukan@dlut.edu.cn (K. Xu), ajith.abraham@ieee.org (A. Abraham), lhb@dlut.edu.cn (H. Liu).

http://dx.doi.org/10.1016/j.neucom.2014.03.079
0925-2312/

The number of the relevance documents is less than 10. The performance of top-k framework is determined by the number of relevant documents. It is ideal to set the value of k equal to the number of relevant documents [25]. But it is very difficult to do so [26]. Inspired by precious researches, we make an attempt to deal with it through group–group pair samples and preference strategy.

3. Loss functions for ranking

The ranking is optimized by minimizing a certain loss function using the training data. Likelihood and cross entropy functions are widely used in learning to rank. Here, we discuss only the basic ideas that are relevant to the present work.

3.1. Likelihood loss function

ListMLE [21] defines its probabilistic listwise loss function as follows:

$$L(f; x^q, y^q) = \sum_{s=1}^{n-1}\left(-f(x^q_{y^q_s}) + \ln\left(\sum_{i=s}^{n}\exp(f(x^q_{y^q_i}))\right)\right) \quad (1)$$

where f is a ranking function, x^q is the document list to be ranked for query q, y^q is a randomly selected optimum permutation for query q, and n is the length of y^q. For any two documents x_i and x_j, x_i is ranked before x_j in y^q if $label(x_i) > label(x_j)$. y^q_i is denoted as the index of the object ranked at the i-th position in y^q. ListMLE is feasible to rank the documents in Letor dataset [23]. However, its ranking performance decreases as the scores of irrelevant documents increase. We will discuss this through some experimental results in Section 5.2.

3.2. Cross entropy loss for ranking

ListNet [17] introduces a probabilistic cross entropy loss function, as defined in the following equation:

$$L(f, y) = D(P(\pi|x; \psi_y) \,\|\, P(\pi|x; f(x))) \quad (2)$$

where D is cross entropy loss. $P(\pi|x;\psi_y)$ and $P(\pi|x; f(x))$ are Luce models based on permutation probabilities. The score vector of the ground truth permutation is produced by a mapping function $\psi_y() : R_d \to R$, which is used to transform the order in a permutation, i.e., if $m > n$, then $\psi_y(m) > \psi_y(n)$. In order to optimize the top-k ranking accuracy, the mapping function only influences the order of documents within the top-k positions of the ground truth permutation. It also assigns a small value ϵ to all the remaining positions. The value is smaller than the score of any object ranked at the top-k positions, i.e. $\psi_y(x_{y_1}), \psi_y(x_{y_2}), \ldots, \psi_y(x_{y_k}), \psi_y(x_{y_{k+1}}), \ldots, \psi_y(x_{y_n})$, which compose a non-increasing sequence. So the loss becomes sensitive to the top-k subgroup order [24]. But all of the documents are considered as non-relevance after k position, which decreases the ranking performance.

4. Group ranking framework

Given a query in training set, the documents can be divided into several groups in which the documents with same labels are gathered together. A document pair is constructed by two groups, i.e., a group of documents with higher level label and a group of documents with lower level label. In this section, we introduce the group–group pair sampling, loss functions. Then our preference strategy is presented and the algorithm is summarized.

4.1. Group–group pair sample

Each query $q^{(i)}$ is associated with a list of documents $D^{(i)} = \{D^{(i)}_1, D^{(i)}_2, \ldots, D^{(i)}_n\}$. $D^{(i)}_j$ denotes the group of documents with the same relevance judgement j. n is the number of relevance degree for the documents. Each list of documents $D^{(i)}$ is associated with a list of judgments (scores) $Y^{(i)} = \{Y^{(i)}_1, Y^{(i)}_2, \ldots, Y^{(i)}_n\}$ where $Y^{(i)}_j$ denotes the judgment on the group document $D^{(i)}_j$ with respect to query $q^{(i)}$. For example, the relevance degree of $D^{(i)}_2$ is 2 when $j=2$. For the query $q^{(i)}$ with the relevance degree $n=3$ (0, 1, 2), the training sample is constructed as $D^{(i)}_g = \{D^{(i)}_{2,1}, D^{(i)}_{2,0}, D^{(i)}_{1,0}\}$. The group sample $D^{(i)}_{2,1}$ includes all the documents in the group $D^{(i)}_1$ and $D^{(i)}_2$. There are two types of label dataset used in this paper, such as OHSUMED with the relevance degree (0, 1, 2) and TD2003 with the relevance degree (0, 1).

4.2. Group ranking with loss functions

Different from the top-k ranking framework with listwise sample, our group ranking framework constructs samples by group pairs with different labels. The true loss of group ranking is defined as follows:

$$l_r(f(x), y) = \begin{cases} 0 & \text{if } \hat{y}_i = y_i \text{ where } \hat{y} = f(x); \\ 1 & \text{otherwise.} \end{cases} \quad (3)$$

where $i \in \{1, \ldots, r\}$, and r is determined by the number of documents with the higher label in the group ranking samples. The expectation of group ranking loss can be re-written as follows:

$$L_g(f) = \int_{X\times Y} l_r(f(x), y)\, dP(x, y) \quad (4)$$

where X is the input space in while the elements are the group samples to be ranked, and Y is the output space in which the elements are permutations of groups. $P(x, y)$ is an unknown but fixed joint probability distribution of x and y. And the optimal ranking function with respect to the group ranking actual loss is

$$f(x) = \underset{G_r(j_1, j_2, \ldots, j_r) \in G_r}{\operatorname{argmax}} P(G_r(j_1, j_2, \ldots, j_r)|x) \quad (5)$$

where $G_r(j_1, j_2, \ldots, j_r)$ denotes a group sample in which all the permutations have the same top-r true loss, which is decided by the number of documents with higher relevance label. G_r denotes the collection of all top-r subgroups. In this paper, likelihood and cross entropy functions are adopted in our group–group samples.

4.2.1. Group ranking with likelihood loss

Considering likelihood loss function, the loss of the group sample is described as follows:

$$L_g(f; x^g, y^g) = \sum_{s=1}^{r}\left(-f(x^g_{y^g_s}) + \ln\left(\sum_{i=s}^{n}\exp(f(x^g_{y^g_i}))\right)\right) \quad (6)$$

where x^g is a group sample, y^g is a ranked list of x^g, in which the documents with higher label are ranked upon the lower label group. In the group–group loss function, r is equal to the number of documents in the group with higher label. n is the length of optimum ranked list. As illustrated in Eq. (6), the loss becomes greater when increasing the scores of irrelevant documents obtained by ranking function. Our group rank framework ignores the increasing scores of the irrelevant documents in the ranked list for the likelihood loss, since the loss only depends on the increasing scores of relevant documents. The bigger scores of the relevant documents are, the smaller the loss is. We refer to the group method based on likelihood loss as GroupMLE.