Learning to Rank for Biomedical Information Retrieval

Bo Xu[1], Hongfei Lin[1], Yuan Lin[2], Yunlong Ma[1], Liang Yang[1], Jian Wang[1], Zhihao Yang[1]
[1] School of Computer Science and Technology, Dalian University of Technology, Dalian, China 116024
[2] School of Public Administration and Law, Dalian University of Technology, Dalian, China 116024
xubo2011@mail.dlut.edu.cn

***Abstract*—Research articles in biomedicine domain have increased exponentially, which makes it more and more difficult for biologists to manually capture all the information they need. Information retrieval technologies can help to obtain the users' needed information automatically. However, it is a great challenge to apply these technologies to biomedicine domain directly because of some domain specific characteristics, such as the abundance of terminologies. To enhance the effectiveness of the biomedical information retrieval, we propose a novel framework based on the state-of-the-art information retrieval methods, called learning to rank, which has been proved effective to rank documents based on their relevance degree. In the framework, we attempt to tackle the problem of the abundance of terminologies by constructing ranking models, which focus on not only retrieving the most relevant documents but also diversifying the searching results to increase the completeness of the resulting list for a given query. In the model training, we propose two novel document labeling strategies, and combine several traditional retrieval models as learning features. Besides, we also investigate the usefulness of different learning to rank approaches in our framework. Experimental results on TREC Genomics datasets demonstrate our proposed framework is effective in improving the performance of biomedical information retrieval.**

***Keywords*—learning to rank; biomedical information retrieval; diversity**

I. Introduction

In recent years, research articles in biomedicine domain have increased exponentially, which makes it difficult for biologists to manually capture all the information they need. To meet biologists' information need better, information retrieval (IR) techniques designed for biomedicine domain have been addressed, focusing on how to effectively retrieve the needed information. Given a query, an IR system can search for its relevant documents, and rank the documents based on their relevance degrees to the query. Unlike traditional IR, biomedical IR faces some domain specific challenges, most of which are due to the abundance of the terminologies. To meet the information need more completely, biomedical IR system should cover the relevance documents from different aspects, where an aspect of relevance documents refers to a subset of relevant documents related to the same terminologies.

Therefore, biomedical retrieval systems not only focus on obtaining the most relevant documents to a given query, but also emphasize the query-related aspects coverage in the document ranked list, which is mostly denoted as the diversity of the searching result.

In recent years, various traditional IR models have been introduced for the biomedical document ranking, and achieve some good results. Learning to rank, as a state-of-the-art IR technique, has been proved effective in many IR tasks, and many learning to rank methods have been proposed [1-8]. However, few studies attempted to employ learning to rank methods to improve the diversity oriented biomedical information retrieval. Learning to rank methods have some advantages over other traditional IR models. For one thing, it can make the most of various ranking information comprehensively to construct a ranking model. For another, the training phase for learning to rank methods iteratively reduce the value of ranking loss (i.e., difference between the predicted ranking and the ground truth ranking), until eventually an optimal ranking model is obtained. Therefore, it seems promising to improve biomedical retrieval using learning to rank methods.

In the paper, we propose a novel framework based on learning to rank methods to study whether learning to rank methods would benefit biomedical retrieval and boost both the relevance and the diversity of results. In the framework, we propose two novel labeling strategies to capture the aspects information of the relevant documents, thus forming the ground truth document ranking. Meanwhile, documents are scored by various traditional IR models, and represented as feature vectors using the scores. Then, we construct an effective ranking model using these feature vectors as training data to improve retrieval performance. Finally, for a new query, we predict its corresponding document ranking using the trained model.

II. Related Work

In biomedical information retrieval, ranking only based on document relevance is not sufficient to meet the information need, because relevant documents may be redundant with each other. Aspect retrieval was proposed to reduce the redundancy and improve result diversity in the Genomics track of Text Retrieval Conference (TREC) [9-10]. In the 2006 TREC Genomics track, University of Wisconsin at Madison proposed a clustering approach, but failed to promote diversity by penalizing redundancy [11]. In the 2007 TREC Genomics track, most submissions are purely based on relevance passage

978-1-4673-6799-8/15/$31.00 ©2015 IEEE

retrieval such as National Library of Medicine (NLM) [12]. Thereafter, some researches focused on modeling the diversity by detecting query-related potential aspects. Yin et al. [13] utilized Wikipedia to detect aspects, and proposed a cost based document re-ranking method to balance the relevance and the diversity of retrieval performance. Based on Wikipedia aspect detection, a survival modeling method was introduced to model the passage diversity [14], and a relevance-novelty model, RelNov, was proposed to improve passage retrieval [15]. In [16], a topic modeling method based on Latent Dirichlet Allocation (LDA) is proposed to measure the novelty of a given passage. In [17], a retrieval model based on Probabilistic Latent Semantic Analysis (PLSA) was proposed to detect latent aspects for diversity retrieval. In [18], Wu et al. directly apply learning to rank methods for medical document retrieval, and achieve good performance. Therefore, we believe that the retrieval performance can be further enhanced by optimizing these methods for biomedicine domain.

In IR, learning to rank, as a powerful technology, has been proved effective in improving relevant-based retrieval performance in the intersection of machine learning and information retrieval [6, 19]. In order to solve ranking problem, many learning to rank methods have been proposed to improve ranking accuracies [1-8]. In particular, learning to rank is grouped into three approaches: the pointwise approach, the pairwise approach and the listwise approach. Different approaches model the learning to rank process in different ways. Besides, Lin et al. [20] proposed group-wise learning to rank framework, and demonstrated its effectiveness. In the paper, we attempt to optimize the learning to rank methods for biomedical information retrieval, and investigate the effectiveness of the learning to rank methods in different approaches.

III. METHOD

A. General Learning Framework

In this section, we will formalize our learning to rank based framework for document retrieval. At the training time, we are given a set of N queries $Q=\{q_1, q_2, \ldots, q_N\}$. To simplify notation, we drop the query index, and refer to a general query q. Each query q is associated with a set of M documents $D=\{d_1, d_2, \ldots, d_M\}$. The documents are manually labeled with relevance labels, denoted as $L=\{l_1, l_2, \ldots, l_M\}$. For each document d_j, label l_j is a integer value indicating the relevant degree of the document d_j to the query q. In addition, each document d_j is represented as a query dependent feature vector, where $f_j[k]$ denotes the k^{th} feature value for the document d_j. The learning goal is to create a scoring function F such that, given a set of documents D with relevance labels L for a query q, the ranking of documents in D produced by F has maximal agreement with L. Then, the scoring function, as the ranking model, is used to rank documents for new queries.

B. Biomedical Document Labeling

In the training phase, learning to rank methods can reduce the ranking loss by measuring the difference between the outputs and the ground truths. The ground truths refer to the relevance labels of the documents, and can be considered as the learning target to train a ranking model.

In biomedical IR, relevant documents are not only judged with relevance labels, but also explicitly annotated with some biomedical terms, and each term stands for one aspect to the query. Therefore, our task is how to generate effective document labels involving both the original relevance label and the aspect information. Based on the idea, we propose two novel labeling strategies to tackle the problem.

1) Optimal Ranking Labeling Strategy.

Firstly, we attempt to construct an optimal ranking list of relevant documents in consideration of their diversity degrees. In the optimal ranking list, more diversified relevant documents are ranked higher than less diversified ones. At the training time, learning to rank methods compute the ranking loss by measuring the difference between the optimal ranking list and the ranking list outputted by the model, and then iteratively adjust the model to reduce the loss continuously. Our first labeling strategy is based on this idea by taking the number of aspects for one document and the frequency of aspects among all the documents into account, where the aspects for a relevant document reflect its diversity degree. The algorithm is shown in Table I.

TABLE I. OPTIMAL RANKING LABELING STRATEGY

```
Input: relevant document set $R=\{r_1, r_2, \ldots, r_n\}$, aspect set $Asp_{ri}$ for
        document $r_i$, the whole aspect set $Asp$
Output: document labels $L=\{l_1, l_2, \ldots, l_n\}$
1  initialize $label=|R|$
2  find documents $S$ with maximum aspects in $R$
3  for each document $r_i$ in $S$, compute $\sum_{aspect_j \in Asp_{r_i}} df(aspect_j)$
4  choose the document $r_k$ with the minimum $\sum_{aspect_j \in Asp_{r_i}} df(aspect_j)$
5  $l_k=label$
6  $label=label-1$
7  update $Asp$ by removing the aspects in $r_k$
8  update $R$ by removing the document $r_k$
Repeat step 2 to step 8 until $Asp$ is empty
9  for the remaining documents in $R$, choose the document $r_k$ with
   minimum $\sum_{aspect_j \in Asp_{r_i}} df(aspect_j)$
10 $l_k=label$
11 $label=label-1$
12 update $R$ by removing the document $r_k$
Repeat step 9 to 11 until $R$ is empty
Return $L$
```

In this algorithm, $df(aspect_j)$ counts the number of relevant documents covering the j^{th} aspect. For a given query, there are a set of relevant documents $R=\{r_1, r_2, \ldots, r_n\}$. One relevant document may cover several query related aspects, and one aspect may be shared by many relevant documents. Therefore, we take these two factors into account to form the ground truth labels of the documents. The final ground truth labels for relevant documents are integers ranging from 1, 2 to *n*, indicating the diversity degrees of these documents from low to high, where *n* represents the total number of relevant documents. Besides, irrelevant documents are labeled as 0.

2) *Group-wise Labeling Strategy*

Optimal ranking strategy provides the target ranking to train the rank models, which may be more suitable for listwise learning to rank approach, because it directly measures the difference between the target ranking list and the output ranking list. However, for pointwise and pairwise learning to rank methods, it may not work well, because they respectively utilize the exact relevance degree of each document and preferences between two documents to compute the ranking loss. Based on this consideration, we propose another labeling strategy to examine learning to rank methods.

Inspired by the group-wise learning to rank framework proposed in [20], we propose a diversity-oriented group-wise learning to rank framework to improve the retrieval diversity. In this framework, documents with different labels are treated as a group, and the ranking task is then reduced from ranking the whole set of documents to ranking a group of documents with different labels. We modify the group-wise learning to rank framework to make it fit into the biomedical diversity-oriented retrieval. This algorithm is presented in Table II.

TABLE II. GROUP-WISE LABELING STRATEGY

```
Input: relevant document set $R=\{r_1, r_2, \ldots, r_n\}$, aspect set $Asp_{ri}$ for
        document $r_i$
Output: document labels $L=\{l_1, l_2, \ldots, l_n\}$,
        document groups $G=\{g_1, g_2, \ldots, g_n\}$
1  find documents $S$ with maximum aspects in $R$
2  for each document $r_i$ in $S$, compute $\sum_{aspect_j \in Asp_{r_i}} df(aspect_j)$
3  choose the document $r_k$ with the minimum $\sum_{aspect_j \in Asp_{r_i}} df(aspect_j)$
4  $l_k=1$, $g_k=k$
5  update $R$ by removing the document $r_k$
6  for each document $r_i$ in $R$
7      if $Asp_{ri} = Asp_{rk}$
8          $l_k=1$, $g_k=k$, update $R$ by removing the document $r_k$
9      if $Asp_{ri} \subset Asp_{rk}$
10         $l_k=0$, $g_k=k$, update $R$ by removing the document $r_k$
11 end for
Repeat step 1 to step 11 until $R$ is empty
Return $L$ and $G$
```

In the algorithm, we firstly divide the relevant documents into groups based on their covered aspects. Each group contains one document with more aspects (label 1) and several documents with less aspects (label 0), and the document with more aspects covers all the aspects in the documents with less aspects. Besides, the documents with the same set of aspects are assigned into the same group with the same labels. After dividing the relevant documents into groups, we assign each group some irrelevant documents. As a result, one or more relevant documents and a group of irrelevant documents constitute the whole of one group, which can be taken as a learning unit at the model training time. Since the division of groups is based on the diversity degrees of the documents, the group-wise framework can be more focused on the diversified documents, and the final ranking model may improve the performance in terms of both relevance and diversity.

C. *Ranking Features*

1) *Features based on Vector Space Model.*

Vector space model (VSM) [21] has been widely used in information retrieval field, and it calculates the cosine similarity between a document *d* and a query *q* as follows.

$$\text{cosine}(d,q) = \frac{\sum_{j \in q} w_d(j) \cdot w_q(j)}{\sqrt{\sum_{j \in q} w_d^2(j) \cdot \sum_{j \in q} w_q^2(j)}} \tag{1}$$

$$\text{tf}(j,d) = \frac{occurrence_d(j)}{|d|+1.0} \tag{2}$$

$$\text{idf}(j) = \log \frac{N - n(j) + 0.5}{n(j) + 0.5} \tag{3}$$

$$w_d(j) = \text{tf}(j,d) \cdot \text{idf}(j) \tag{4}$$

2) *Features Based on BM25 Model.*

Okapi BM25 model [22] takes into account the document length to overcome the shortcoming of vector space model (VSM) as follows.

$$\text{BM25}(d,q) = \sum_{j \in q} \text{idf}(j) \cdot \frac{(k_3 + 1.0) \cdot \text{tf}(j,q)}{k_3 + \text{tf}(j,q)} \cdot \frac{\text{tf}(j,d) \cdot (k_1 + 1.0)}{\text{tf}(j,d) + k_1 \cdot (1 - b + b \cdot |d| / avgdl)} \tag{5}$$

3) *Features Based on Language models.*

The unigram language model (LM) [23] is often used in traditional IR, and different smoothing methods are adopted to optimize the ranking model. The language model with Jelinek-Mercer smoothing can be calculated as follows:

$$w_d(j) = (1-\lambda) \cdot \text{tf}(j,d) + \lambda \cdot \text{tf}(j,C) \tag{6}$$

Language model with Bayesian smoothing can be calculated as follows.

$$w_d(j) = \frac{\text{tf}(j,d) + \mu \cdot \text{tf}(j,C)}{\sum_i \text{tf}(i,d) + \mu} \tag{7}$$

D. *Learning Methods*

In this paper, we examine the usefulness of our framework by extending three learning to rank approaches: the pointwise approach, the pairwise approach and the listwise approach.

For the pointwise approach, the ranking loss is computed based on the difference between the score obtained from the trained model and its ground truth label. Take Regression [24] as an example, and its loss function is as follows.

$$\mathrm{loss}(f(x_i), y_i) = \sum_i (f(x_i) - y_i)^2 \tag{8}$$

For pairwise approach, the loss function is computed based on the preferences in each document pair. For example, RankBoost [4] combines preferences based on the boosting approach to machine learning, and its loss function is as follows.

$$\mathrm{loss}(f(x_i), f(x_j), y_{i,j}) = \sum_{i,j} \exp(-y_{i,j} \bullet (f(x_i) - f(x_j))) \tag{9}$$

For listwise approach, the loss function is measured in terms of the difference between the target ranking list and the output ranking list of documents [25]. For example, LambdaMART [1], as a listwise method, is the boosted tree version of listwise LambdaRank [2], which is based on RankNet [26]. Its ranking loss can be accumulated with the loss gradient replacement λ as follows.

$$\lambda_i = \sum_{j:(i,j)\in I} \lambda_{i,j} - \sum_{j:(j,i)\in I} \lambda_{i,j} \tag{10}$$

The loss function of LambdaMART has the same form as RankNet with a particular gradient replacement λ [27].

IV. EXPERIMENTS AND ANALYSIS

A. Experimental Settings

We examine our learning framework on TREC Genomics track 2006 & 2007 datasets. The dataset consists of 162,259 documents from 49 genomics-related journals. These documents are divided into more than 10 million passages based on the pre-defined passage legal spans [9]. There are totally 62 queries, 26 queries of which are from 2006's track (we remove two queries with no relevant documents in advance) and 36 queries are from 2007's track.

We perform 5 fold cross validation to examine the performance. Specifically, queries from 2006 and 2007 TREC Genomics tracks are respectively divided by the query number into training set, validation set and test set, where 60% queries are used for training, 20% for validation and 20% for testing. The reporting results are averaged over all the folds. The retrieval units for the datasets are passages, so we will replace the phrase "document retrieval" with the phrase "passage retrieval" in our experiments, but in practice, they are the same.

We take the evaluation measures used in TREC Genomics Track, Document MAP, Aspect MAP, Passage MAP and Passage2 MAP, to examine the retrieval performance [9-10]. These variations of Mean Average Precision (MAP) can help measure both the diversity and the relevance of retrieved passages.

TABLE III. RETRIEVAL PERFORMANCE OF RANKING MODELS ON THE TREC 2006 GENOMICS COLLECTION

MAP	Document	Passage	Aspect	Passage2
Okapi	0.3466	0.0282	0.2362	0.0325
Survival	0.3523	0.0290	0.2450	0.0331
LTR	0.3490	0.0291	0.2351	0.0312
Opi_Rank	0.3522	0.0300	0.2443*	0.0349*
	(+1.62%)	(+6.28%)	(+3.43%)	(+7.41%)
Group	0.3780*	0.0450*	0.2494*	0.0619*
	(+9.07%)	(+59.45%)	(+5.58%)	(+90.27%)

TABLE IV. RETRIEVAL PERFORMANCE OF RANKING MODELS ON THE TREC 2007 GENOMICS COLLECTION

MAP	Document	Passage	Aspect	Passage2
Okapi	0.2562	0.0659	0.1948	0.0800
Survival	0.2654	0.0720	0.2022	0.0853
LTR	0.2579	0.0707	0.1947	0.0812
Opi_Rank	0.2640	0.0715*	0.2138*	0.0821
	(+3.06%)	(+7.83%)	(+9.77%)	(+2.61%)
Group	0.3555*	0.1125*	0.2822*	0.1226*
	(+38.77%)	(+57.33%)	(+44.90%)	(+53.30%)
NLMinter	0.3286	0.0968	0.2631	0.1148
Survival	0.3243	0.0969	0.2695	0.1183
LTR	0.3270	0.0953	0.2644	0.1135
Opi_Rank	0.3309	0.0972*	0.2638*	0.1152
	(+0.69%)	(+0.37%)	(+0.27%)	(+0.35%)
Group	0.4264*	0.1211*	0.2896*	0.1425*
	(+29.75%)	(+25.07%)	(+10.09%)	(+24.12%)
MuMSHfd	0.2906	0.0840	0.2068	0.0895
Survival	0.2844	0.0844	0.2256	0.0918
LTR	0.2903	0.0791	0.2097	0.0860
Opi_Rank	0.2941	0.0975*	0.2152*	0.0988*
	(+1.22%)	(+16.13%)	(+4.04%)	(+10.34%)
Group	0.2991*	0.0977	0.2220*	0.1033*
	(+2.94%)	(+16.28%)	(+7.32%)	(+15.38%)

B. Ranking model compared

To compare the retrieval effectiveness of our proposed framework, we evaluate the following ranking models in our experiments; there are totally 5 kinds of ranking models.

(a) The ranking model obtained from original submission run. For 2007 queries, we select two high-performance official submission runs and an Okapi run. The two official runs are NLMinter [12] and MuMshFd [28], and Okapi run is solely based on the probabilistic weighting model BM25. For 2006 queries, we only select the Okapi run as our baseline, because other official submissions are not available.

(b) The ranking model obtained using survival modeling approach in [14], which models aspects using survival analysis to promote the ranking diversity, which can be considered as a strong baseline.

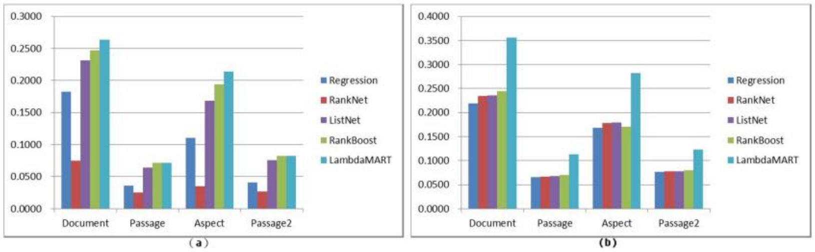

Fig. 1. Performance of different learning to rank methods based on the 2007's Okapi baseline. (a) The optimal ranking strategy. (b) The group-wise strategy.

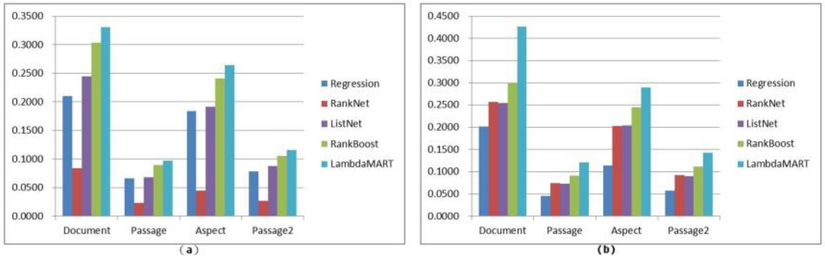

Fig. 2. Performance of different learning to rank methods based on the 2007's NLMinter baseline. (a) The optimal ranking strategy. (b) The group-wise strategy.

(c) The ranking model obtained from traditional learning to rank methods with all the defined features and binary relevance labels, denoted as traditional LTR models.

(d) The ranking model obtained by optimal labeling strategy and the group-wise labeling strategy with all the defined features.

Besides, we compare three learning to rank approaches, namely the pointwise approach, the pairwise approach and the listwise approach in our study.

C. Comparisons on Retrieval Performance

In this section, we evaluate our methods based on the learning to rank method LambdaMART in comparison with all baseline runs, and show their performance in Table III and Table IV, where Survival refers to the method in [14], LTR refers to the original learning to rank method, the Opi_Rank represents the method based on optimal ranking labeling strategy, and Group represents the methods based on group-wise learning to rank. The values in parentheses are the relative rates of improvement over the original results. Besides, we compare the results using statistical test (i.e., two-tailed paired Student's t tests), where '*' indicates that improvement of term ranking over original run is significant with 95% confidential level ($p<0.05$).

From the tables, we can see that our methods achieve consistent improvement over all the baseline runs in terms of all levels of MAP evaluation measures. In comparison, the results based on optimal ranking labeling strategy outperforms most of the baseline measures, and the group-wise learning to rank framework achieves better results, and improve the passage retrieval performance further.

D. Performance of Different Learning to Rank Approaches

In this section, we compare the effectiveness of our framework among five state-of-the-art learning to rank methods based on four baseline runs mentioned above. These methods belongs to three learning to rank approaches, which are Regression [24] (pointwise), RankNet [26] (pairwise) and RankBoost [4] (pairwise), ListNet [29] (listwise) and LambdaMART [1] (listwise). The comparisons of results on two standard submission runs are shown in Fig. 1 and Fig. 2.

From the figures, we find that, compared with all the other methods, LambdaMART performs the best on the baseline runs in terms of most of the evaluation measures, and the performance of RankNet varies a lot on the two strategies. For optimal ranking strategy, RankNet does not perform very well, but for the group-wise strategy, its performance is almost between Regression and other methods.

E. Discussion

In this section, we will further discuss and analyze our experimental results to find the advantages and disadvantages of our methods.

The optimal ranking labeling strategy can set a learning target for learning to rank algorithms to tune the model, and it seems effective to improve the original results. Meanwhile, the learning target may focus too much on the most diversified passages, so its performance is less significant on all the evaluation measures. In comparison, group-wise learning to rank can better meet the requirements for diversity-oriented retrieval by taking groups as a training unit, and each group consists of one or more diversified passage, some less diversified passages and a group of irrelevant passages. Based on this idea, learning to rank algorithm can be focused on the passages with more aspects, and tend to choose different aspects in various ways, resulting in more effective ranking models. Therefore, the ranking models can contribute more to the performance in terms of both relevance and diversity.

Besides, from the algorithms in Table I and Table II, we can also find that time complexity of group-wise learning to rank framework is much lower than the optimal ranking one. Above all, we believe that group-wise learning to rank framework is more effective than the optimal ranking framework for biomedical document retrieval to improve the performance in terms of relevance and diversity.

V. CONCLUSIONS AND FUTURE WORK

In this paper, we propose a learning to rank based framework for biomedical information retrieval. The proposed methods are respectively based on optimal ranking strategy and the group-wise learning to rank, and we investigate the effectiveness of our methods on various learning to rank methods belonging to three approaches. Experimental results on TREC Genomics track datasets demonstrate our proposed framework is effective in improving the performance of biomedical retrieval. Learning to rank method, LambdaMART, outperforms other methods in our framework for biomedical retrieval. The optimal ranking strategy and the group-wise strategy can both contribute to the performance, and group-wise learning to rank can improve the performance better.

We will extend our future work in some directions. Since our proposed method needs explicit aspect annotations to train a ranking model, we will attempt to explore an approach for automatic aspect mining when the dataset contains no aspect annotations, and we will also develop and examine the performance of other features, especially some domain specific features, to make the framework more applicable for biomedical document retrieval.

ACKNOWLEDGMENT

This work is partially supported by grant from the Natural Science Foundation of China (No. 61272370, 61402075, 61572102, 61572098, 61272373), Natural Science Foundation of Liaoning Province, China (No.201202031, 2014020003), State Education Ministry and The Research Fund for the Doctoral Program of Higher Education (No.20090041110002), the Fundamental Research Funds for the Central Universities.

REFERENCES

[1] Burges, C. J. (2010). From ranknet to lambdarank to lambdamart: An overview. Learning, 11, 23-581.

[2] Burges, C.J., Ragno, R., Le, Q.V. (2006). Learning to rank with nonsmooth cost functions. In: NIPS. pp. 193-200.

[3] Cao, Y., Xu, J., Liu, T. Y., Li, H., Huang, Y., & Hon, H. W. (2006). Adapting ranking SVM to document retrieval. In Proceedings of the 29th annual international ACM SIGIR conference on research and development in information retrieval (pp. 186-193). ACM.

[4] Freund, Y., Iyer, R., Schapire, R. E., & Singer, Y. (2003). An efficient boosting algorithm for combining preferences. The Journal of machine learning research, 4, 933-969.

[5] Joachims, T. (2002). Optimizing search engines using clickthrough data. In Proceedings of the eighth ACM SIGKDD international conference on knowledge discovery and data mining (pp. 133-142). ACM.

[6] Liu, T. Y. (2009). Learning to rank for information retrieval. Foundations and Trends in Information Retrieval, 3(3), 225-331.

[7] Wu, Q., Burges, C. J., Svore, K. M., & Gao, J. (2008). Ranking, boosting, and model adaptation. Tecnical Report, MSR-TR-2008-109.

[8] Xu, J., & Li, H. (2007). Adarank: a boosting algorithm for information retrieval. In Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval (pp. 391-398). ACM.

[9] Hersh, W. R., Cohen, A. M., Roberts, P. M., & Rekapalli, H. K. (2006). TREC 2006 genomics track overview. In TREC.

[10] Hersh, W.R., & Voorhees, E. (2009). TREC genomics special issue overview. Information Retrieval, 12(1), 1-15.

[11] Goldbery, A., Gael, D. A. J., Settles, B., Zhu, X., & Craven, M. (2006). Ranking biomedical passages for relevance and diversity. University of Wisconsin, Madison at TREC Genomics 2006; Proc of TREC, 15.

[12] Demner-Fushman, D., Humphrey, S. M., Ide, N. C., Loane, R. F., Mork, J. G., Ruch, P., et al. (2007). Combining Resources to Find Answers to Biomedical Questions. In TREC.

[13] Yin, X., Huang, X., & Li, Z. (2010). Promoting ranking diversity for biomedical information retrieval using wikipedia. In Advances in Information Retrieval (pp. 495-507). Springer Berlin Heidelberg.

[14] Yin, X., Huang, J. X., Li, Z., & Zhou, X. (2013). A survival modeling approach to biomedical search result diversification using Wikipedia. Knowledge and Data Engineering, IEEE Transactions on, 25(6), 1201-1212.

[15] Yin, X., Li, Z., Huang, J. X., & Hu, X. (2010). A relevance-novelty combined model for genomics search result diversification. In Bioinformatics and Biomedicine (BIBM), 2010 IEEE International Conference on (pp. 692-695). IEEE.

[16] Chen, Y., Yin, X., Li, Z., Hu, X., & Huang, J. X. (2011). Promoting Ranking Diversity for Biomedical Information Retrieval based on LDA. In Bioinformatics and Biomedicine (BIBM), 2011 IEEE International Conference on (pp. 456-461). IEEE.

[17] An, X., & Huang, J. X. (2013). Boosting novelty for biomedical information retrieval through probabilistic latent semantic analysis. In Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval (pp. 829-832). ACM.

[18] Wu, J., & Huang, J. (2014). York University at CLEF eHealth 2014: A Learning-to-Rank Approach for Medical Document Retrieval. Proceedings of the ShARe/CLEF eHealth Evaluation Lab.

[19] Qin, T., Liu, T. Y., Xu, J., & Li, H. (2010). LETOR: A benchmark collection for research on learning to rank for information retrieval. Information Retrieval, 13(4), 346-374.

[20] Lin, Y., Lin, H., Ye, Z., Jin, S., & Sun, X. (2010). Learning to rank with groups. In Proceedings of the 19th ACM international conference on Information and knowledge management (pp. 1589-1592). ACM.

[21] Baeza-Yates, R., & Ribeiro-Neto, B. (1999). Modern information retrieval (Vol. 463). New York: ACM press.

[22] Robertson, S. E., Walker, S., Jones, S., Hancock-Beaulieu, M. M., & Gatford, M. (1995). Okapi at TREC-3. NIST SPECIAL PUBLICATION SP, 109-109.

[23] Zhai, C., & Lafferty, J. (2001). A study of smoothing methods for language models applied to ad hoc information retrieval. In Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval (pp. 334-342). ACM.

[24] Cossock, D., & Zhang, T. (2006). Subset ranking using regression. In Learning theory (pp. 605-619). Springer Berlin Heidelberg.

[25] Xia, F., Liu, T. Y., Wang, J., Zhang, W., & Li, H. (2008). Listwise approach to learning to rank: theory and algorithm. In Proceedings of the 25th international conference on Machine learning (pp. 1192-1199). ACM.

[26] Burges, C.J., Shaked, T., Renshaw, E., Lazier, A., Deeds, M., Hamilton, N., et al. (2005). Learning to rank using gradient descent. In Proceedings of the 22nd international conference on Machine learning (pp. 89-96). ACM.

[27] Ganjisaffar, Y., Caruana, R., & Lopes, C. V. (2011). Bagging gradient-boosted trees for high precision, low variance ranking models. In Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval (pp. 85-94). ACM.

[28] Stokes, N., Li, Y., Cavedon, L., Huang, E., Rong, J., & Zobel, J. (2007). Entity-Based Relevance Feedback for Genomic List Answer Retrieval. In TREC.

[29] Cao, Z., Qin, T., Liu, T. Y., Tsai, M. F., & Li, H. (2007). Learning to rank: from pairwise approach to listwise approach. In Proceedings of the 24th international conference on Machine learning (pp. 129-136). ACM.