Global Ranking via Data Fusion

Hong-Jie Dai[1,2] **Po-Ting Lai**[3] **Richard Tzong-Han Tsai**[3*] **Wen-Lian Hsu**[1,2*]

[1]Department of Computer Science, National Tsing-Hua University,
[2]Institute of Information Science, Academia Sinica,
[3]Department of Computer Science & Engineering, Yuan Ze University
hongjie@iis.sinica.edu.tw
s951416@mail.yzu.edu.tw
thtsai@saturn.yzu.edu.tw
hsu@iis.sinica.edu.tw

Abstract

Global ranking, a new information retrieval (IR) technology, uses a ranking model for cases in which there exist relationships between the objects to be ranked. In the ranking task, the ranking model is defined as a function of the properties of the objects as well as the relations between the objects. Existing global ranking approaches address the problem by "learning to rank". In this paper, we propose a global ranking framework that solves the problem via data fusion. The idea is to take each retrieved document as a pseudo-IR system. Each document generates a pseudo-ranked list by a global function. The data fusion algorithm is then adapted to generate the final ranked list. Taking a biomedical information extraction task, namely, interactor normalization task (INT), as an example, we explain how the problem can be formulated as a global ranking problem, and demonstrate how the proposed fusion-based framework outperforms baseline methods. By using the proposed framework, we improve the performance of the top 1 INT system by 3.2% using the official evaluation metric of the BioCreAtIvE challenge. In addition, by employing the standard ranking quality measure, NDCG, we demonstrate that the proposed framework can be cascaded with different local ranking models and improve their ranking results.

1 Introduction

Information Retrieval (IR) involves finding documents that are relevant to a given query in a large corpus. The task is usually formulated as a ranking problem. When a user submits a query, the IR system retrieves all documents that contain at least one query term, calculates a ranking score for each of the documents using a ranking model, and sorts the documents according to the ranking scores. The scores represent the relevance, importance, and/or diversity of the retrieved documents. Thus, the quality of a search engine can be determined by the accuracy of the ranking results.

Recently, a machine learning technology called *learning to rank* has been applied extensively to the task. Several state-of-the-art machine learning-based ranking algorithms have been proposed, e.g., RankSVM and RankNet. These algorithms differ substantially in terms of the ranking models and optimization techniques employed, but most of them can be regarded as "local ranking" approaches in the sense that each model is defined on a single document without considering the possible relations to other documents to be ranked. In many applications, this is only a loose approximation as there is always relational information among documents. For example, in some cases, users may prefer that two similar documents have similar relevance scores; even

[*] Corresponding author

though one of the documents is not as relevant to the given query as the other; this problem is similar to Pseudo Relevance Feedback (Kwok, 1984). In other cases, web pages from the same site form a sitemap hierarchy in which a parent document should be ranked higher than its child documents (referred to as Topic Distillation at TREC (Chowdhury, 2007)). To utilize all available information, more advanced ranking algorithms define a ranking model as a function of all the documents to be ranked, i.e., a global ranking model (Qin et al., 2008a; Qin et al., 2008b).

Unlike conventional ranking and learning to rank models, such as BM25 and RankSVM, whose ranking functions are defined on a query and document pair, global ranking models utilize both content information and relation information. Qin et al. (2008) proposed the first supervised learning framework for the global ranking problem. They formulated the problem as an optimization problem that involves finding an objective function to minimize the trade-off between local consistence and global consistence and implemented it on SVM. Subsequently, they defined the global ranking problem formally in (Qin et al., 2008) and solved it by employing continuous conditional random fields (CRF).

In this paper, we propose a new framework for the global ranking problem. The major difference between our work and that of Qin et al. (2008a; 2008b) is that we do not compile a feature vector of relational information directly to construct a new machine-learned ranking model for global ranking. Instead, we use the ranking results generated by the original ranking model and then employ an algorithm with the relational information to transform the global ranking problem into a data fusion problem; that is also known as a rank aggregate problem. The proposed framework is flexible and can be cascaded with conventional ranking models or learning to rank models.

The remainder of this paper is organized as follows. In Section 2, we present a formal definition of global ranking. In Section 3, we describe the proposed framework and consider three fusion algorithms that can be used with our framework. We also explain how the algorithms can be adapted to solve the global ranking problem. In Section 4, we introduce a biomedical text mining task called the interactor normalization task (INT) (Krallinger et al., 2009) and show why it should be formulated as a global ranking problem. In Section 5, we report extensive experiments conducted on the INT dataset released by BioCreAtIvE (Krallinger et al., 2009). Section 6 contains some concluding remarks.

2 Global Ranking Problem

The global ranking problem was first defined formally by Qin et al. (2008). In this paper, we propose a new global ranking framework based on their definition. Although we developed the framework independently, we adopt Qin et al.'s terminology.

Let q denote a query. In addition, let $x^{(q)} = \left\{x_1^{(q)}, x_2^{(q)}, \dots, x_{n^{(q)}}^{(q)}\right\}$ denote the documents retrieved by q , and let $y^{(q)} = \left\{y_1^{(q)}, y_2^{(q)}, \dots, y_{n^{(q)}}^{(q)}\right\}$ denote the ranking scores assigned to the documents. Here, $n^{(q)}$ represents the number of documents retrieved by q. Note that the numbers of documents varies according to different queries. We assume that $y^{(q)}$ is determined by a ranking model.

If a ranking model is defined on a single document, i.e., in the form of

$$y_i^{(q)} = f\left(x_i^{(q)}\right), \mathrm{i} = 1, \dots, n^{(q)}, \qquad (1)$$

it is called a "local ranking" model.

Let $\left\{g_k\left(y_i^{(q)}, y_j^{(q)}, x^{(q)}\right)\right\}_{k=1}^{K}$ be a set of real-value functions defined on $y_i^{(q)}$, $y_j^{(q)}$, and $x^{(q)}$ $(i, j = 1, \dots, n^{(q)}, i \neq j)$. The functions

$$g\left(y_i, y_j, x\right) \qquad (2)$$

represents the relations between documents. Equation 2 is defined according to the requirements of different tasks. For example, for the Pseudo Relevance Feedback problem, Qin et al. (2008) defined Equation 2 as the similarities between any two documents in their CRF-based model.

If a ranking model takes all the documents as its input and exploits both local and global information (Equation 2) in the documents, i.e., in the form of

$$y^{(q)} = F\left(x^{(q)}\right),$$

it is called a "global ranking" approach.

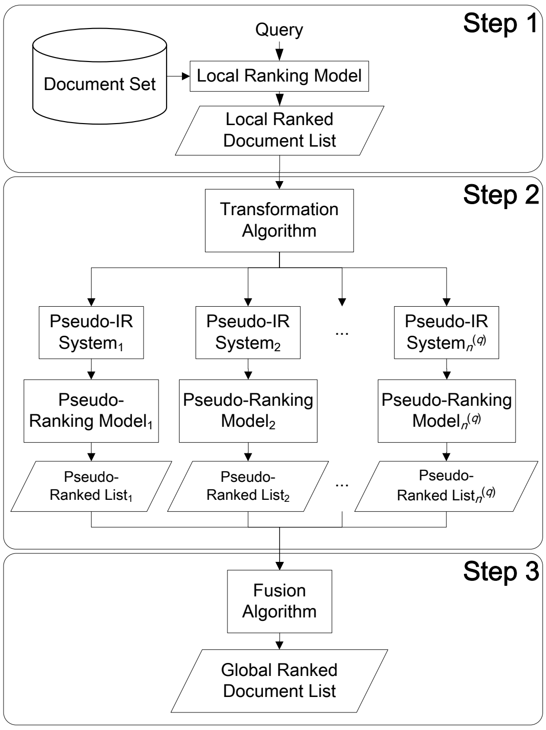

Figure 1. The Proposed Framework for Global Ranking.

3 Fusion-based Global Ranking Framework

It is usually difficult to develop a global ranking algorithm that can fully utilize all the local and global information in documents to produce a document rank and also consider the score ranks. One example of a global ranking algorithm that satisfied these criteria is the one proposed in (Qin et al., 2008) in which the modified CRF algorithm handles context (local) features and relational (global) features in the documents. Without solving a ranking problem directly, however, the modified CRF algorithm is more like a regression algorithm since it optimizes the CRF parameters in a maximum likelihood estimate without considering the score ranks. With respect to the ranking feature, in this section, we describe our framework based on the idea of data fusion for solving the global ranking problem.

3.1 Framework Description

The flow chart of the proposed framework is illustrated in Figure 1. The first step is the same as that of the traditional local ranking model. Given a query, the local ranking model $y_i^{(q)}$ defined in Equation 1 is used to calculate the ranking score for each document, and return a document list sorted according to the local scores.

The second step transforms the global ranking problem into a data fusion problem. Our idea is to take each retrieved document as a pseudo-IR system, and the pseudo-ranking model, $y'^{(q)}_i$, used by each system is the function defined in Equation 2. For each pseudo-IR system, $x_j^{(q)}$, the pseudo-ranking model for a document $x_i^{(q)}$ is defined as follows:

$$y'^{(q)}_i = f\left(x_i^{(q)}\right) = g\left(y_i^{(q)}, y_j^{(q)}, x^{(q)}\right), \quad i = 1, \dots, n^{(q)}. \qquad (3)$$

There are totally $n^{(q)}$ pseudo-IR systems, which generate $n^{(q)}$ pseudo-ranked lists. As a result, the global ranking problem is transformed into a data fusion problem, that is to aggregate the pseudo-ranked lists. Figure 2 shows the steps of the transformation algorithm.

The final step is to adapt fusion algorithms to aggregate the pseudo-ranked lists. A canonical data fusion task is called *meta-search* (Aslam and Montague, 2001; Fox and Shaw, 1994; Lee, 1997; Nuray and Can, 2006), which aggregates Web search query results from several engines into a more accurate ranking. The origin of research on data fusion can be traced back to (Borda, 1781). In recent years, the process has been used in many new applications, including aggregating data from microarray experiments to discover cancer-related genes (Pihura et al., 2008), integration of results from multiple mRNA studies (Lin and Ding, 2008), and similarity searches across datasets and information merging (Adler et al., 2009; Zhao et al., 2010).

Liu et al. (2007) classified data fusion technologies into two categories: order-based fusion and score-based fusion. In the first category, the orders of the entities in individual ranking lists are used by the fusion algorithm. In the second category, the entities in individual ranking lists are assigned scores and the fusion algorithm uses the scores. In the following sub-sections, we adapt three fusion algorithms

```
function transform ($x^{(q)}$: the documents retrieved
with query $q$)
{generate pseudo-ranked lists for $x^{(q)}$}
# a dictionary that maps the pseudo-IR systems to
# their corresponding pseudo-ranked lists
pseudoRankedLists = {}
for $x_i^{(q)}$ in $x^{(q)}$:
    # a dictionary that maps the relation score (real
    # value) to a list of documents.
    relation = {}
    for $x_j^{(q)}$ in $x^{(q)}$:
        relation[$g\left(y_i^{(q)}, y_j^{(q)}, x^{(q)}\right)$].append($x_j^{(q)}$)
    # relation.keys() returns all keys stored in the
    # dictionary relation. The key of relation is the
    # relation score.
    Sort relation.keys() in decreasing order
    # a dictionary that maps a new rank to a list of
    # documents.
    pseudoRankedList = {}
    newRank = 0
    for score in sorted relation.keys():
        # relation[score] returns the document list
        # corresponding to the given score
        for doc in relation[score]:
            pseudoRankedList[1+newRank]
                                   .append(doc)
        newRank = newRank + 1
    pseudoRankedLists [$x_i^{(q)}$] = pseudoRankedList
return pseudoRankedLists
```

Figure 2. The Dependent Ranked List Generation Algorithm (represented using python syntax).

for the proposed framework. The first is the Borda-fuse model (Aslam and Montague, 2001), an order-based fusion approach based on an optimal voting procedure. The second is a linear combination (LC) model (Vogt and Cottrell, 1999), which is a score-based fusion approach.

3.2 Borda-fuse

The Borda-fuse model (Aslam and Montague, 2001) is based on a political election strategy called the Borda Count. For our framework, the rationale behind the strategy is as follows. Each pseudo-IR system $x_j^{(q)}$ is an analogy for a voter; and each voter ranks a fixed set of $n^{(q)}$ documents in order of preference (Equation 3). For each voter, the top ranked document is given $n^{(q)}$ points, the second ranked document is given $n^{(q)}$-1 points, and so on. If some documents left unranked by the voter, the remaining points are divided equally among the unranked documents. The documents are ranked in descending order of the total points.

In our framework, we implement two Borda-fuse-based models. The first is the modified Borda-fuse (MBF) model. In MBF, the number of points given for a voter's first and subsequent preferences is determined by the number of documents they have actually ranked, rather than the total number of ranked. Because the ranking model, $y'^{(q)}_i$, used by the pseudo-IR system may only retrieve m documents where m is smaller than $n^{(q)}$, we penalize systems that do not rank a full document set by reducing the number of points their vote distributes among the documents. In other words, if there are ten documents, but the pseudo-IR system only retrieves five, then the first document will only receive 5 points; the second will receive 4 points, and so on.

The second is the weighted Borda-fuse (WBF) model. The original Borda-fuse model reflects a democratic election in which each voter has equal weight. However, in many cases, we prefer some voters because they are more reliable. We employ a simple weighting scheme that multiplies the points assigned to a document determined by system $x_j^{(q)}$ by a weight $w_{x_j^{(q)}}$.

3.3 LC Model

The LC model has been used by many IR researchers with varying degrees of success (Bartell et al., 1994; Knaus et al., 1995; Vogt and Cottrell, 1999; Vogt and Cottrell, 1998). In our framework, it is defined as follows. Given a query q , a document $x_i^{(q)}$, the weights $\mathbf{w} = \left(w_1, w_2, w_3, \ldots, w_{n^{(q)}}\right)$ for $n^{(q)}$ individual pseudo-IR systems, and jth pseudo-IR system's ranking score ${y_j}'^{(q)}_i$, the LC model calculates the ranking score ρ of $x_i^{(q)}$ against all pseudo-IR systems as follows:

$$\rho\left(\mathbf{w}, x_i^{(q)}\right) = \sum_{j=1}^{n^{(q)}} w_i {y_j}'^{(q)}_i \qquad (4)$$

This score is then used to rank the documents. For example, for two pseudo-IR systems, this reduces to:

$$\rho\left(w_1, w_2, x_i^{(q)}\right) = w_1 {y_1}'^{(q)}_i + w_2 {y_2}'^{(q)}_i$$

Compared with MBF, Equation 4 requires both relevance scores and training data to determine the weight w_i given to each pseudo-IR system.

4 Case Study

In this section, we describe the task examined in our study. We also explain how we formulate the task as a global ranking problem. The experiments results are detailed in Section 5.

4.1 Interactor Normalization Task

The interactor normalization task (INT) is a complicated text mining task that involves the following steps: (1) It recognizes gene mentions in a full text article. (2) It maps the recognized gene mentions to corresponding unique database identifiers which is similar to the word sense disambiguation task in computational linguistics. (3) It generates a ranked list of the identifiers according to their importance in the article and their probability of playing the interactor role in protein-protein interactions (PPIs). Such ranked lists are useful for human curators and can speed up PPI database curation.

Dai et al. (2010) won first place in the BioCreAtIvE II.5 INT challenge (Mardis et al., 2009) by using a SVM-based local ranking model in which they treat gene mentions' identifiers in an article as the document set, and the query is a constant string "interactor". Based on their feature sets and evaluation results, we can find that their local ranking model tends to rank focus genes higher (Dai et al., 2010). However, the primary objective of INT is to generate a ranked list of interaction gene identifiers. According to (Jenssen et al., 2001), co-mentioned genes are usually related in some way. For example, if two gene mentions frequently occur alongside each other in the same sentence in an article, they probably have an association and influence each other's rank. Take a low-ranked interactor that is only mentioned twice in an article as an example. If both mentions are next to the highest-ranked interactor in the article, then the low-ranked interactor's rank should be boosted significantly. Therefore, the ranking task for each article can be formulated as a global ranking problem; the global ranking algorithm should consider both the local information from Dai et al.'s model and the global information from the associations among identifiers.

4.2 Global Ranking in INT

Let q be a constant "interactor." The identifier set generated by an INT system for a full-text article is analogous to the document set $x^{(q)} = \left\{x_1^{(q)}, x_2^{(q)}, \dots, x_{n^{(q)}}^{(q)}\right\}$. Here $n^{(q)}$ denotes the number of identifiers. Note that the number of identifiers varies for different articles. Let $y^{(q)} = \left\{y_1^{(q)}, y_2^{(q)}, \dots, y_{n^{(q)}}^{(q)}\right\}$ denote the ranking scores assigned to the identifiers given by a local ranking model. In this study, we used the INT system and SVM-based local ranking model released by (Dai et al., 2010) to generate the identifier set and ranking scores.

To obtain the global information, we consider the co-occurrence of identifiers and employ mutual information (MI) to measure the association between two identifiers as follows:

$$g(y_i, y_j, x) = \mathrm{MI}(x_i, x_j) =$$
$$P(x_i, x_j) / \left(P(x_i) \times P(x_j)\right).$$

In the above formula, the identifier probabilities $P(x_i)$ and $P(x_j)$ are estimated by counting the number of occurrences in an article normalized by N, i.e., the number of sentences containing identifiers. The joint probability, $P(x_i, x_j)$, is estimated by the number of times x_i co-occurs with x_j in a window of k words normalized by N. Note that, in practice, other advanced approaches can be used to calculate the association score.

For the proposed framework, each identifier $x_i^{(q)}$ is a pseudo-IR system with MI as its pseudo-ranking model $y'^{(q)}_i$. The identifiers that co-occur with $x_i^{(q)}$ become candidates on $x_i^{(q)}$'s pseudo-ranked list.

5 Experiments

In the following sub-sections, we introduce the dataset used in the experiments, describe the evaluation methods, report the results of the experiments conducted to compare the performance of different methods, and discuss the efficiency of the proposed global ranking framework.

5.1 Dataset

We used the BioCreAtIvE II.5 Elsevier corpus released by BioCreAtIvE II.5 challenge in the experiments. The corpus contains 1,190 full-text journal articles selected mainly from FEBS Letters. Following the same format as the BioCreAtIvE II.5 INT challenge, we used articles published in 2008 (61 articles) as our training set and articles published in 2007 or earlier (61 articles) as our test set.

5.2 A Fusion-based Global Ranking Framework for INT

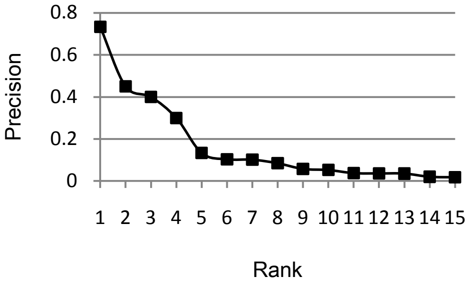

Figure 3. Precision of Different Ranks.

Before applying the proposed framework, we preprocess the articles in the dataset to identify all gene mentions, and map them to their corresponding identifiers. After preprocessing, each full-text article is associated with a list of identifiers (Step 1 in Figure 1). The transform and fusion algorithm is then applied on each article (Steps 2 and 3 in Figure 1).

To apply the WBF and LC models, we need to determine the weight assigned to each pseudo-IR system. To obtain the weight, we calculate the precision of each rank of the ranked lists generated by Dai et al.'s INT system. Figure 3 shows the precision of ranks 1 to 15 calculated by applying three-fold cross validation on the INT training set. We observe that the precision declines as the rank increases, which implies that the higher ranks predicted by their SVM-based local ranking model are more reliable than the lower ranks.

5.3 Evaluation Metrics

Our evaluations focus on two comparisons: the first compares the ranking of the proposed framework with the original local ranking model by using the area under the curve of the interpolated precision/recall (iP/R) curve. This is the evaluation metric used in the BioCreAtIvE II.5 challenge and is a common way to depict the degradation of precision as one traverses the retrieved results by plotting interpolated precision numbers against percentage recall. The area under the iP/R function f_{pr} is defined as follows:

$$Area\_iPR\left(f_{pr}\right) = \sum_{j=1}^{n}\left(p_{i_j} \times \left(r_j - r_{j-1}\right)\right),$$

$$p_i(r) = \max_{r' \geq r} p(r')$$

where n is the total number of correct identifiers and p_i is the highest interpolated precision for the correct identifier j at r_j, the recall for that hit. The interpolated precision p_i is calculated for each recall r by taking the highest precision at r or any $r' \geq r$.

In the second comparison, we use a standard quality measure in IR to estimate the ranking performance of local ranking models and the proposed framework. We adopt Normalized Discounted Cumulative Gain (NDCG) to measure the performance. The NDCG score of a ranking is computed based on DCG (Discounted Cumulative Gain) as follows:

$$\mathrm{DCG}(r) = g(1) + \sum_{i=2}^{r} \frac{g(i)}{\log_2(i)},$$

where r is the rank position, and $g(i) \in \{0,1\}$ is the relevance grade of the ith identifier in the ranked result set. In our experiment, $g(i) = 1$ corresponds to an interaction identifier, and $g(i) = 0$ corresponds to other identifiers. NDCG is then computed as follows:

$$\mathrm{NDCG}(r) = \frac{\mathrm{DCG}(r)}{\mathrm{IDCG}(r)},$$

where IDCG denotes the results of a perfect ranking. The NDCG values for all articles are averaged to obtain the average performance of the proposed framework.

5.4 INT Test Set Performance

Figure 4 shows the *Area_iPR* scores of four configurations. In the baseline configuration (Local/Rank1), the SVM-based local ranking model released by Dai et al. is employed. In the configuration Global+LC, Global+MBF, and Global+WBF, the proposed global ranking framework is cascaded with the local ranking model and with three data fusion models: the LC model, the modified Borda-fuse (MBF) model, and the weighted Borda-fuse model. The figure also shows the *Area_iPR* scores of

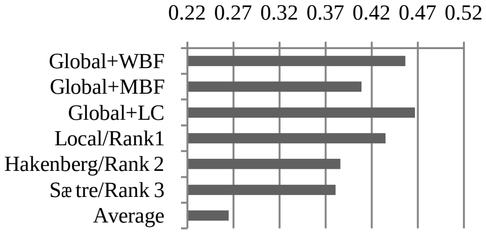

Figure 4. The *Area_iPR* Results of Different Ranking Models

the top three teams and the average *Area_iPR* score of all BioCreAtIvE II.5 INT participants (Average).

The results show that under the global ranking framework, *Area_iPR* performance is improved in addition to Global+MBF. The highest *Area_iPR* (Global+LC: 46.7%) is 3.2% higher than the Rank 1 score in the BioCreAtIvE II.5 INT challenge. According to our analysis, before global ranking, identifiers whose feature values rarely appear in the training set are often ranked incorrectly because their feature values are under-estimated by the ranking model. However, if the identifiers cooccur with higher-ranked identifiers whose feature values appear frequently, the proposed framework is very likely to increase their ranks. This results in an improved *Area_iPR* score.

5.5 Global Ranking Performance

Based on	Global Ranking	NDCG1	NDCG3	NDCG5
Local Ranking /Rank1	Global+LC	+0.908	+1.323	-0.003
	Global+MBF	-3.279	-1.034	-0.020
	Global+WBF	-0.016	+3.630	+2.071
Freq	Global+LC$_f$	+1.639	+3.152	+2.817
	Global+MBF$_f$	-6.860	-4.275	-4.839
	Global+WBF$_f$	+2.549	+2.390	+3.043

Table 1. The NDCG Gain (%) of Different Ranking Models.

To illustrate the effectiveness of the proposed global ranking framework and assess its performance when it is cascaded with other conventional ranking models, we implement a simple term frequency-based ranking function, which is based on the identifier frequency in an article as another local ranking model. If two or more identifiers have the same frequency, two heuristic rules are employed sequentially to rank them: (1) the identifier with the highest frequency in the Results section of the article, and (2) the identifier mentioned first in the article.

Table 1 shows the NDCG percentage gain of different ranking models. It compares the ranked list generated by our global ranking framework and by the local ranking models. We observe that (1) irrespective of whether the local ranking model is a conventional model or a learning to rank model, Global+LC and Global+WBF models achieve NDCG gains over the original rankings of the local ranking models; (2) the results show that our global ranking framework can improve the performance by only exploiting MI analysis. However, it is expected that employing more advanced relation extraction methods to determine the global information (Equation 3) would yield more reliable pseudo-ranked lists and lead to a further improvement in the final ranking; and (3) similar to the results in Section 5.4, the performance of Global+MBF does not improve. Global+MBF has a negative NDCG gain and the *Area_iPR* decreases by 2.61%. We believe this is due to MBF gives equal weight to each pseudo-IR system. As mentioned in Section 4.1, the document set in INT is comprised of the identifiers of the gene mentions derived by Dai et al.'s system. Unfortunately, there must be incorrect identifiers (the errors may be due to their gene mention recognition or identifier mapping processes). As in the meta-search, the best performance is often achieved by weighting the input systems unequally. Reasonable weights allow the algorithm to concentrate on good feedback from pseudo-IR systems and ignore poor feedback. As shown by the average precision results in Figure 3, the identifiers (corresponding to the pseudo-IR systems in our framework) in the higher ranks are more reliable; however, MBF cannot use this information, which leads to a negative NDCG gain and a lower *Area_iPR* score.

6 Conclusion

We have presented a new global ranking framework based on data fusion technology. Our approach solves the global ranking problem in three stages: the first stage ranks the document set by the original local ranking model; the second stage transforms the problem into a data fusion task by using global information, and the final stage adapts fusion algorithms to solve the ranking problem. The framework is flexible and it can be combined with other mature ranking models and fusion algorithms. We also show how the BioCreAtIvE INT can be formulated as a global ranking problem and solved by the proposed framework. Experiments on the INT dataset demonstrate the effectiveness of the proposed framework and its superior performance over other ranking models.

In our future work, we will address the following issues: (1) the use of advanced data fusion algorithms in the proposed framework; (2) assessing the performance of the proposed framework on other tasks, such as Pseudo Relevance Feedback and Topic Distillation; and (3) design an advanced supervised learning relation extraction algorithm to replace MI in INT to evaluate the system performance.

References

Adler, P., R. Kolde, M. Kull, A. Tkachenko, H. Peterson, J. Reimand and J. Vilo (2009). *Mining for coexpression across hundreds of datasets using novel rank aggregation and visualization methods*. *Genome Biology* 10(R139).

Aslam, J. A. and M. Montague (2001). *Models for metasearch*. *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*, New Orleans, Louisiana, United States, ACM.

Bartell, B. T., G. W. Cottrell and R. K. Belew (1994). *Automatic combination of multiple ranked retrieval systems*. *Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval*, Dublin, Ireland Springer-Verlag New York, Inc.

Borda, J. (1781). *Mémoire sur les élections au scrutin*. *Histoire del'Acad´emie Royale des Sciences* 2: 13.

Chowdhury, G. (2007). *TREC: Experiment and Evaluation in Information Retrieval*. *Online Information Review* 31(5): 462.

Dai, H.-J., P.-T. Lai and R. T.-H. Tsai (2010). *Multi-stage gene normalization and SVM-based ranking for protein interactor extraction in full-text articles*. *IEEE TRANSACTIONS ON COMPUTATIONAL BIOLOGY AND BIOINFORMATICS* 14 May. 2010. IEEE computer Society Digital Library. IEEE Computer Society.

Fox, E. A. and J. A. Shaw (1994). *Combination of Multiple Searches*. *1994*, Proceedings of the Second Text REtrieval Conference (TREC 2)

Jenssen, T.-K., A. Lagreid, J. Komorowski and E. Hovig (2001). *A literature network of human genes for high-throughput analysis of gene expression*. *Nature Genetics* 28(1): 21-28.

Knaus, D., E. Mittendorf and P. Schäuble (1995). *Improving a basic retrieval method by links and passage level evidence*. *NIST Special Publication 500-225: Overview of the Third Text REtrieval Conference (TREC-3).*

Krallinger, M., F. Leitner and A. Valencia (2009). *The BioCreative II.5 challenge overview*. *Proceedings of the BioCreative II.5 Workshop 2009 on Digital Annotations*, Madrid, Spain.

Kwok, K. L. (1984). *A document-document similarity measure based on cited titles and probability theory, and its application to relevance feedback retrieval*. *SIGIR'84*.

Lee, J. H. (1997). *Analyses of multiple evidence combination*. *Proceedings of the 20th annual international ACM SIGIR conference on Research and development in information retrieval*, Philadelphia, Pennsylvania, United States, ACM.

Lin, S. and J. Ding (2008). *Integration of Ranked Lists via Cross Entropy Monte Carlo with Applications to mRNA and microRNA Studies*. *Biometrics* 65(1): 9-18.

Liu, Y.-T., T.-Y. Liu, T. Qin, Z.-M. Ma and H. Li (2007). *Supervised rank aggregation*. *Proceedings of the 16th international conference on World Wide Web*, Banff, Alberta, Canada, ACM.

Mardis, S., F. Leitner and L. Hirschman (2009). *BioCreative II.5: Evaluation and ensemble system performance*. *Proceedings of the BioCreative II.5 Workshop 2009 on Digital Annotations*, Madrid, Spain.

Nuray, R. and F. Can (2006). *Automatic ranking of information retrieval systems using data fusion*. *Inf. Process. Manage.* 42(3): 595-614.

Pihura, V., S. Dattaa and S. Datta (2008). *Finding common genes in multiple cancer types through meta–analysis of microarray experiments: A* *rank aggregation approach Genomics* 92(6): 400-403

Qin, T., T.-Y. Liu, X.-D. Zhang, D.-S. Wang and H. Li (2008). *Global Ranking Using Continuous Conditional Random Fields*. *Proceedings of the Twenty-Second Annual Conference on Neural Information Processing Systems (NIPS 2008)*, Vancouver, Canada.

Qin, T., T. Liu, X. Zhang, D. Wang, W. Xiong and H. Li (2008). *Learning to rank relational objects and its application to web search*, ACM.

Vogt, C. and G. Cottrell (1999). *Fusion via a linear combination of scores*. *Information Retrieval* 1(3): 151-173.

Vogt, C. C. and G. W. Cottrell (1998). *Predicting the performance of linearly combined IR systems*. *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*, Melbourne, Australia ACM.

Zhao, Z., J. Wang, S. Sharma, N. Agarwal, H. Liu and Y. Chang (2010). *An Integrative Approach to Identifying Biologically Relevant Genes*. *Proceedings of SIAM International Conference on Data Mining (SDM)*.