

# **RANKING BASED ON COLLABORATIVE FEATURE-WEIGHTING APPLIED TO THE RECOMMENDATION OF RESEARCH PAPERS**

Amir E. Sarabadani Tafreshi<sup>1</sup>, Amirehsan Sarabadani Tafreshi<sup>1</sup> and Anca L. Ralescu<sup>2</sup>

<sup>1</sup>ETH Zürich, CH-8092 Zürich, Switzerland

<sup>2</sup>University of Cincinnati, Cincinnati, Ohio, USA

## ***ABSTRACT***

Current research on recommendation systems focuses on optimization and evaluation of the quality of ranked recommended results. One of the most common approaches used in digital paper libraries to present and recommend relevant search results, is ranking the papers based on their features. However, feature utility or relevance varies greatly from highly relevant to less relevant, and redundant. Departing from the existing recommendation systems, in which all item features are considered to be equally important, this study presents the initial development of an approach to feature weighting with the goal of obtaining a novel recommendation method in which features which are more effective have a higher contribution/weight to the ranking process. Furthermore, it focuses on obtaining ranking of results returned by a query through a collaborative weighting procedure carried out by human users. The collaborative feature-weighting procedure is shown to be incremental, which in turn leads to an incremental approach to feature-based similarity evaluation. The obtained system is then evaluated using Normalized Discounted Cumulative Gain (NDCG) with respect to a crowd-sourced ranked results. Comparison between the performance of the proposed and Ranking SVM methods shows that the overall ranking accuracy of the proposed approach outperforms the ranking accuracy of Ranking SVM method.

## ***KEYWORDS***

Ranking, recommendation system, feature weighting, support vector machine.

## **1 INTRODUCTION**

With the widespread of online publications and the growth of the internet use as a potential source of knowledge, recommendation systems play an increasingly important role in helping researchers to find the papers of their interest. Most recommendation systems return results ranked by their relevancy [1] and users are then free to consider as many of the returned results, usually the top ranked [7, 8, 20]. Learn-to-rank techniques have been proposed and used in various applications of information retrieval including in recommendation systems [19]. During the process of ranking, provided a group of items, their relative order may well depict their levels of relevance, preference, or importance, depending on the application [6]. An important issue in recommendation systems is that of similarity. Usually, the similarity of a pair of items is defined as an aggregate of the similarities along their features [6]. Features may be numeric as well as nominal representing certain facets of an item. For example, in content based recommendation systems characteristics such as color, price, etc. are used [6]. Similarity computation methods [6] must take into account the different types (numerical, nominal) of features.

The application domain also may determine the importance of a feature: the same feature might have different levels of importance in different domains of applications [21]. Depending on

the application, a feature might be redundant to some other features, less relevant or on the other hand, highly relevant. For instance, in recommendation systems for publications, while selecting a paper as a query, keywords of the paper might be more important to some user than the author names, conference, and location of the conference.

Many feature selection techniques have been proposed to address the issue of relevancy [11]. In these techniques, redundant and less relevant features are removed and the remaining features are considered to be (highly) relevant and used in the ranking process. However, the state-of-the-art learning to rank approaches merely analyze the relation among the features from the aspects of feature selection [12], which may not reflect the importance of a particular feature. This might be problematic since features which may increase the accuracy of the system may get removed, while, at the same time, some features which are not as useful are preserved.

On the other hand, it has been shown that good weighting techniques are more critical and better than the feature selection process [5]. Ultimately, it is very important to weigh features based on their level of importance in a specific domain [21].

The present work develops a recommendation system for digital paper libraries in which features are weighed. The input to the systems is a collection of features used to describe a paper. Such features include key-words, author names, conference name, location, and year. These features are weighted based on their individual level of importance to identify correctly ranked recommended papers.

The steps of the proposed approach, applied to a research paper recommendation system, are as follows: (1) **Feature extraction.** Predefined publication features (e.g., Title, Author, etc.) are extracted for each paper in the database considered; (2) **Identification of ground-truth relevance.** Ground truth (GT) relevance is extracted from correctly ranked relevant papers (ground-truth) for a possible user query (e.g. a paper); (3) **Collaborative feature weighting.** Feature importance, and hence weight, is obtained from a survey of human users (in the experiment presented here, 20 computer science graduate students using digital libraries of their domain of expertise participated in this survey); (4) **Evaluate similarity.** The similarities between query and the database of papers along features are weighted by the corresponding feature weight and aggregated to produce the final relevance scores. (5) **Return results.** A list of results, ranked by their relevance scores is returned.

The performance of the proposed method is evaluated by comparing the results, sorted by their relevance, with the corresponding ground-truth ranked papers, which were already identified, using Normalized Discounted Cumulative Gain (NDCG) [13]. An empirical study on Ranking SVM [14] is also conducted. Then, the achieved performance by proposed and Ranking SVM methods is compared.

## 2 RELATED WORK

Recent work has explored three classes of methods for learning to rank, including point-wise, pair-wise and list-wise methods [16]. Existing methods [9, 11, 17, 18] of ranking systems, have in common a two-stage paradigm [16]: (1) Select a subset of features via a feature selection method, and (2) Learn a ranking predictor from the selected features. However, these separated stages cannot guarantee that the selected features in the first stage are optimal for learning the ranking predictors in the second stage [16]. Indeed, the feature selection process can lead to removal of some of the features, which if processed together, might increase the accuracy of the recommending system. Commonly used similarity measures (e.g. Euclidean distance based, or, for documents, the cosine similarity) assume implicitly that all features are equally important. However, human common sense in evaluation of similarity between items frequently gives different weights to various features [10]. Therefore, feature selection is replaced in this work with *feature weighting* such that

all features are preserved and individual feature effectiveness in the ranking process is conveyed by a feature weight.

### 3 MATERIALS AND METHODS

This section details the current approach, including the data set used, and the feature weighting process.

#### 3.1 Ranking of research papers using feature weighting

First, a simple survey was conducted to obtain the weight of the features and implicitly the importance of each feature in the community under consideration. Weighting the features not only helps to achieve a promising ranking result but also addresses collaboration of the users in recommendation systems and helps avoid the cold start issue. To weigh the relevant features in a specific domain, the evaluators need deep knowledge and experience in that field and to know the relative importance of the features. In general, utilizing expert judgments to directly allocate weights is an issue in any weighting decision [2, 4]. Therefore, for the domain of application of the method described here, a community of 20 graduate students in computer science was used to weigh the features of research articles in computer science. The final weight of each feature was calculated by averaging the weights assigned by all the subjects:  $\bar{W}_N = \frac{w_1 + \dots + w_N}{N}$ , where  $\bar{W}_N$  denotes a feature average weight computed as an average of the weights given by  $N$  subjects and  $W_i$ ,  $i = 1, \dots, N$  is the weight assigned by user  $i$ . Averaging of weights to aggregate them across all users has the advantage that it can be done incrementally. Indeed, when an additional user assigns a weight,  $w_{N+1}$ , to an item then the updated average,  $\bar{W}_{N+1}$ , can be easily obtained as  $\bar{W}_{N+1} = \frac{N}{N+1}\bar{W}_N + \frac{1}{N+1}w_{N+1}$ . Table 1 shows the average weight (and normalized average weights) of each feature achieved by conducting the survey.

Table 1: Feedback from the Survey

Features	Weight	Normalized Weight
Keywords	0.85	0.3036
Year	0.35	0.1250
Conference	0.75	0.2679
Location	0.34	0.1214
Author	0.51	0.1821

Given a query paper, the similarity between it and each paper of a list of publications is computed based on their feature similarities. These are calculated using the Boolean model for single-valued features (e.g. conference) and probabilistic model of the traditional Robertson/Sparck Jones formalization model [23] for multiple-valued features (e.g. keywords). The boolean model assigns 1 or 0 according to whether features match or not. The probabilistic model evaluates features similarity as follows: assume that feature  $k$  takes on the values  $k_l$ , with  $l = 1, \dots, m_k$ . Let  $p_{ijk_l}$  be the probability that value  $k_l$  of feature  $k$  of the paper  $O_i$  is also the value of feature  $k$  of the paper  $O_j$ , and let  $q_{ijk_l}$  the probability that the value  $k_l$  of feature  $k$  in  $O_i$  is not a value of feature  $k$  in  $O_j$ . Then, the similarity along feature  $k$  of the two papers,  $O_i$  and  $O_j$ , is defined as in equation (1) [22].

$$Sim(F_{ik}, F_{jk}) = \sum_{l=1}^{m_k} \left( \log \frac{p_{ijk_l}}{1-p_{ijk_l}} + \log \frac{1-q_{ijk_l}}{q_{ijk_l}} \right) \quad (1)$$

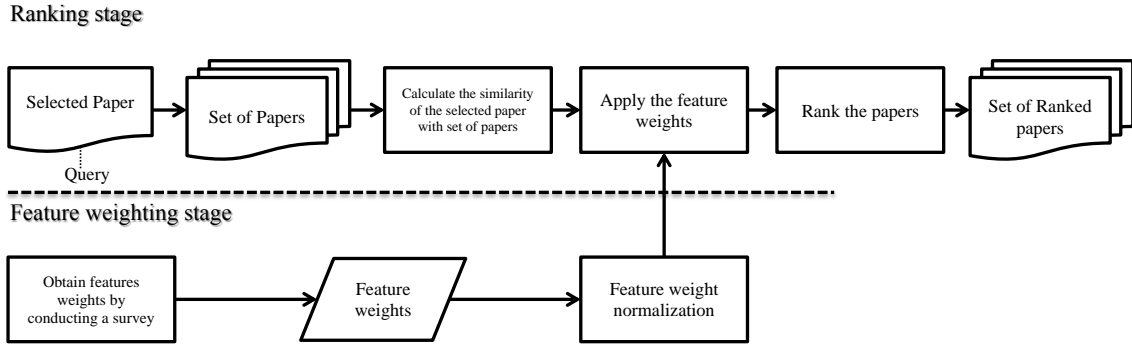


Figure 1: System overview.

The papers are sorted and ranked based on their overall similarity score,  $SimScore_N(O_i, O_j)$ , obtained by the weighted aggregation of their individual feature similarities, shown in equation (2).

$$SimScore_N(O_i, O_j) = \sum_{k=1}^n Sim(F_{ik}, F_{jk}) \times \bar{W}_{k,N} \quad (2)$$

where  $F_{ik}$  is feature  $k$  in paper  $O_i$ , and  $\bar{W}_{k,N}$  is its normalized weight, and  $Sim(F_{ik}, F_{jk})$  computed from equation (1). Incrementally, considering the  $N + 1$ st subject leads to

$$SimScore_{N+1}(O_i, O_j) = \sum_{k=1}^n Sim(F_{ik}, F_{jk}) \times \bar{W}_{k,N+1}$$

and letting  $\alpha = \frac{N}{N+1}$ , to

$$\begin{aligned} SimScore_{N+1}(O_i, O_j) &= \sum_{k=1}^n Sim(F_{ik}, F_{jk}) \times [\alpha \bar{W}_{k,N} + (1 - \alpha) W_{k,N+1}] \\ &= \alpha \sum_{k=1}^n Sim(F_{ik}, F_{jk}) \bar{W}_{k,N} + (1 - \alpha) \sum_{k=1}^n Sim(F_{ik}, F_{jk}) W_{k,N+1} \\ &= \alpha SimScore_N(O_i, O_j) + (1 - \alpha) \sum_{k=1}^n W_{k,N+1} Sim(F_{ik}, F_{jk}) \end{aligned}$$

Figure 1 shows the system overview of the proposed approach.

### 3.2 The Dataset

The dataset used for experiments contains 20 conference papers from 2009 to 2012 on database and information systems from the Global information systems group (ETH Zurich) website<sup>1</sup>. Each paper is described using the author name, Title, conference name, year, and location, the publisher, and keywords.

### 3.3 Performance Evaluation

In order to evaluate the performance of the proposed method, ground-truth ranked publications for each selected paper were needed. To find the ground-truth ranked list for each paper, a simple crowd-sourcing algorithm was run. Students were asked to rank the most relevant papers to each single paper independently. The results were aggregated by using a Local Keminization algorithm [15] to obtain a single ground-truth ranked relevant paper list for each paper. The ground-truth ranked list for each paper essentially reflects ranking which the considered community generally assumes to be correct. A simple linguistic summarization of the returned results is implemented by labeling the 1st 25% of the ranked papers as ‘definitely relevant’, the next 25% are labeled as

<sup>1</sup><http://www.globis.ethz.ch/publications>

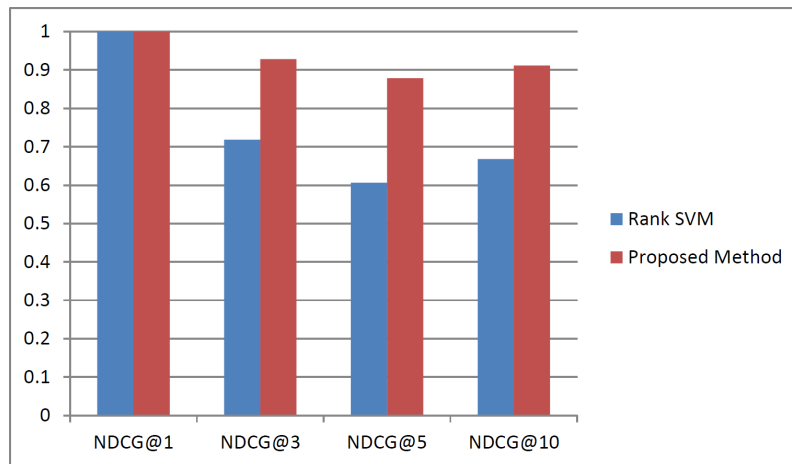


Figure 2: Ranking accuracies: evaluation and comparison of the proposed and the rank SVM methods using NDCG at positions 1, 3, 5 and 10.

‘partially relevant’. The second half of the ground-truth ranked papers are labeled as ‘irrelevant’. Ranking SVM was chosen as a reference to evaluate the relative performance of the proposed method. To rank the papers relevant to each single paper using this method, a leave-one-out strategy was used. Training was done using ground truth rankings of all except one paper. The trained algorithm was then used to predict the ranking list of the left out paper. This procedure was repeated for all the papers.

Many measures have been proposed to evaluate the ranking results, such as MAP [3] and Normalized Discounted Cumulative Gain (NDCG) [13]. In this paper, NDCG at the positions of 1, 3, 5, and 10 was used to evaluate the ranking lists produced by the proposed and Ranking SVM methods with respect to ground truth ranking lists resulted from aggregation of personalized rankings from each student.

## 4 RESULTS

To analyze the performance of the methods in terms of NCDG, the results were averaged over the twenty trials, as shown in Fig. 2.

As it can be seen from Figure 2 the approach adopted in this paper yields a much higher accuracy with respect to the ground truth, on average in excess of 90%. Moreover, when compared with the SVM approach, it can be seen that the two approaches agree in the accuracy of the highest ranked item (paper). However, they strongly disagree in the accuracies at positions 3, 5, and 10, where SVM accuracy varies between less than 60% to at most slightly over 70%.

## 5 DISCUSSION AND CONCLUSION

We described a method that uses feature weights to reflect their importance in digital library systems to effectively recommend ranked related papers in response to a user query (selected paper). The approach relies on *explicit feature weighting* in the ranking process, rather than implicitly considering equal significance for the features. Being based on input from a community of human users, the feature weighing approach can be thought of as *collaborative feature weighting approach*. Moreover, the incremental aspect of the procedure for feature weighting and for feature-based similarity computation ensures that collaborative weighting and similarity evaluation can

be easily adapted when more users provide their weights for the same set of items. This method shows a high retrieval performance which also matches the human user perception. The current work was specifically focused on the ranked style recommendation systems results and takes into account the importance of features directly, rather than conducting a feature selection and then leaving the rest of the process for learning to rank algorithms. The results show that a general feature weighting procedure can help the ranking process to closely reach the ground-truth results.

Comparison to Ranking SVM using NDCG, shows that the collaborative weighting results in a higher ranking accuracy and thus it conveys its effectiveness. The results obtained on the small scale experiment described in this paper are sufficiently promising to warrant further exploration of the proposed approach on a large database of papers. Although using a survey for feature weighting leads to consistency with human user preferences, at the same time, running such a survey can be very time consuming and expensive if it were required for the members of a research community. Selecting such a group and how it can affect the objectives of a research community is an issue to be explored in a future study, as are incremental approaches for collaborative weight elicitation in which new survey subjects can be used to update feature weights.

## References

- [1] G. Adomavicius and Y. Kwon. Improving aggregate recommendation diversity using ranking-based techniques. *Knowledge and Data Engineering, IEEE Transactions on*, 24(5):896–911, may 2012.
- [2] B. S. Ahn and K. S. Park. Comparing methods for multiattribute decision making with ordinal weights. *Computers & Operations Research*, 35(5):1660–1670, 2008.
- [3] R. Baeza-Yates, B. Ribeiro-Neto, et al. *Modern information retrieval*, volume 463. ACM press New York, 1999.
- [4] F. Barron and B. Barrett. Decision quality using ranked attribute weights. *Management Science*, 42(11):1515–1523, 1996.
- [5] C. Buckley. The importance of proper weighting methods. In *Proceedings of the workshop on Human Language Technology, HLT '93*, pages 349–352, Stroudsburg, PA, USA, 1993. Association for Computational Linguistics.
- [6] Y. Cao, J. Xu, T.-Y. Liu, H. Li, Y. Huang, and H.-W. Hon. Adapting ranking svm to document retrieval. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval, SIGIR '06*, pages 186–193, New York, NY, USA, 2006. ACM.
- [7] S. Cl  men  on and N. Vayatis. Ranking the best instances. *The Journal of Machine Learning Research*, 8:2671–2699, 2007.
- [8] D. Cossock and T. Zhang. Subset ranking using regression. In *Learning theory*, pages 605–619. Springer, 2006.
- [9] S. F. Da Silva, M. X. Ribeiro, J. d. E. Batista Neto, C. Traina-Jr, and A. J. Traina. Improving the ranking quality of medical image retrieval using a genetic feature selection method. *Decision Support Systems*, 51(4):810–820, 2011.
- [10] S. Debnath, N. Ganguly, and P. Mitra. Feature weighting in content based recommendation system using social network analysis. In *Proc. Intl. Conf. on World Wide Web (WWW)*, pages 1041–1042. ACM, 2008.

- [11] X. Geng, T.-Y. Liu, T. Qin, and H. Li. Feature selection for ranking. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 407–414. ACM, 2007.
- [12] G. Hua, M. Zhang, Y. Liu, S. Ma, and L. Ru. Hierarchical feature selection for ranking. In *Proceedings of the 19th International Conference on World Wide Web, WWW '10*, pages 1113–1114, New York, NY, USA, 2010. ACM.
- [13] K. Järvelin and J. Kekäläinen. Ir evaluation methods for retrieving highly relevant documents. In *Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval*, pages 41–48. ACM, 2000.
- [14] T. Joachims. Optimizing search engines using clickthrough data. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 133–142. ACM, 2002.
- [15] J. G. Kemeny. Mathematics without numbers. *Daedalus*, 88(4):577–591, 1959.
- [16] H.-J. Lai, Y. Pan, Y. Tang, and R. Yu. Fsmrank: Feature selection algorithm for learning to rank. 2013.
- [17] C. Li, L. Shao, C. Xu, and H. Lu. Feature selection under learning to rank model for multimedia retrieve. In *Proceedings of the Second International Conference on Internet Multimedia Computing and Service*, pages 69–72. ACM, 2010.
- [18] F. Pan, T. Converse, D. Ahn, F. Salvetti, and G. Donato. Greedy and randomized feature selection for web search ranking. In *IEEE 11th International Conference on Computer and Information Technology (CIT)*, pages 436–442. IEEE, 2011.
- [19] T. Qin, T.-Y. Liu, J. Xu, and H. Li. Letor: A benchmark collection for research on learning to rank for information retrieval. *Information Retrieval*, 13:346–374, 2010.
- [20] C. Rudin. Ranking with a p-norm push. In *Learning Theory*, pages 589–604. Springer, 2006.
- [21] A. G. A. Saeid M. and S. H. Rank-order weighting of web attributes for website evaluation. *IAJIT*, Vol. 8, 2011.
- [22] E. W. Selberg. Information retrieval advances using relevance feedback. *UW Dept. of CSE General Exam*, 1997.
- [23] C. J. van Rijsbergen. *Information Retrieval*. Butterworth-Heinemann, 1979.