An Intelligent Recommendation Service for Student-Selection on Research Social Network: Bridging the Gap Between Students and Supervisors—Research-in-Progress

Ming-Yu ZHANG¹ and Jian-Shan SUN²

¹School of Management, Putian University, PTU, Putian, China
²School of Management, Hefei University of Technology, HFUT, Hefei, China

Keywords: Student-Selection, Research Social Network, Intelligent Recommendation

Abstract. Student-selection is of critical importance for research supervisors in the higher education environment. On account of the information asymmetry, it poses a significant challenge for supervisors to find the most appropriate students. Current studies are limited to the context of one university. They are not suitable in web 2.0 era which are inundated with vast online information. This article proposes an intelligent approach with the help of recommendation system techniques which have emerged in both research social network to provide services for scholars and e-commerce applications to serve for consumers. Furthermore, the method distinguishes a supervisor according to his co-author network firstly. Then, it applies respective recommendation strategy to provide student recommendation services for the target supervisor. A prototype is implemented on Scholarmate, which is a research social network website emphasizing communication between students and supervisors. The evaluation of our proposed method will be completed in the future.

1. Introduction

Postgraduate study is a primary way for undertaking the training of scientific research. To start a student-supervisor relationship, students/supervisors need to make the decision of selecting a suitable supervisor/student for themselves first. Student-selection is profoundly important for supervisors in the higher education environment. Traditionally, supervisors choose students among the ones who contact them proactively. Or, a supervisor can actively show his intention to the students who he prefers after the entrance interview. However, it shows restrictions for supervisors to select students among limited ones. Supervisors suffer from finding appropriate students due to asymmetric information.

In current studies, evidences have shown that the mismatch between students and supervisors has an effect on students’ academic achievements [1-3]. Various methods are applied to solve this problem, such as multiple criteria decision method [4], analytic hierarchy process method [5], analytic network process method [6], and generic algorithm [7]. However, they share a common limitation of merely focusing on the context of one university. They are not suitable to deal with vast online information in web 2.0 era. They do need a channel to comprehensively communicate before making the crucial decision.

This article proposes an approach making full use of the research social network to solve the problem of student-supervisor selection. It helps to bridge the gap between students and supervisors. Scholarmate is one of research social network websites and it promises the authenticity of all information of publications which are a typical feature to represent researchers. Firstly, it makes sure the authenticity of all confirmed publications by checking authors’ names, registered emails, research fields and so on. Simultaneously, users can self-claim the research fields by choosing the keywords among suggested standard ones. Secondly, every researcher has a structured personalized profile on the research social network. Thirdly, users can interactively communicate through social activities, such as liking others’ works, sharing articles, endorsing their research fields and so on.
Fourthly, it facilitates the communication between students and supervisors, which aims at enhancing the process of student-supervisor selection.

This paper develops an intelligent approach for providing supervisors with student recommendation services on research social network. It aims at providing a platform for communication in order to facilitate the process of student-selection. According to the proposed method, supervisors are clustered by k-means algorithm based on their own co-author diversity. Then, different group of supervisors receive student candidates by respective recommendation strategies. Finally, the well-known entropy method is employed to rank the candidates.

2. Literature Review

This section reviews the related works on the methods of student-supervisor selection, recommendation system applications in similar situations and the concepts of exploration and exploitation which are utilized in our proposed method.

2.1 Student-Supervisor Selection

In the existing literature, solid evidences have shown that the “style war” [8] will occur when the compatibility between a supervisor and a student is poor. Furthermore, the match/mismatch between students and supervisors has great influence on students’ achievements [1-3]. When students are accepted by the university, they are usually evaluated by several common parameters facilitating supervisors to make the decision. These parameters are GPA, undergraduate university, research interest, passed courses and specific skills [7]. The studies show that supervisors tend to work with the students who share high compatibility with them [7, 9]. Maedeh develops an advisory agents modeling system in University of Tehran to enhance the decision making of student-supervisor selection [9]. Additionally, researchers propose an approach based on genetic algorithm for student-supervisor assignment [7]. There are some other methods used to facilitating the process of student-supervisor selection, such as a strategic approach adapted from multiple criteria decision method [4], analytic hierarchy process method [5], and analytic network process method [6].

However, the extant methods merely pay attention to the context of one university. Moreover, another study finds that traditional methods result in some decisions which are not the best ones for both students and supervisors [10]. The underlying reason is the limited human rationale and asymmetry information. With the development of information technologies, it is a challenge to utilize social information to provide recommendation services to enhance the student-supervisor selection. We propose an intelligent approach which helps to consider various aspects for decision making and provides more communication opportunities for both students and supervisors.

2.2 Recommendation Systems in Similar Situations

Recommendation techniques are employed in various applications. It can be divided into item recommendation and people recommendation. On the one hand, the items can be movies, music, news, products [11] and so on. On the other hand, people-to-people recommendation attracts more and more attention in academic. Various articles study the collaborator recommendation [12] and expert recommendation for researchers. Meanwhile, some papers elucidate feasible approaches for the scenarios such as recruitment [13], online dating [14] and helper-finding [15]. This type of recommender system makes efforts to satisfy both parties and it is called reciprocal recommendation in academic [16].

In the prior studies, most of the studies provide a same recommendation strategy for different uses, which neglect users’ differences in characteristics. However, Hong and Zheng [17] design a job recommendation system which can automatically provide different recommendation strategies for different groups of users. Similarly, different kinds of supervisors may prefer different kinds of recommendation strategies. A good recommendation system should take users’ characteristics into consideration and select the personalized recommendation strategy for each one.
2.3 Exploration and Exploitation

The concepts of exploration and exploitation were initially introduced by March in 1991 [18]. It shows the ambidexterity of organizational learning orientations: exploration for new knowledge, skills and processes, while exploitation of existing ones [19]. Yang [20] applies the two concepts into alliance strategies and compares the impacts of exploration alliance and exploitation alliance with large firms. Similar with the studies on organization level, some studies on individual level show that the diversity of team members will influence the firm’s performance. They have been classified into newcomers, who come with new knowledge and always solve a problem in a novel perspective, and old-timers, who stick to make full use of and refine the existing knowledge to solve the current problems. Newcomers foster the exploration and old-timers enhance the exploitation [21].

In the context of co-author network, exploration behavior means preferring to collaborate with different researchers as much as possible and exploitation refers to cooperate with same scholars. Some researchers prefer to broaden the perspectives by collaborating with different scholars, while some researchers prefer to deepen the common understanding with the same collectives. Eldon [22] develops a co-author diversity index to measure the preferred collaboration style from the target researcher’s co-author network.

In summary, the student-supervisor relationship is regarded as one of the collaboration relation. They do also cooperate with each other for writing research papers. Therefore, the different preference of collaboration style should be taken into consideration when we design a recommendation approach. Most of current recommendation approaches are the same for all users. In this context, the authors intend to develop a method based on supervisor clustering primarily. Moreover, the extant methods focus on the context of one university. It is both a chance and a challenge to develop an intelligent approach and provide recommendation services on research social network.

3. Proposed Method

3.1 Overview of the Proposed Method

According to subjective expected utility theory (SEU), this article holds the idea that people make decision in uncertain circumstances based on the product of utility and subjective value probability \((\pi)\). The subjective value probabilities towards to different candidates depend on individual differences. In the context of student-supervisor selection, the major factor which should be considered is knowledge background. Some prefer to choose the student whose knowledge background is consistent with theirs. Therefore, they will show higher subjective value probability to the students whose backgrounds are harmonious with their research fields. Oppositely, some supervisors don’t pay so much attention on the factor of knowledge background. They are risk lovers and prefer to have a chance to collaborate with the students whose knowledge backgrounds are not so consistent with their research expertise. In light of this consideration, the method is proposed. Firstly, it mines the historical collaboration diversity of the target supervisor. Then, it distinguishes that which cluster the target supervisor should belong to. “Exploitation” group represents the supervisors who prefer to utilize their existing knowledge. Therefore, they prefer the students whose knowledge background is consistent with their expertise. It employs content-based method which focuses on this factor to provide recommendation services for this group. “Exploration” group shows the ones who intend to collaborate with somebody whose knowledge are different with theirs in some extent. They prefer to make use of new knowledge to innovate. It applies collaborative filtering method which pays not so much attention on the consistent knowledge background but pays enough attention on other factors to obtain candidate students. Finally, the hybrid method will be applied to “moderation” group.

Firstly, the data will be prepared and processed. The co-author information will be extracted from the target supervisor’s publications. Secondly, the supervisors will be classified into three different groups by simple K-means algorithm [23] based on the co-author diversity index. The three groups
are defined as “exploration group”, “exploitation group” and “moderation group” respectively. Thirdly, different recommendation strategies will be employed for different groups. Fourthly, the candidates will be measured by several criteria and the final ranking list will be formed. Finally, the top-n recommendation results will be generated based on the ranking list. And the feedback of final results will be compared in the end.

3.2 Clustering Process

According to the target supervisor’s research publications, we can extract the information of his co-authorship network. Then, the collaboration diversity index (CDI) [22] is used to measure the supervisor’s preferred style of collaboration. Based on the value of this index, supervisors will be grouped into different clusters by k-means algorithm.

$$CDI(a_i) = \frac{\sum_{j=1}^{n} coauthor_j - \sum_{j=1}^{n} duplicate_j}{\sum_{j=1}^{n} relation_j}$$  \hspace{1cm} (1)

Where, \(a_i\) represents the \(i^{th}\) author. \(j\) is the number of articles. \(\sum_{j=1}^{n} relation_j\) stands for total collaborative relations of all publications. Meanwhile, \(\sum_{j=1}^{n} coauthor_j - \sum_{j=1}^{n} duplicate_j\) refers to the total of unique co-authors. If the value of CDI is equal to 1, it shows that the author is extremely explorative. Meanwhile, it is completely exploitative if the value is near zero.

3.3 Recommendation Strategies

The different recommendation strategies will be applied to the supervisors in different groups. There are three recommendation approaches: content-based recommendation (CB), collaborative filtering recommendation (CF), and hybrid recommendation (Hyb). The principle of CB is recommending items which are similar with the given item in the content information [24]. In contrast, CF method which is regarded as user-to-user correlation method, searches similar users who have similar taste with the target user firstly. Then it recommends the items which similar users like to the given user [25]. Additionally, the hybrid method is a combination of CB and CF in order to overcome the drawbacks of the former two methods. And the detailed algorithms in specific approaches are showed respectively as following.

3.3.1 Content-Based Recommendation Approach

Student’s information includes the student’s major, taken courses, references which he prefers, and publications (optional). In the meantime, the information of a supervisor embraces his department, taught courses, publications and funded projects. All the relevant information will be arranged into the student’s and the supervisor’s documents respectively. The keywords will be extracted from the student’s document and all the keywords will formed as a query (\(Q_{(i,j)}\)). Therefore, the similarity between a student and a supervisor will be calculated by the BM25 algorithm as follows. BM25 is widely used in information retrieval. It is utilized by search engines to rank documents based on the relevance to a given query.

$$FScore_{cb}(Q_{(i,j)}, D_{(i,j)}) = \sum_{k} W_k R(q_k, D_{(i,j)})$$  \hspace{1cm} (2)

Where,

$$W_k = IDF(q_k) = \log \frac{N - n(q_k) + 0.5}{n(q_k) + 0.5}$$  \hspace{1cm} (3)

$$R(q_k, D_{(i,j)}) = \frac{Df_k \cdot (m_1 + 1) \cdot Qf_k \cdot (m_2 + 1)}{Df_k + M \cdot Qf_k + m_2}$$  \hspace{1cm} (4)

$$M = m_1 \cdot (1 - b + b \cdot \frac{DI}{avgDI})$$  \hspace{1cm} (5)
represents the corresponding document of a supervisor. A query contains keywords \( q_1, q_2, \ldots, q_n \). \( n \) is the number of keywords. \( w_i \) is the weight of the \( k^{th} \) query. It is measured by inverse document frequency (\( IDF(q_i) \)). \( N \) is the total number of supervisors’ documents. \( n(q_i) \) is the number of documents which contain keyword \( q_i \). \( D_l \) is the length of the document. \( \text{avg}D_l \) is the average document length. \( m_1, m_2 \), and \( \beta \) are free parameters.

### 3.3.2 Collaborative Filtering Recommendation Approach

It firstly finds the supervisors who are similar with the target supervisor. According their interaction on the research social network, we trace the students which the candidate supervisors prefer.

In order to find the similar supervisors, it will calculate similarity among supervisors according to the information of their department, taught courses, publications and projects. The calculation process is similar with that in the content-based method. The difference is that the similarity here is between the target supervisor and other supervisors.

\[
\text{Score}(Q_{t(i)}, D_{n(j)}) = \sum_k^n W_k R(q_k, D_{n(i)})
\]

After obtaining a set of similar supervisors, it traces their preferred students according to the interactions on research social network, such as the behavior of like, share, and endorse. By the following formula, we can get a list of candidate students.

\[
\text{Score}(S_{s(i)}, S_{t(j)}) = \sum_m^3 n_w(s(i), t(j)) \quad m = 1, 2, 3 \quad (1 - \text{endorse}, 2 - \text{share}, 3 - \text{like})
\]

\[
F\text{Score}_c(s(i), t(j)) = \alpha \text{Score}(Q_{t(i)}, D_{n(j)}) + \beta \text{Score}(S_{s(i)}, S_{t(j)})
\]

### 3.3.3 Hybrid Recommendation Approach

The hybrid approach is a combination of the CB method and the CF method. On the one hand, the candidate students are traced based on the similarity with the target supervisor by CB method. On the other hand, CF method is employed to seek out similar supervisors. According to the social interactions, the preferred students of the similar supervisors can be found. Therefore, two candidate sets will be generated according to the CB method and CF method. We utilize CombMNZ, a score-based fusion method, to aggregate the two candidate lists.

\[
F\text{Score}_{hyb}(s(i), t(j)) = \tau (\text{list1}(i), \text{list2}(i)) \ast \text{SUM}(\text{Score}_{cb}(Q_{t(i)}, D_{n(j)}) + \text{F}\text{Score}_c(s(i), t(j)))
\]

If \( \text{list1}(i) \) and \( \text{list2}(i) \) are both more than zero, \( \tau \) equals to 2; if only one of them is more than zero, \( \tau \) equals to 1; otherwise \( \tau \) equals to 0.

### 3.4 Ranking Process

After the filtering stage, the candidate students will be obtained. The candidate students will be ranked by the measurements in the ranking stage. The student’s GPA, publication (if applicable), and institution will be taken into consideration.

Firstly, GPA is a major factor which is widely used to measure a student’s school performance. It represents students’ long-term and persistent efforts, the degree of knowledge acquisition and ability. However, we cannot deny that grading standards vary across different universities. Therefore, GPA should be transferred into a same form. It is always in the form of \( \triangleleft x_i / y_i \). To facilitate the further calculation, it will be changed into the form of \( \triangleleft z_i / 4.0 \) by the following expression.

\[
z_i = 4x_i / y_i
\]
Where, \( \langle a_i, a_2, a_3, a_4, a_5 \rangle \) is a decreasing function. \( \langle m, m_2, m_3, m_4 \rangle \) is a decreasing function.

Secondly, most of Master programs and Ph.D. programs are research-oriented and share the goal of qualified and plentiful scientific productivity. Research productivity can be indicated by the quality and quantity of published journal papers and conference papers which a student has produced. But, we should admit that the majority of new research students have no publications and a few have published several papers. Therefore, the initial value is set as 1 to facilitate the calculation process. Furthermore, we set the weight of SCI/SSCI journal higher than that of other kinds of journals.

\[
(PGA) = \begin{cases} 
\alpha_1 & \text{if } m_1 \leq z_1 \\
\alpha_2 & \text{if } m_2 \leq z_1 \leq m_1 \\
\alpha_3 & \text{if } m_3 \leq z_1 \leq m_2 \\
\alpha_4 & \text{if } m_4 \leq z_1 \leq m_3 \\
\alpha_5 & \text{if other}
\end{cases}
\]

(11)

Where, \( \langle \alpha_i, \alpha_2, \alpha_3, \alpha_4, \alpha_5 \rangle \) is a decreasing function.

In the ranking stage, the candidate students are ranked by the well-known entropy method. It gets a decision matrix \((A)\) representing respective scores from three aspects of every candidate student.

\[
\begin{pmatrix}
a_{11} & a_{12} & a_{13} \\
a_{21} & a_{22} & a_{23} \\
\vdots & \vdots & \vdots \\
a_{n1} & a_{n2} & a_{n3}
\end{pmatrix}
\]

(15)

Where, \( n \) represents the number of candidates. \( a_{ij} \) is candidate \( i \) ’s value of \( j^{th} \) aspect.

Firstly, the data is normalized, and matrix \((B)\) is got. Subsequently, the entropy value will be calculated by expression (17).

\[
E_j = - \left( \sum_{i=1}^{n} b_{ij} \ln b_{ij} \right) / \ln n
\]

(17)

Then, let \( d_j = 1 - E_j \) \((1 \leq j \leq m)\). Where \( m \) means the number of criteria. \((m = 3, \text{here})\). Normalize it and we obtain the entropy weight of the \( j^{th} \) attribute through formula (18). Consequently, the final ranking score will be calculated by formula (19).
\[ w_j = d_j / \sum_{j=1}^{m} d_j \]  
\[ RS_i = w_u \sum_{j=1}^{m} p_{ij} w_j \]  

(18)  
(19)

4. Future Work: Evaluation

The data can be collected on ScholarMate. In order to evaluate the proposed method, it has to do an offline user survey of satisfaction. It will evaluate the performance of different recommendation strategies by supervisors’ satisfaction towards the top-n recommended students. By comparing the satisfaction result of three different methods for a specific group accordingly, it can verify the assumptions. Whether the CB method is most suitable for the exploitation group, the CF method is most suitable for the exploration group, and the hybrid method is most suitable for the moderation group?

5. Conclusion

In this paper, an intelligent approach is proposed to recommend students for supervisors on research social network which enhances communication process between students and supervisors. Firstly, the target supervisor’s preference is mined from his diversity of collaborators. And it classifies him into different groups: exploration group, exploitation group and moderation group. Secondly, different recommendation strategies are applied to do recommendation for the target users in different groups. For the exploitation group, the content-based method is employed. For the exploration group, the collaborative filtering method is utilized. And the hybrid method is applied into the moderation group. Furthermore, the social activities are utilized in the proposed method as well. In a nutshell, it puts forward an approach to provide student recommendation service for supervisors on research social network, which plays the crucial role of bridging the gap between students and supervisors.

Certainly, there are several limitations in our study. Firstly, it didn’t consider the semantic meanings of keywords when we calculate the similarity score. In the future, it will pay attention to this part and modify our algorithms. Secondly, the evaluation is not completed by now. It will evaluate the performance of the proposed method by users’ satisfaction towards the recommended candidates. Thirdly, ScholarMate is a regional research social network platform and it will consider to do further experiments on other scientific communities in order to verify the generalization of the proposed method.

Acknowledgements

This paper is sponsored by the project from Fujian Provincial Education Department (No.JAS180414) and a School-level Research Project of Putian University (No.2018061).

References


