A personalized information recommendation system for R&D project opportunity finding in big data contexts

Wei Xu [a], Jianshan Sun [b,*], Jian Ma [c], Wei Du [c]

[a] School of Information, Renmin University of China, Beijing, PR China
[b] School of Management, Hefei University of Technology, Hefei, Anhui 230009, PR China
[c] Department of Information Systems, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong

ARTICLE INFO

ABSTRACT

With the rapid proliferation of online information, how to find useful information, such as suitable jobs, appropriate experts, and proper projects, is really an important problem. Recommendation technique, as one of emerging tools to deal with information overload and information asymmetry, is critically important for providing personalized online information services. With the increase of R&D investment in government and industry, such as high-tech companies and advanced manufacturing enterprises, more and more R&D project information are launched in public websites for cooperation. When the number of online information and users is extremely huge, how to effectively recommend R&D project opportunities to related researchers and practitioners is a challenging and complex task. In this paper, a novel two-stage method is proposed for R&D project opportunity recommendation. An information filtering method is first offered to identity proper R&D projects as a candidate set. Then, an information aggregation model with various constraints is suggested to recommend appropriate R&D projects for applicants. The proposed method has been implemented in an online research community – ScholarMate (www.scholarmate.com). An online user study has been conducted and the evaluation results exhibit that the proposed method is more effective than existing ones.

1. Introduction

With the rapid development of the Internet and information technologies, the amount of online information is growing in an explosive speed. Users are flooded in the sea of too much information and are struggling to make good decisions, which refer to information overload problem (Aljukhadar et al., 2012). Furthermore, users are often unaware of useful opportunistic information (such as job opportunity and funding opportunity) due to information asymmetry (Chen et al., 2013). In industrial environment, the continuous R&D investments have brought out more and more R&D project opportunities. For example, some high-tech companies and advanced manufacturing enterprises have posted difficult issues on the public websites to call for cooperation. However, relevant researchers and practitioners have no channels to find the useful project opportunity information. There is a dilemma between two parties (information publisher and information searcher). To attach this critical challenge, we propose a two-stage recommendation approach for making personalized R&D project opportunity information for online users.

The recommendation system, as a personalization tool, is a computer-based system that matches potential goods and information with user preferences (Adomavicius and Tuzhilin, 2005). The system can reduce the search efforts of users and mitigate the encumbrance of information overload (Aljukhadar et al., 2012).

There are several works in personalized information recommendation services. In the E-recruitment platform, Al-Otaibi and Ykhlef (2012) built recommender systems to provide job information for jobhunters. In the research community, Balog et al. (2009) proposed language models to find appropriate experts which could help users solve tricky problems. In tourism, a personalized route recommendation service was designed for theme parks and RFID information and tourist behaviors were explored to support decisions (Tsai and Chung, 2012). In mobile environment, Park et al. (2008) employed probabilistic multi-criteria decision making methods to provide restaurant recommendations for group of users. Generally, there are two types of research methods for these information recommendations: content-based methods and network-based methods (Liang et al., 2008). Content-based methods recommend useful information services that are similar to those in which the user has shown interest in the past. Network-based methods recommend useful information services to the user based on other like-minded individuals who have similar preferences or tastes as the target user. Since content-based methods

* Corresponding author. Tel.: +86 13265835100; fax: +86 55162904991.
  E-mail address: sunjs9413@gmail.com (J. Sun).

and network-based method had their own advantages and disadvantages, a hybrid method is needed to combine them for better recommendation performance. Furthermore, both of these approaches have the scalability and efficiency problems in the big data context where huge amount of information is involved. Big data analytics tools should be employed and leveraged to support more intelligent information recommendation services. Also, existing literatures did not consider contextual factors of online users for information recommendations while these factors often had crucial effects on the decision making.

To remedy the dilemma of R&D project cooperation and address issues from the literatures of information recommendation services, we proposed a novel two-stage recommendation method to help researchers and practitioners find proper R&D project opportunities launched by governments and enterprises. In the first stage, a candidate set of R&D project opportunities is identified by the information filtering method. In the second stage, an information aggregation model with various constraints is designed to recommend most suitable R&D project opportunities for applicants. The designed R&D information recommendation system has been implemented in the research social network website – ScholarMate.[1] A real online survey has been conducted to verify the effectiveness of our recommendation method. The results show that the proposed method outperforms the baseline methods in terms of various recommendation accuracy metrics.

The rest of the paper is organized as follows. In Section 2, we review the related work on information recommendation services. In Section 3, we propose the two-stage approach for R&D project opportunity recommendation. The proposed recommender system is developed and implemented in Section 4. Section 5 introduces evaluation metrics and experimental study. Finally, Section 6 concludes the paper and points out further research directions.

2. Literature review

Information recommendation services have attracted much attention in recent years since amount of web information appear in the Internet, which online users cannot inquire easily. They have been applied into various contexts such as job search (Al-Otaibi and Ykhlef, 2012; Malinowski et al., 2006; Paparrizos et al., 2011), expert finding (Balog et al., 2009; Cao et al., 2005; Deng et al., 2008, 2012; Fang and Zhai, 2007), travel planning (Staab and Werthner, 2002; Tsai and Chung, 2012; Yu and Chang, 2009), and restaurant recommendation (Lee et al., 2006; Mui et al., 2001; Park et al., 2008). Generally, two types of research methods have been suggested for information recommendation service: one is content-based methods, by using text mining techniques and from the relevance perspective; the other is network-based methods, by using graph theory and from the connectivity perspective.

The type of content-based methods could be further classified into two categories: profile-based methods and document-based methods (Fang and Zhai, 2007). Profile-based methods directly built online user profile based on associated documents and then generated the ranking score according to the profile in response to a user query. For example, Balog et al. (2009) first built a term-based expertise profile (virtual document vector) for each candidate, and ranked the candidate experts based on the relevance scores of their profiles for a given topic by using traditional ad hoc retrieval models. On the other hand, document-based methods first ranked documents in the corpus given a query topic. Then they found the associated candidates from the subset of retrieved

documents. Furthermore, Cao et al. (2005) used a probabilistic approach to rank experts by combining relevance model and co-occurrence model. The topic-based model and the hybrid model were exploited by Deng et al. (2008) to achieve better performance than basic language model. Macdonald and Ounis (2008) presented yet another approach based on a voting model for expert search. Profile-based methods operated efficiently due to smaller documents in size for modeling while they performed less effectively than other approaches because they could not measure each document individually. Document-based methods allowed the application of advanced text modeling techniques in ranking individual documents while they showed inconvenient data management. With the rapid development of cloud computing (Armbrust et al., 2010; Vouk, 2004) and big data analytics (LaValle et al., 2011; Zikopoulos and Eaton, 2011), some useful tools (such as Hadoop) to speed up computational performance have been proposed. In this paper, we employ MapReduce platform to improve the efficiency of content-based methods for information recommendation service.

The type of network-based methods benefited from the success of PageRank (Page et al., 1999) and hyperlink-induced topic search (HITS) (Kleinberg, 1999) algorithms in search engines. Paparrizos et al. (2011) applied a directly graph to construct a job transition graph, and then used a machine learning method to recommend jobs. Zhou et al. (2007) proposed a coupled random walk model between authorship networks and citation networks for ranking authors and documents together. Recently, heterogeneous bibliographic networks were modeled and exploited for expertise ranking (Deng et al., 2012). Although graph-based techniques above were advanced and sophisticated for information recommendation service, their success is mainly dependent on the accuracy of the constructed networks. In this research, we offer multi-level social networks including personal-level network and institutional-level network and also employ MapReduce platform to speed up the calculation of the similarities between the nodes in networks.

Meanwhile, some researchers have argued that it is not enough to recommend information by only looking at the queries' without taking the online users into consideration, and several contextual factors (e.g. information quality) may have effects on the decision concerning (Hofmann et al., 2010). So, several combined approaches in which many of these factors considered in the modeling process have been proposed to improve the quality of information recommendation (Ganu et al., 2013; Weng et al., 2009; Zhang et al., 2012). For example, Zhang et al. (2012) combined content and relation analysis for recommendation in social tagging systems. In addition, existing studies show that topic relevance, social connectivity and expertise level are all valuable information sources for information recommendation (Hofmann et al., 2010). However, very limited research combines them to give a better solution to the problem. In this research, we define a research analytics framework (RAF) by combining quality, relevance, and connectivity for information recommendation.

Moreover, various information recommendation services, such as job search, expert finding, and travel planning, have been developed, and some of them have been implemented in various platforms, such as ScholarMate (www.scholarmate.com), an online research community. However, with the increase of R&D investment in government and industry, such as high-tech companies and advanced manufacturing enterprises, more and more R&D project information are launched in public websites for cooperation. How to recommend R&D project opportunities to related researchers and practitioners is yet not well studied in the existing literatures. Furthermore, R&D project opportunity recommendation has some new challenges, comparing with other information recommendation service. For example, they may have some

---

[1] www.scholarmate.com.