A Profile-Boosted Research Analytics Framework to Recommend Journals for Manuscripts

Thushari Silva and Jian Ma
Department of Information Systems, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong.
E-mail: {tpsilva2, isjian}@cityu.edu.hk

Chen Yang and Haidan Liang
School of Management, University of Science and Technology of China, Hefei 23000, PR China, and
Department of Information Systems, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong.
E-mail: yangc0201@gmail.com, haidan1@qq.com

With the increasing pressure on researchers to produce scientifically rigorous and relevant research, researchers need to find suitable publication outlets with the highest value and visibility for their manuscripts. Traditional approaches for discovering publication outlets mainly focus on manually matching research relevance in terms of keywords as well as comparing journal qualities, but other research-relevant information such as social connections, publication rewards, and productivity of authors are largely ignored. To assist in identifying effective publication outlets and to support effective journal recommendations for manuscripts, a three-dimensional profile-boosted research analytics framework (RAF) that holistically considers relevance, connectivity, and productivity is proposed. To demonstrate the usability of the proposed framework, a prototype system was implemented using the ScholarMate research social network platform. Evaluation results show that the proposed RAF-based approach outperforms traditional recommendation techniques that can be applied to journal recommendations in terms of quality and performance. This research is the first attempt to provide an integrated framework for effective recommendation in the context of scientific item recommendation.

Introduction

Apart from advancing knowledge by presenting valid and acceptable innovative research outcomes, getting things published is an important task as it opens ways for the occupational success of academics (Albers, Floyd, Fuhrmann, & Martínez, 2011; Nihalani & Mayrath, 2008). At the same time, journal citation counts and the number of publications in refereed journals are widely accepted measurements of researcher and institution rankings. Thus, there is a steadily growing demand for publishing in refereed journals—the "publish or perish" phenomenon (Nihalani & Mayrath, 2008).

Researchers in all fields face difficulty finding appropriate journals in terms of relevance and quality for their manuscripts. *Research relevance* refers to the exact alignment of topics covered by journals with the manuscript contents, and *quality* refers to academic rewards generated by publications. This is especially the case for young researchers who are impatient and averse to taking risks, for example, research students and untenured professors (Heintzelman & Nocetti, 2009). The choice of submission is particularly difficult given the large number of potential journals. Gregory and Mankiv have proposed an intuitive theory of submission: "develop a submission tree, aiming at the top journals in the first submission, and if rejected, follow the path established" (Heintzelman & Nocetti, 2009, p. 1). Although this is a simple strategy and easy to follow, it makes the submission process more complicated. For example, assume that there are 15 significant journals in the information systems (IS) discipline in which many senior scholars have published. Then we have 1,307,674,368,000 (15!) possible paths. Generally, in the field of IS there are about 135 *Science Citation Index* (*SCI*) indexed journals. Therefore, practical application of this theory in is ineffective, especially for impatient young

Received January 10, 2013; revised September 14, 2013; accepted September 30, 2013

 • Published online 7 May 2014 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/asi.23150

researchers. Thus, understanding the manuscript submission process and recommending journals for given manuscripts is important for researchers trying to increase their productivity as well as journal managers (e.g., editors), as it affects the efficiency of scholarly research production. However, identification of suitable and relevant publication outlets with the highest value and visibility for a research article/manuscript is more challenging due to the lack of a scientific way to determine the relevance of an article to a field and its appropriateness for a journal (VandenBos et al., 2010).

A manuscript can be characterized by a set of qualitative and quantitative, tangible and intangible attributes. Although there have been several attempts by computer scientists, management scientists, IS practitioners in the area of personalized recommendation of research articles, books, news articles, and webpages (Chen, Chen, & Sun, 2001; Watanabe, Ito, Ozono, & Shintani, 2005), almost none is presented in the journal recommendation literature. Economists have proposed a two-player game model (Faria, 2005) as well as economic models (Heintzelman & Nocetti, 2009) to model the process of manuscript submission from different angles. Economic models that maximize an author's payoff via balancing the consideration of journal quality and acceptance probabilities are useful for identifying journals but fail to capture the relatedness of the manuscript to the journal.

The problem of recommending scientific items (e.g., research articles, journals, research collaborators) for researchers may differ significantly from the traditional recommendation problem that involves consumer goods or movies in terms of how preferences for recommended items are distributed and what the researcher's main informational needs are (Im & Hars, 2007). Traditional collaborative filtering recommendation techniques are useful when suggesting journals for researchers based on other similar researchers' choice of journals, but fail to take into account the characteristics of the current manuscript.

Content-based recommendation approaches in similar applications such as article recommendation mainly consider matching research relevance in terms of keywords or disciplines while ignoring the social connections (e.g., collaboration and co-authorship) and productivity (e.g., quality, quantity, and citations of published journal articles) of users. Thus, it is desirable to incorporate all these aspects into a unified framework. To achieve this goal, this research work proposes a novel three-dimensional research analytics framework (RAF) that holistically considers relevance, connectivity, and productivity aspects for effective journal recommendation. The relevance dimension considers the research relevance of manuscript and the potential journals; the productivity dimension considers the level of matching between author's quality and quality of the recommended journals. Better judgment on the quality of the authors of the manuscript as well as the publication outlets will help to identify journals with more career rewards and a higher possibility of acceptance for a given manuscript. The connectivity dimension analyzes the research social network of the authors of the manuscript and identifies potential journals in which similar researchers have published. Better identification of social connections can effectively cluster researchers based on topics of interests. Once the communities of researchers with similar interests are identified, it should be easy to identify publication outlets that relevant researchers have published in by analyzing researcher-topics and journal-researcher two mode networks. Most research in the literature only focus on either one type of subjects (e.g., researchers) or single relationship (co-authorship) among them. In this study we construct two researcher-topics and researcher-journal two-mode networks and perform joint analysis for better publication outlet identification.

The research question that this study addresses is how to determine the relevance and appropriateness of a research article to publication outlets. To achieve this objective, profiles of research entities (e.g., journals and manuscript) are constructed by integrating three different types of information sources, that is, subjective information, objective information, and social information. For example, when constructing journal profiles, information from the journal homepage (i.e., subjective), published articles in the journal (i.e., objective information), and opinion of scholars, for example, PhD supervisors (i.e., social information), are aggregated. Once the profiles are constructed, a unique matching algorithm based on RAF is constructed to match the profiles of manuscript with journal profiles.

In summary, this research makes the following important contributions: First, a novel three-dimensional research analytics framework that holistically considers relevance, productivity, and connectivity to provide effective journal recommendation is developed. Second, this research is the first attempt to employ two-mode research collaboration network analysis in the research recommendation environment.

The rest of the paper is organized as follows: The next section reviews the relevant literature. The Proposed Method section provides the novel three-dimensional aspects, which is referred to as the RAF, together with journal and manuscript profiling. The multi-objective decision model for journal recommendation is presented in the Recommending Journals for Manuscript section. The Implementation and Evaluation section describes the system and the predictive accuracy of the proposed approach, comparing it with a few benchmark models.

Literature Review

Most scientific recommender systems adopt twofold paradigms either recommending a user to an object or an

DOI: 10.1002/asi

object to a user (e.g., reviewer to a proposal or article to a reviewer) (Adomavicius, Tuzhilin, & Zheng, 2011). Examples for such scientific recommender systems are reviewer/expert assignment system, citation recommendation, and article recommendation. In this section, we also review the literature related to scientific recommendation, as journal recommendation is one of the scientific recommendation applications.

Reviewer assignment focuses on assigning relevant reviewers for suitable manuscript/proposal evaluation while article recommendation is focused at recommending related articles for researchers. Citation recommendation concentrates on suggesting similar citations by considering content of the main text (McNee, 2006).

Information retrieval-based approaches proposed for reviewer and article recommendation as well as citation recommendation can be classified into content-based approaches, collaborative filtering approaches, and hybrid approaches (Wang, 2010).

Content-Based Approaches in Scientific Recommendation

Content-based approaches in reviewer recommendation focus on matching research relatedness in terms of keywords and use multiple techniques including Latent Semantic Indexing (LSI) (Manning, Raghavan, & Schütze, 2008), the Vector Space Model (VSM) (Biswas & Hasan, 2007), and data mining (Hettich & Pazzani, 2006). In LSI, reviewer expertise is represented as a reviewer-keyword matrix in which elements represent factor weights computed by automatically using keyword frequency. Similarly, an article/proposal-keyword matrix is computed. The dot product of the two matrices is used to determine the matching degree between reviewer and article/proposals. In VSM, proposals are represented as a bag-of-words occurring in the proposal/article. The number of proposals determines the dimension of the vector space. A reviewer's expertise is also represented as a vector in this space assuming that they share the same vocabulary. The angle between a reviewer expertise vector and a proposal/article vector that can be captured by the cosine is used to judge the similarity of the proposal/article to the reviewer. Liang et al. (2008) have applied the spread activation model to match keywords via semantic network construction in the article recommendation. The VSM is a lexical matching technique that assumes keywords are independent but it is not a realistic assumption, and hence suffers from keyword ambiguity problems and keyword sparsity problems that lead to a match irrelevant problem. The LSI overcomes problems inherent in lexical matching by taking into account the latent semantic structure in word usage. However, the serendipity issue, the problem of discovering new expertise, is the major bottleneck for LSI in reviewer assignment.

Collaborative Filtering Approaches in Scientific Recommendation

Traditional collaborative filtering techniques used in scientific recommender systems are accurate when suggesting items to target users based on what other similar users have previously preferred, but less effective when there are not enough ratings from others. These have been used in article, citation, and reviewer recommendation. The scale-free network approach (Watanabe et al., 2005) is one of the collaborative filtering approaches used in reviewer recommendation. It constructs a weighted keyword map based on a reviewer's publications and calculates matching degrees between articles/proposals and reviewers based on the sum of the weights of the keywords appearing in them. It ignores the relatedness of the content of the article/proposal to reviewer and thus suffers from match irrelevant problem i.e. finding unrelated reviewers. To overcome the "data sparsity" problem commonly appearing in content-based approaches, Vellino (2010) proposed usage-based and citation-based methods for recommending research articles. McNee et al. (2002) recommend citation for research papers following three different rating matrices: author-citation, paper-citation, and co-citation matrix.

Hybrid Approaches in Scientific Recommendation

The hybrid approach, which combines content-based and collaborative filtering approaches, has performed better (He, Kifer, Pei, Mitra, & Giles, 2011). Hwang and Chuang (2004) combined article content and web usage information for literature recommendation in a digital library context. He et al. (2011) combined the language model with citation analysis to recommend citations for research papers.

Furthermore, with the proliferation of factors that influence the recommendation process especially within a contextually rich environment, it has urged multidimensional recommender systems that consider all other possible constraints or relevant external expertise rather than only "user" and the "object" to increase the accuracy of the recommendation process.

User Profiling

The major challenge in recommending publication outlets for a manuscript is identifying key areas and themes covered by the manuscript as well as the potential journals. In this study, we propose a profile-boosted approach to recommend suitable publication outlets for a given research article. We construct journal profiles and article profiles integrating three different types of information sources: subjective, objective, and social information. Although it is rare in the literature to work that focuses on

DOI: 10.1002/asi

how to conduct article profiling or journal profiling, there are many studies that focus on researcher profiling. Among other approaches proposed for researcher profiling, two approaches are significant. One relies on subjective self-claimed information declared by the researchers themselves. The other is based on objective measurement obtained through automated inferences about the researcher's behavior patterns related to publications and citations derived from relevant resources (Vivacqua, Oliveira, & De Souza, 2009). The first approach uses qualitative methods (e.g., surveys, questionnaires, or interviews) and traditional information retrieval models, for example, term-based modeling (Joachims, 2001) and rough-set modeling (Li, Zhang, & Swan, 2000) to gain knowledge of a researcher's interests and resultant profiles. The latter approach uses various feature selection techniques from machine learning to learn a user profile (Fan, Gordon, & Pathak, 2005). We employ an amalgamation of these two approaches for article and journal profiling.

The machine learning approaches used in user profiling tend to learn the mapping between the incoming set of documents relevant to user input and real numbers which represent the strength of user preferences. The features of the documents are first extracted by widely used techniques including information gain (Mitchell, 1997; Yang & Pedersen, 1997) and correlation coefficient (Strzalkowski, 1994). Then the key features are used as attributes in the mapping functions. Some studies focus on techniques such as neural networks (Mostafa & Lam, 2000), Support Vector Machine (SVM) (Joachims, 2001; Robertson & Soboroff, 2002), KNearest Neighbors (K-NN), and logistic regression (Caulkins, Ding, Duncan, Krishnan, & Nyberg, 2006; Zheng, Chen, Sun, & Zha, 2007) before generating a mapping with a set of real numbers. Li, Zhou, Bruza, Xu, and Lau (2012) proposed a Rough Threshold Model (RTM) to analyze and extract keywords from the scientific publications. In our approach, we augment the original RTM with a phrase analysis algorithm to resolve semantic ambiguity that is not handled by the original rough threshold model for topic generation.

Manuscript Submission

Given the importance of manuscript submission to a journal, extensive research has been carried out in the field of economics and management related to manuscript submission. A framework for benchmarking journals from a submission author's point of view has been proposed by Bjork and Holmstrom (2006). It consists of eight main factors that could influence an author's submission decision and 21 other underlying factors. This framework has been used by Brochner and Bjork (2008, p. 1) for their empirical investigation that aims at "analyzing how an author's choice of journals in construction management are affected by quality and service perceptions." This study adopts some of the criterion proposed in the framework for benchmarking journals. Heintzelman and Nocetti (2009) analyzed the problems faced by impatient researchers when selecting suitable publication outlets for their submission. The authors justified that impatient risk-averse young researchers deviate from the traditional principle for journal submission that is proposed by Gregory and Mankiv. An author-editor two-player game model has been proposed for journal submission by Faria (2005). This model assumes that, authors need to maximize the number of publications targeting high academic impact and editors need to maximize quality of the papers that they publish to achieve high academic reputation. Following the existing literature, the main influential criteria for journal selection including impact factor, acceptance rate, and submission delay have been used to judge the quality of the journal.

Proposed Method: Profile-Boosted RAF

Figure 1 illustrates the proposed model for journal recommendation based on RAF.

Three different key performance indicators (KPIs), namely, relevance index, productivity index, and connectivity index, are constructed to measure the strength of each dimension of the RAF. The relevance index calculates the research relatedness of a manuscript to a journal in terms of discipline and keywords. The productivity index measures the quality, quantity, citations, and impact of the author's research. The connectivity index determines the popularity of the journals (i.e., widely used journals by similar experts) to be recommended. Finally, a unique matching algorithm based on the three KPIs is developed to achieve optimal assignment of journals to manuscripts.

Figure 2 illustrates the detailed process models in journal recommendation and how the three different indexes can be integrated to determine the most relevant and highly productive and widely accepted journals (i.e., a journal that satisfies the author's needs, such as a journal with high impact factor, high acceptance rate, Institute for Scientific Information [ISI] indexed journal, etc.). Initially, the system constructs the profiles of the articles and journals. Article profiling and journal profiling are discussed in detail in the Profiling section. As the next step, the connectivity index (c_{ij}), which is defined as a decision parameter used to determine whether collaborators who share common knowledge as the authors of the article *i* have published similar contents in journal *j,* is derived by analyzing the collaboration network properties including network structure and closeness of neighbors. The collaboration network is constructed by using the ScholarMate platform services developed by the authors of this paper. As the third step, the productivity index (e_a) of a manuscript is generated by considering the quality of the

DOI: 10.1002/asi

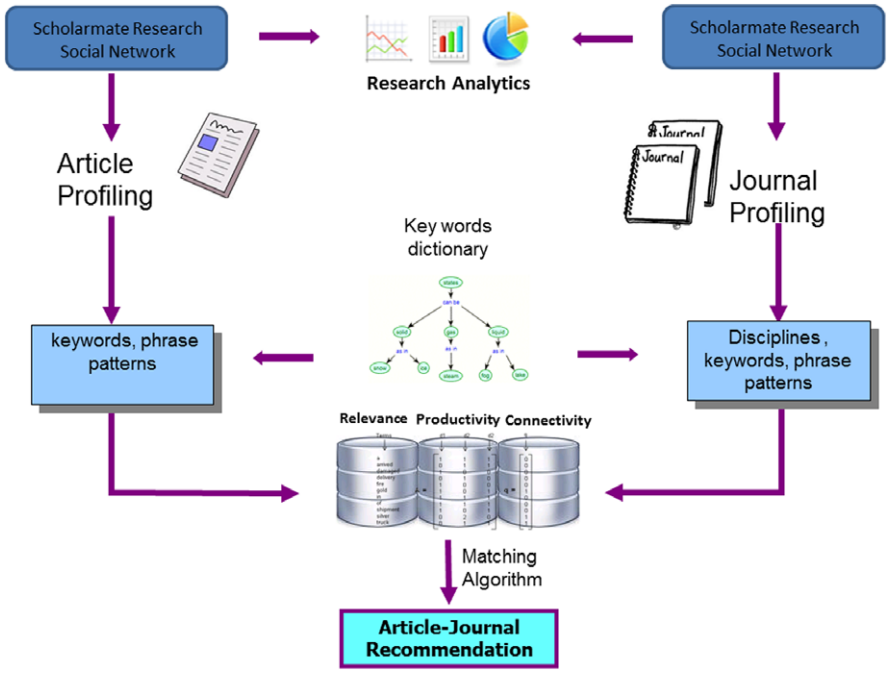


FIG. 1. The framework of a profile-based manuscript–journal recommendation. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

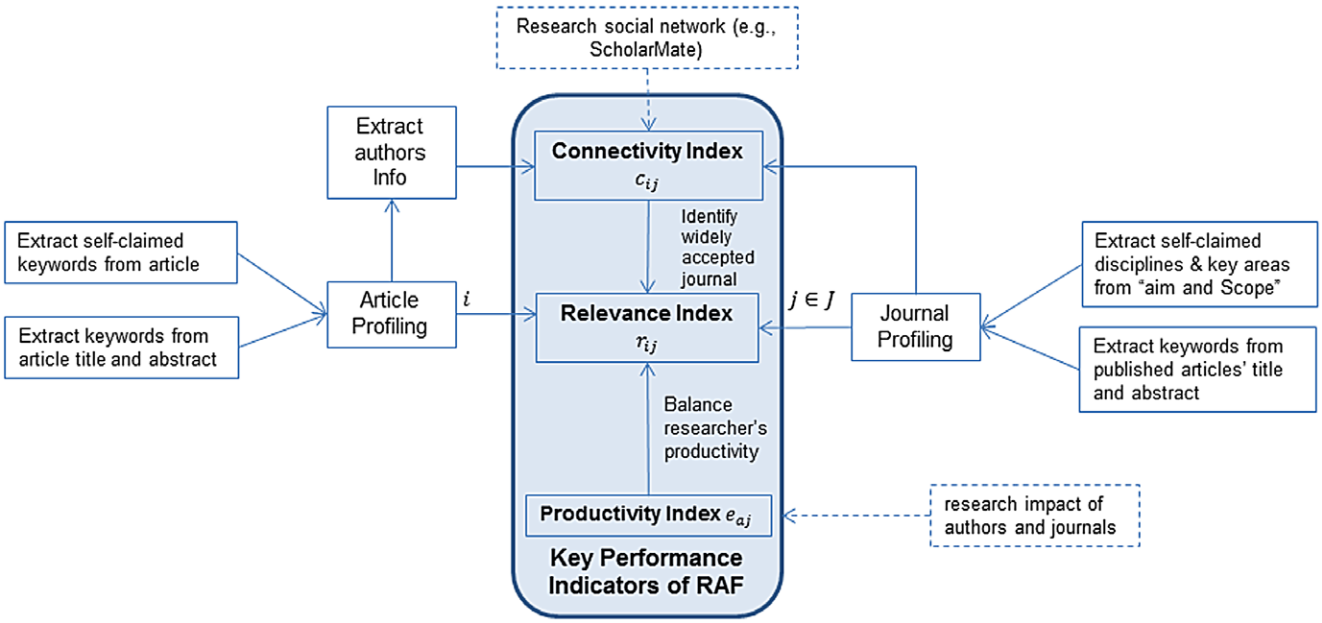


FIG. 2. The process models and measuring indexes in article–journal recommendations. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

publications made by its authors, their (the authors) citation impact, and academic achievements. A component-based matching algorithm is developed to calculate the relevance index (r_{ij}) which denotes the degree of matching between the i^{th} manuscript profile and the j^{th} journal profile. Detailed descriptions on the calculation of the connectivity index and the productivity index are presented in their respective sections.

DOI: 10.1002/asi

In summary, once r_{ij}, the matching between manuscript i and journal j, is calculated, c_{ij} is used to figure out whether researchers with similar interests as the authors have published in journal j. Then we balance the quality of the authors and the quality of the corresponding journal using e_a to recommend journal j for article i. Finally, a matching algorithm that takes into account all three-dimensional measures is proposed for recommending suitable publication outlets.

Profiling of Journals and Manuscripts

In this section, we present manuscript profiling and journal profiling, which are achieved by integrating three different information sources: subjective, objective, and social information. Specifically, subjective information refers to the information that is found in a manuscript or a journal (e.g., keywords presented in a manuscript's keyword section). The information that is derivable from the title and abstract of a manuscript is considered objective information. From a journal's perspective, objective information is what can be derived from the titles and abstracts of articles that are published in a journal. What the authors who publish articles in a journal and what their peers say about a journal is referred to as the social information of a journal. The quality of profiling of both articles and journals directly affects the quality of journal recommendation. Thus, it is necessary to build comprehensive profiles by integrating all these information sources in order to support effective, high-quality recommendation.

Manuscript Profiling

The objective of manuscript profiling is to extract relevant features including keywords and subject categories that can be used to represent manuscripts uniquely. Keywords are extracted from both standard and nonstandard sections of a manuscript. The keywords mentioned in the keyword section are extracted and are treated as subjective information in this study. Authors can also claim the subject category to which their manuscript may belong. We express the subject categories and extracted keywords using the subjective key vector as follows:

$$<\text{ArticleNo}, \text{subjcat}_1, \text{subjcat}_2 \ldots \text{subjcat}_{m1}, \text{Key}_1, \text{Key}_2 \ldots \text{Key}_{n1}>, \quad (1)$$

where *ArticleNo* is the unique article number assigned to an article. *n1* is the number of extracted keywords from the keyword section of the manuscript and *m1* is the number of subject categories presented in the manuscript.

By analyzing the nonstandard key sections such as the title and the abstract of an article, we extract the objective key vector as follows:

$$<\text{ArticleNo}, \text{key}_1, \text{key}_2 \ldots \text{key}_m>, \quad (2)$$

where m is used to denote the number of keywords in the objective key vector and that number is to be determined empirically. Here we use simple lower-case letters to represent keys in the objective key vector.

Importantly, these two key vectors might have an overlapping set of keywords. For effective comparison, we extract m number of keywords from each document. The algorithm that we discuss in the Extracting Key Phrases From Nonstandard Key Areas section determines the preferred number of keywords. Although it is desirable to use the entire article text for extracting relevant keywords, we found that this adds more computational complexity without adding more insight. Therefore, we analyzed keywords only from the title, keywords, and abstract.

Journal Profiling

The objective of journal profiling is to extract a journal's key attributes including key disciplines that it belongs to as well as keywords covered by the published articles. Thus, all subject categories are extracted in order to construct a comprehensive journal profile. Analyzing subjective information including subject category and keywords that appear on the scope section of a journal home page, we generate the subjective key vector as follows:

$$<\text{JournalID}, \text{subjcat}_1, \text{subjcat}_2 \ldots \text{subjcat}_{m2}, \text{Key}_1, \text{Key}_2 \ldots \text{Key}_n> \quad (3)$$

where *JournalID* is used to uniquely identify a journal. The parameter *m2* denotes the number of disciplines that the specific journal belongs to and the value of n can be varied from one journal to another. Generally, *m2* will be two (Leydesdorff & Rafols, 2008). As before, n is used to represent the number of keywords that are extracted from the "scope and aim" sections of a journal.

By analyzing titles and keywords of the articles that are published in a specific journal, the objective keyword vector for that journal is constructed. This objective keyword vector is used to verify the areas that a specific journal belongs to, and it also helps in distinguishing journals. The keyword vector that represents objective information can be written as follows:

$$<\text{JournalID}, \text{key}_1, \text{key}_2 \ldots \text{key}_m> \quad (4)$$

In addition, journals have social tags. As mentioned before, social tags are labels about the scope of a journal and usually

DOI: 10.1002/asi

are the judgments of experts in the corresponding field. For example, senior scholars may determine the areas covered by a journal in his/her field based on his experience. Information extracted from the social tags of external expertise can be aggregated and expressed as

$$<\text{JournalID}, \text{key}_1, \text{key}_2 \ldots \text{key}_m> \quad (5)$$

Hybrid Rough Threshold Model for Scientific Key Phrase Extraction

During the process of objective information extraction, it is necessary to analyze nonfree text areas such as the titles and abstracts of research articles. The determination of a set of topic features from nontext fields follows several steps that include extracting of phrases, filtering out nonkey phrases, and resolving semantic heterogeneity. In this study, we combine several techniques including Rough Threshold Model, Keyword Co-Occurrence, and Database Tomography and develop an algorithm to calculate the document phrase weight distribution.

Extracting Key Phrases From Nonstandard Key Areas

According to RTM (Li et al., 2012), documents are represented in terms of a weight distribution over topic features. In the following, we describe our augmented topic filtering algorithm that is used to generate key topics from articles. In this algorithm, we use the word co-occurrence technique together with multiple keyword phrases to address the semantic heterogeneity issue, which is the main drawback of the RTM algorithm. Phrases (a combination of multiple words) rather than a single word are used to solve semantic ambiguity, as single words are rarely sufficient to accurately distinguish standing research expertise (Strzalkowski, 1994). We have found that a phrase with a length of two to four keywords is strong enough to capture the meaning effectively.

The standard areas of scientific publications (i.e., title, abstract, and keywords) are analyzed and technical phrases are extracted using the Database Tomography (DT) process (Kostoff, Braun, Schubert, Toothman, & Humenik, 2000; Kostoff, del Rio, Humenik, Garcia, & Ramirez, 2001). DT is a textual database analysis system that provides algorithms for extracting multiword phrase frequencies with their proximities. We apply the DT algorithm to extract all adjacent double, adjacent triple, and adjacent quadruple word phrases from the text (i.e., title, abstract, and keywords) along with their frequencies. We discard those phrases with extremely high frequencies, as they are not useful in distinguishing documents and those with extremely low frequencies, as they are not useful in comparing documents. Finally, these phrases are built into the keyword dictionary.

Thus, a given article i can be represented as $d_i = \{(p_{i1}, tf_{i1}), (p_{i2}, tf_{i2}) \ldots (p_{im}, tf_{im})\}$ where tf_{i1} is the number of times the key phrase p_{i1} appears in the document i. Then we calculate the normalized weight for each term frequency to convert them into common comparable quantities. The formula used to calculate the normalized weight can be presented as follows:

$$w_{ik} = \frac{tf_{ik}}{\sum_{r=1}^{m} tf_{ir}}, r = 1, 2, \ldots, m \quad (6)$$

Thus, any given article is represented by using a set of key phrases and their associated normalized weights, that is, $di = \{(p_{i1}, w_{i1}), (p_{i2}, w_{i2}) \ldots (p_{im}, w_{im})\}$.

Resolving Semantic Heterogeneity of Key Phrases

Further, to resolve semantic heterogeneity we use the keyword co-occurrence model. Therefore, for each $di = \{(p_{i1}, w_{i1}), (p_{i2}, w_{i2}) \ldots (p_{im}, w_{im})\}$ we first construct its key phrase vector $P = \{p_1, p_2 \ldots p_n\}$. Then we construct the phrase co-occurrence matrix (PC) using those identified phrases. The elements of the co-occurrence matrix represent the relativity of pairwise phrases. The relativity of any keyword pair is measured quantitatively using the co-occurrence model that uses mutual information, $MI(p_{i1}, p_{j1})$, directly. According to Wang et al. (2003), the fundamental co-occurrence model can be represented as follows:

$$MI(p_{i1}, p_{j1}) = P(p_{i1}, p_{j1}) log\left(\frac{P(p_{i1}, p_{j1})}{P(p_{i1}) \cdot P(p_{j1})}\right) \quad (7)$$

$$P(p_{i1}, p_{j1}) = \frac{C(p_{i1}, p_{j1})}{\sum_{p_i, p_j} C(p_i, p_j)} \quad (8)$$

$$P(p_{i1}) = \frac{C(p_{i1})}{\sum_{p_i} C(p_i)} \quad (9)$$

$C(p_{i1}, p_{j1})$ in Equation (8), denotes the frequency that any two phrases p_{i1} and p_{j1} occur in the same publication and p_i and p_j are variables, representing any other phrases in the document. In Equation (9), $C(p_{i1})$ refers to the frequency of phrase p_{i1} occurring in the document. The probability value of occurring phrase pair p_{i1} and p_{j1} in one document is denoted by $p(p_{i1}, p_{j1})$. $p(p_{i1})$ represents the statistical probability of the occurrence of phrase p_{i1} independently.

The dot product of vector d_i and matrix PC is calculated and the elements of the dot product matrix are referred to as the normalized weighted frequency of the phrases in the document. This process can be summarized as in Algorithm 1. The top m phrases with their normalized weighted frequencies are selected to represent the objective key phrase vector.

DOI: 10.1002/asi

ALGORITHM 1. Calculate Document Phrase Weight Distribution

```
Input: A set of documents
Output: Document's phrase weight distributions
RP = Φ;
for each d_i ∈ D
     d_i = {(p_1, f_i1 ), . . . (p_m, f_im)};
     for each (p_i, f_i1) ∈ d_i do
          Calculate w_im using equation (6);
          rp_i = {p_j | w_im > 0};
          RP = RP ∪ {rp_i};
     end
     i=1;
     for each rp_i ∈ RP do
          j = i + 1;
          for each rp_j ∈ RP do
               if (rp_i = rp_j) or MI(rp_i, rp_j) > threshold
               then
                  ⊕(rp_i, rp_j); //(rp_i, a) ⊕ (rp_j, b) = (rp_i, a + b)
               end
          end
     end
     dp_i = {(rp_1, (dp_i. PC_1), (rp_2, PC_2), . . . , (rp_m, PC_m)};
     d_pI =∪ dp_i;
end
return dp_I
```

Relevance Index: Similarity Between Journal and Manuscript Profiles

The relevance index determines the matching degree between both manuscript and journal profiles. The task of calculating the relevance index can be viewed as deciding whether a sequence of key phrases that are attributes of an article profile matches key phrases that describe traits of a journal profile. We use two widely used approaches, the Jaccard similarity measure (Hidenao & Shusaku, 2010) and the cosine similarity measure (Dong, Sun, & Jia, 2006), to generate the matching degree of both profiles. The data extracted by the expressions (1) and (3) can be matched using the Jaccard method. As subject categories represent more general concepts than the keyphrases, they cannot be matched together. Therefore, we perform a component-based matching strategy and the keywords presented in expression (1) are matched with the keywords presented in expression (3) while the subject categories presented in expression (1) are matched with the subject categories presented in expression (3). Hence, the Jaccard index between manuscript i and journal j is expressed as:

$$J_{ij} = \rho \frac{F\left[(\mathrm{Key}_{i1} \ldots \mathrm{Key}_{in1}) \bigcap (\mathrm{Key}_{j1} \ldots \mathrm{Key}_{jn})\right]}{F\left[(\mathrm{Key}_{i1}, \ldots \mathrm{Key}_{in1}) \bigcup (\mathrm{Key}_{j1} \ldots \mathrm{Key}_{jn})\right]} + (1-\rho) \frac{F\left[(\mathrm{Subcat}_{i1}, \ldots \mathrm{Subcat}_{im1}) \bigcap (\mathrm{Subcat}_{j1}, \ldots \mathrm{Subcat}_{jm2})\right]}{F\left[(\mathrm{Subcat}_{i1}, \ldots \mathrm{Subcat}_{im1}) \bigcup (\mathrm{Subcat}_{j1}, \mathrm{Subcat}_{jm2})\right]} \quad (10)$$

where Key_{ik} represent keywords that appear on subjective key vectors, $\wedge$ and $\wedge$ represent subject categories that appear on subjective key vectors in an article profile and a journal profile, respectively. According to our empirical observation, the maximum number of elements in a subjective key vector is 6 and the maximum number of subject categories is 2. ρ is the weighted factor and $0 < \rho < 1$.

To identify the best parameter value for ρ, we selected 35 published journal articles and their corresponding journals. We varied the value ρ from 0.1 to 1 and for each ρ value the accuracy (i.e., precision) of the top 3 predicated journals were computed. Additionally, the rank of the corresponding journal (Mean Reciprocal Rank [MRR]) was computed. The best case was achieved at $\rho = 0.6$ and precision was 65.6% and MRR was 0.611.

The relevance degree between the objective key vectors as well as the social key vector is determined by using the cosine similarity measure. Data extracted from the formulae (2), (4), and (5) are matched using this measure. For article profile i and journal profile j, the similarity can be calculated as follows:

$$C_{ij} = \frac{w_i w_j}{\|w_i\| \|w_j\|} = \frac{\sum_{k=1}^{m} w_{ik} w_{jk}}{\sqrt{\sum_{k=1}^{m} w_{ik}^2 \sum_{k=1}^{m} w_{jk}^2}} \quad (11)$$

where w_i and w_j are the normalized frequencies of the keyphrases in the two profiles and they are generated by using the algorithm presented in Figure 3.

Denote r_{ij} as the relevance degree matching between manuscript i and journal j. An aggregate measure in the relevance dimension can be obtained as follows:

$$r_{ij} = \alpha J_{ij} + (1-\alpha) C_{ij}, \text{ where } 0 < \alpha < 1 \quad (12)$$

To identify the best parameter value for α, we selected 30 published journal articles and their corresponding journals. We varied the value of α from 0.1 to 1 by fixing ρ at 0.4, and

DOI: 10.1002/asi

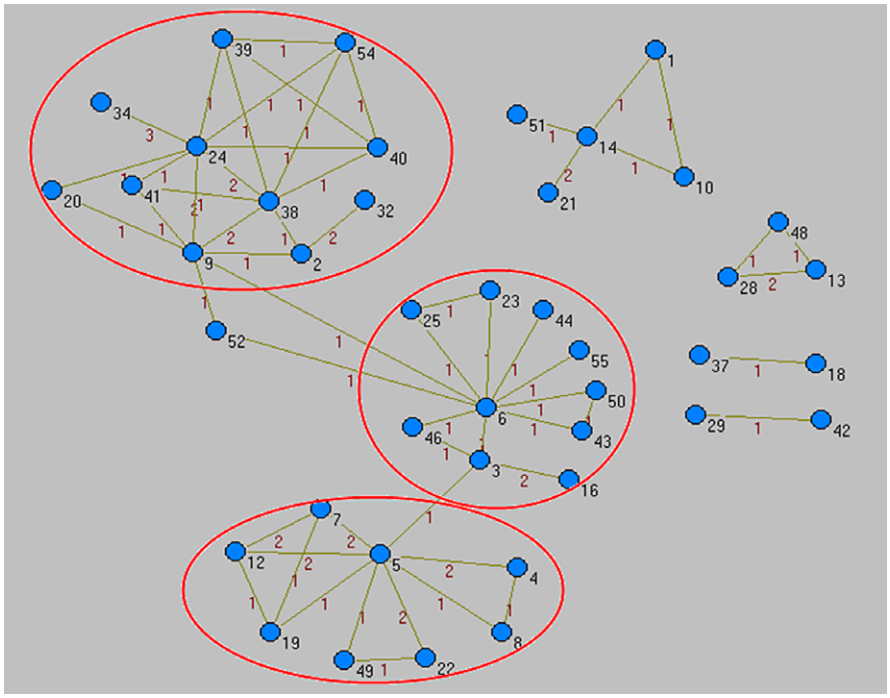


FIG. 3. Example of constructed researchers' collaboration networks. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

for each α value the accuracy (i.e., precision) of the top 3 predicated journals were computed. Additionally, the rank of the corresponding journal (MRR) was computed. The best case was achieved at $\alpha = 0.4$ and the precision was 66% with MRR 0.678. Similarly, we varied the value of ρ from 0.1 to 1 and for each ρ the precision and MRR were computed. The best value for ρ was achieved at 0.4 with precision = 71% and MRR = 0.635.

Connectivity Index: Identifying Widely Accepted Journals

To identify widely accepted journals by similar researchers, a collaboration network analysis was performed and the connectivity index was calculated. The connectivity index measures the strength of the connection between researchers, and it is used to identify potential journals in which similar researchers (e.g., co-authors) have published.

We focus on the collaboration network of authors and their expertise network. A node in the collaboration network represents a researcher. An edge between two nodes is constructed when one researcher has co-authored with the other researcher. We first assign the weight w_{ij} for a pair of vertices, which is defined as the frequency of collaboration between two researchers. High weight implies more connectivity between the two researchers. We first use a graph clustering method to identify groups of similar researchers (i.e., communities) in the collaboration network. Hierarchical clustering, which is a traditional method for detecting community structure, is followed to derive an optimal community structure. Assume that there are K predefined communities. Define w_{IJ} as the fraction of the collaboration frequency among researchers in community II compared to those in community JJ. Denote $a_I = \Sigma_J \; W_{IJ}$, which represents the weighted fraction of edges that connect to vertices in the other communities (i.e., the fraction of collaborations that the researchers in a community collaborate with researchers in other communities). Following Newman's (2001) fast algorithm, which is based on the idea of modularity (Dong et al., 2006), we define the modularity measure for a network with communities as

$$Q_k = \sum_{i=1}^{k} \left(w_{II} - a_I^2 \right) \tag{13}$$

where w_{II} is the weighted fraction of edges in the network that connect vertices in the same community. A high value of Q_k represents a good community division. However, optimizing Q_k over all possible divisions is not feasible in practice for networks larger than 30 vertices. Various approximation methods are available, such as simulated annealing, genetic algorithms, and so on. A standard "greedy" optimization algorithm is used (Ratnayaka, Wang, Anamalamudi, & Cheng, 2012). The hierarchical clustering method also enables us to define the community structure according to the required granularity level.

DOI: 10.1002/asi

TABLE 1. Decision rules for identifying journals in which similar researchers have published.

```
IF (RK and A are in same community)
 IF (a link connecting RK and J exits in researcher-journal network)
  IF (RK and A connected via manuscript topics in researcher-topic network)
   THEN
    cij = 1 (Journal j is more related to the manuscript i)
   ELSE
    cij = 0.05 (Journal j is less related to the manuscript i)
  END IF
 END IF
END IF
```

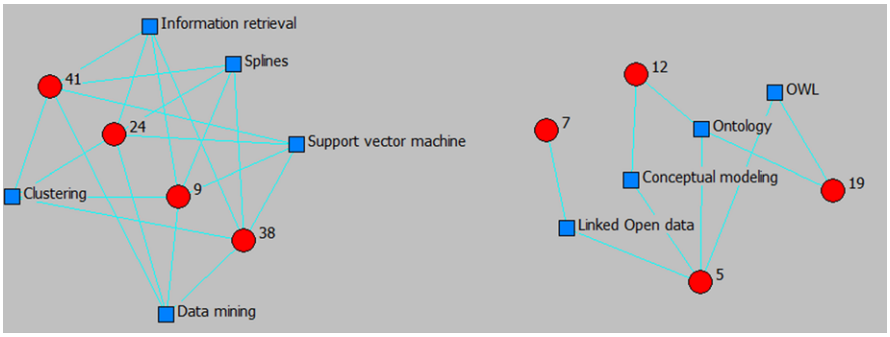


FIG. 4. Two-mode networks representing researchers and their expertise. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

As the next step, two two-mode networks, researcher-topic and researcher-journal, are constructed for each community. The purpose of the researcher-topic network is to determine similar researchers who share similar topics and the purpose of the researcher-journal networks is to identify journals in which similar researchers have published. The decision rule for identifying similar researchers as authors and determining their published journals can be viewed as follows. Here, *RK* represents any researcher, A represents a manuscript author and *J* represents any journal. According to this decision rule (Table 1), the connectivity index c_{ij} which represents the connection strength between author *i* and journal *j* is set to 1 if journal *j* is used by researchers in the same community as the manuscript author. Otherwise, c_{ij} is set to 0.05, indicating less relevance of journal *j* to manuscript *i*.

Illustrative Example for Connectivity Index Calculation

In this section, we illustrate the connectivity index calculation by using an example. Suppose that we have analyzed the publication records and a collaboration network has been constructed. After applying the aforementioned algorithm, corresponding communities have been derived as shown in Figure 3. Derived communities are marked by using red circles. Next we construct the two two-mode networks, researcher-topic and researcher-journal networks for each community. A sample of the generated researcher-topic network for one community is presented in Figure 4. It illustrates a partition of researchers who are closely connected with some of the discovered topics. The cluster on the left is about "knowledge management." The cluster on the right is about "Semantic Web." This shows that the expertise of researchers is highly clustered on these topic areas.

In parallel, we construct the journal-researcher network as presented in Figure 5. This network helps in identifying journals in which similar researches have published. For example, suppose author 24 has written an article titled "Dynamically Integrating Knowledge in Social Teams: Transforming Resources Into Performance." As this article belongs to the areas covered under knowledge management, researcher 41, researcher 38, and researcher 9 are identified as researchers with similar expertise. Then based on the journal-researcher network, journals in which researcher 41, researcher 38, and researcher 9 have published in can be selected as potential publication outlets, that is, *Academy of Management Journal* (*AMJ*), *Journal of Management Information Systems* (*JMIS*), *Journal of Management Studies* (*JMS*), *Management Science* (*MS*), *International Journal of Organizational Behavior* (*IJOB*), *MIS Quartely* (*MISQ*), *IEEE Transactions on Knowledge and Data Engineering* (*TKDE*), *Journal of Information Science* (*JIS*), and *International Journal of Electronic Commerce* are candidate

DOI: 10.1002/asi

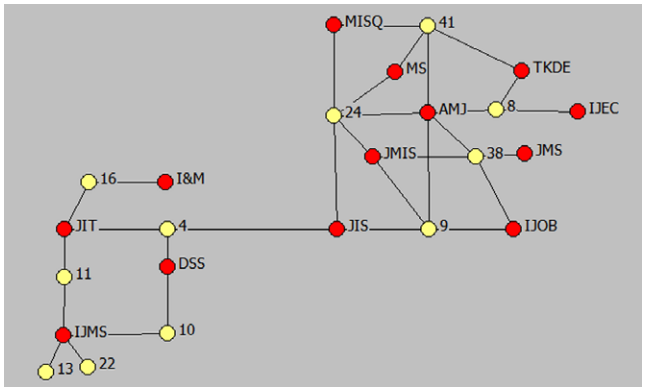


FIG. 5. Journal-author two-mode network. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

publication outlets for the author 24. Therefore, we set the parameter value of c_{ij} to one following the above decision rule. In this example, $j \in \{AMJ, JMS, IJOB, JMIS, MS, MISQ, TKDE, JIS, IJEC\}$ and i represents the manuscript "Dynamically Integrating Knowledge in Social Teams: Transforming Resources Into Performance."

The use of the researcher-topic network and researcher-journal network generates several advantages. First, it helps in identifying more localized potential publication outlets. Second, this greatly improves the efficiency in determining the relevance of journals to the article.

Productivity Index: Identifying Highly Productive Journals

The productivity matrix captures the quality of the publication outlet and the author's contribution to the field. We explore the quality of the authors of manuscripts and the quality of the journals using the productivity index as described in the following two subsections.

Measuring the Quality of Manuscript Authors

Generally, it is difficult to judge the quality of a manuscript, and the authors of an article are not in the best position to judge the quality of their manuscript. Therefore, we consider the level of expertise of the author/authors instead of the quality of a manuscript before recommending the most productive publication outlet for their manuscript. We further assume that professionalism is reflected in the quality of research output and a patient researcher who has established a reputation can submit their manuscript to top-level journals and if unsuccessful they can submit it to the journal with lower rewards but with a high acceptance rate (Heintzelman & Nocetti, 2009). In contrast, young researchers (i.e., untenured and who are likely to be impatient and risk averse) should consider submitting to nonpremier journals with a high acceptance rate first (Heintzelman & Nocetti, 2009) to achieve high academic rewards.

The main aim of the productivity index in journal recommendation is to help researchers find the most effective journals (i.e., journals with higher probability of acceptance) for their submissions. Therefore, for unbiased decision making, we recommend all relevant journals in which quality exceeds the author's quality for a given manuscript. This guarantees that the final list of journal recommendations for a young researcher will include high-ranking journals with low priorities as well as average-rank journals with high priorities. By doing so, we open room for younger researchers to select high-level journals for their submissions if they want. We measure the productivity of the authors of an article in terms of the number of publications as well as quality of their publications and their academic achievement (e.g., reputation and h-index). Thus, we compute the productivity index as an aggregation of quality and quantity measurements.

Generally, academic journals are classified into different disciplines and they are assigned a rank, such as level A journals, level B journals, or level C journals. As in Sun, Ma, Fan, and Wang (2008), we assume that the journal rank reflects the quality of the articles published in that journal as it is widely used in many research performance measuring activities related to merit increases and for the allocation of research funding in university settings (Turban, Zhou, & Ma, 2004). Following Sun et al. (2008), we adopt a weighted scheme to generate the quality index as a measure of contribution made by the author towards the field. Let q_A, q_B, and q_C be author a's total number of publications in level A, level B, and level C journals respectively. The publication quality of author a is represented by D_a and can be expressed as:

$$D_a = w_A q_A + w_B q_B + w_C q_C \tag{14}$$

where $w_A > w_B > w_C$, indicating the emphasis on quality work. There are different ways to define the weights. For example, the average impact factors for all the journals classified at the same level can be used to define the corresponding weight. Professional titles (e.g., senior scholar such as professor and associate professor, or junior scholar such as assistant professor) and h-index can also be taken into consideration for recommending journals. We assign a higher rank score to higher professional titles. Let R_a and H_a be author a's rank score and h-index, respectively. An integrated Productivity Index of author a can be obtained as follows.

$$e_a = uD_a + vR_a + tH_a, \quad \text{where } u + v + t = 1. \tag{15}$$

To identify the best parameter value for u, we selected 30 published journal articles and their corresponding journals. We varied the value of u from 0.1 to 1 by fixing v and t at 0.2 and 0.3, respectively, and for each value of u the accuracy (i.e., precision) of the top 3 predicated journals were computed. Additionally, the rank of the corresponding journal

DOI: 10.1002/asi

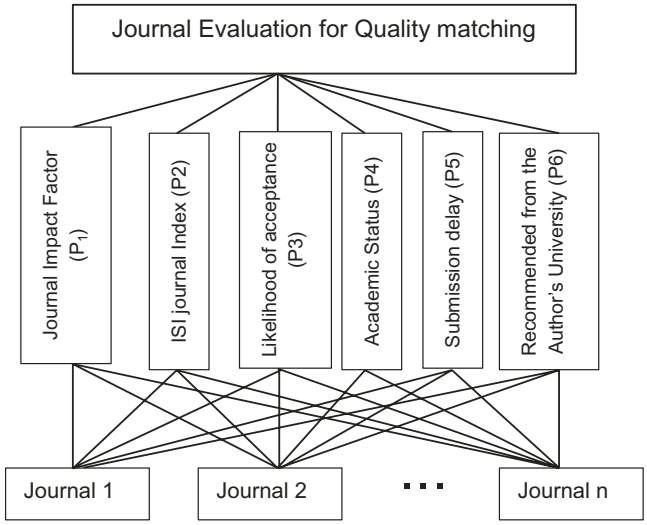


FIG. 6. Hierarchical structural model for journal quality evaluation.

(MRR) was computed. The best case was achieved when $u = 0.5$ with precision = 71% and MRR = 0.589. Similarly, we varied the value of v from 0.1 to 1 and for each v the precision and MRR were computed while fixing the values of u and t at 0.5 and 0.3, respectively. The best value for v was achieved at 0.2 with precision = 69.5% and MRR = 0.745. Similarly, we found that the parameter t got its maximal value at 0.3. The recorded precision and MRR were 68% and 0.656, respectively.

Measuring the Quality of Journals

It is also important to judge the quality of the publication outlet when deciding on the appropriate journals for submission. Different matrices including impact factors and publishing houses are used to evaluate the journals (Andonie, Dzitac, Agora, & Tineretului, 2010; Bjork & Holmstrom, 2006). Bjork and Holmstrom (2006) proposed a journal benchmarking framework from a submission authors' point of view. We follow their framework and have identified six of the most critical factors that can be used to determine the quality of a journal within the context of manuscript submission. They are journal impact factor, ISI indexed journals, likelihood of acceptance, academic status, submission delay, and recommended from the author's university (Figure 6). All information related to these criteria needs to be aggregated to perform an overall evaluation of journal quality. Here, the Analytical Hierarchy Process (AHP) is implemented to determine the importance of one journal over another through pairwise comparison. We selected AHP for this task for the following reasons. First, it is user friendly because users can directly input judgment without in-depth knowledge of mathematics. Second, relevant inconsistency in individual judgments is dealt with appropriately. Last, the power of AHP has been validated by

TABLE 2. Comparison scale.

Absolute value	Definition
1	Equal importance
3	Moderate importance over another
5	Strong or essential importance of one over another
7	Very strong or demonstrated importance of one over another
9	Extreme importance of one over another
2, 4, 6, 8	Intermediate values
Reciprocals	Reciprocals for inverse comparison

empirical applications in diverse areas (Sun et al., 2008). A scale of absolute values of 1 to 9 is used for making the pairwise comparison judgments (Sun et al., 2008). The scales are listed and explained in Table 2. Thus, once making the pairwise comparison judgment among potential journals, weights of the journals are generated and are used to judge the quality matching degree of journals and the manuscript as presented in the following section. The detailed algorithms of AHP can be found in Sun et al. (2008).

Suppose that x_i denotes the weight of the criterion i ($i = 1, 2, 3, 4, 5, 6$) and J_j represents the overall evaluation on a journal j, then J_j can be calculated as follows:

$$J_j = \sum_{i=1}^{6} x_i P_i,\ j = 1, 2, \ldots, n \tag{16}$$

Recommending Journals for Manuscripts

The key objective of our approach is to recommend journals which maximize the relevance, that is, the matching degree between the contents of a manuscript and the contents of the publications of a journal and maximize the connectivity index. Because the quality of the recommendation largely depends on the quality of the publication outlets, and their usefulness or appropriateness to the author, there is a need to maintain a balance between an author's productivity and journal quality. As we argued before, impatient young researchers seek publication outlets which have a high probability of acceptance for their manuscripts. Based on the argument presented by Heintzelman and Nocetti (2009), we recommend journals with lower rewards but with a high acceptance rate. Thus, we recommend all journals whose quality exceeds the productivity of the authors.

A multiobjective optimization model for journal recommendation can be developed as follows. Let x_{ij} be the integer decision variable indicating the recommendation of a given research article i to a potential publication outlet j. Therefore, $x_{ij} = 1$ implies that the assignment is recommended and $x_{ij} = 0$ implies otherwise. We want to maximize the relevance (i.e., relevance index) as well as the connectivity index, which represents widely accepted journals for similar

DOI: 10.1002/asi

researchers while satisfying the flow constraints. Thus, the multiobjective optimization model for journal recommendation can be presented as follows:

$$\begin{array}{lll} \textit{Maximize} & \sum_{j\in J} c_{ij} r_{ij} x_{ij} & \text{for a given i} \\ \textit{s.t.} & e_a x_{ij} - J_j \geq 0 & \text{for } j \in J \\ & \sum_{j\in J} x_{ij} \leq d & \text{for } j \in J \\ & x_{ij} \in \{0,1\} & \text{for } i \in I,\ j \in J \end{array} \quad (17)$$

The coefficients in the objective function ensure that we maximize the overall relevance measure in the article and publication outlet pools. c_{ij} is the indicator for preferential assignment of a journal to a given manuscript. The second constraint is used to balance the authors' productivity e_a on potential journals. J_j represents the quality index of the journal j. The third constraint is to determine the maximum number of journals per article. Note that $d > 0$ can be chosen by empirical observation.

With the increment of the number of potential journals, the exact multiobjective optimization algorithms fail to handle the above model efficiently. Thus, we use the heuristic approach based on the particle swarm optimization technique to solve the above-mentioned multiobjective optimization problem effectively, especially when the number of candidate journals is large, that is, when the solution space is large.

Solving Journal Recommendation Model

We follow the Multi Objective Particle Swarm Optimization (MOPSO) as in Haeri and Tavakkoli-Moghaddam (2012) to solve our multiobjective journal recommendation model presented above, for it can generate two main advantages over the others. First, it is a very simple approach when compared to the other traditional optimization algorithms such as genetic algorithms (GA), as it considers only one operator for creating a new solution (Reyes-Sierra & Coello, 2006) leading to less computational complexity. Second, it exhibits superior performance when producing optimal results with low computational cost (Kennedy, 2006). It also has a fast convergence rate toward the global optimal solutions.

MOPSO follows three main steps: initialization, evaluation of particles representing candidate journals, and update velocity and position of candidate journal particles in this journal recommendation context. Figure 7 shows these main steps together with their corresponding subtasks.

Step 1: Initialize candidate journal particle swarm. The objective of this step is to initialize the particles representing candidate journals with corresponding velocity and positions. Traditional particle swarm optimization deals with continuous variables. As the problem of journal recommendation consists of numerical values, it is necessary to denote particles which represent candidate journals with numerical values. Therefore, as the first step, a string with n-real numbers is defined and named as the original string. Each number in the original string corresponds to one candidate journal and its value is in the [0, 1] closed interval. Then, following the Rank Ordered Value (ROV) method (Liu, Wang, Jin, & Huang, 2006), the numbers of the original string are sorted in ascending order. Then a rank is assigned to each real number in the ascending order. The corresponding final list of candidate journals is prepared by assigning the corresponding ranks (in the sorted list) to the real numbers in the original string.

For example, consider a situation where we have five candidate journals and assume that the original string A is [0.23, 0.56, 0.81, 0.34, 0.13]. The sorted order of A is as follows: A-sorted = [0.13, 0.23, 0.34, 0.56, 0.81]. Then the ranks are assigned to the sorted list.

$$\text{A-sorted with ranks assigned} = \begin{bmatrix} 0.13 & 0.23 & 0.34 & 0.56 & 0.81 \\ 1 & 2 & 3 & 4 & 5 \end{bmatrix}$$

Thus, the corresponding final list of candidate journals is [2, 4, 5, 3, 1].

As the second step of swarm initialization, the velocities of each particle are set to zero.

Step 2: Evaluation of candidate journal particles. All generated particles are evaluated via processing them through the optimization model. Based on the evaluation results the nondominated solutions from the swarm are selected. A solution j is said to be nondominated, if there are no other solutions that exceed the objective values of solution j (Haeri & Tavakkoli-Moghaddam, 2012).

The selected nondominated solutions are stored in an external repository. The repository serves as a pool of nondominated solutions and its capacity is to be determined before the execution of the algorithm. Before adding new solutions to the repository, it is necessary to compare them with the existing solutions. If there is any solution that is dominated by the new solutions, then it is deleted from the repository. Due to its limited capacity, in some situations it is necessary to remove some solutions from the repository. As a strategy to identify the solution that ought to be removed from the repository, we use their crowding distance (CD) factor as in Deb, Pratap, Agarwal, and Meyarivan (2002). This is the indicator that is used to denote how much a solution is crowded with other solutions. The solutions with lower CD values are excluded from the list. This implies that solutions which are more crowded with other solutions are removed to maintain more diversification in the space search process when executing the algorithm (Haeri & Tavakkoli-Moghaddam, 2012). This guarantees that the final recommended list of journals consists of a diversified set rather localizing only to a very similar set of journals.

Step 3: Update velocity and position of candidate journal particles. With each iteration, the velocities of the particles

DOI: 10.1002/asi

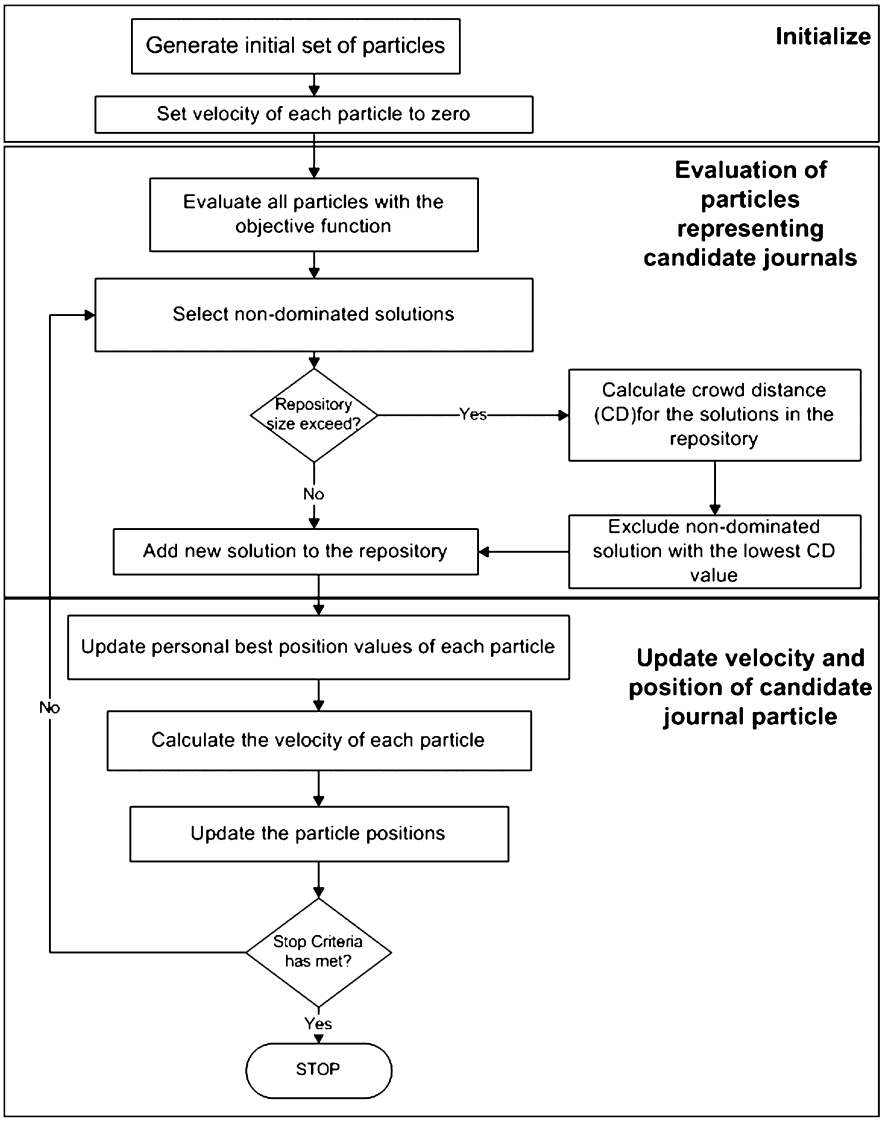


FIG. 7. Flow chart for assignment model solving using particle swarm optimization.

are updated. Following the equation presented in Haeri and Tavakkoli-Moghaddam (2012) we update the velocity of each particle as follows:

$$v_{i+1} = \left(w \cdot v_i + c_1 r_1 \left(pbest_i - x_i\right) + c_2 r_2 \left(rep_H - x_i\right)\right) \quad (18)$$

where v_{i+1} and v_i represent the velocity vectors in the $i+1^{th}$ and i^{th} iteration, respectively, $pbest_i$ is the best position of the particle of its i^{th} iteration, x_i is the position vector in the i^{th} iteration. c_1 and c_2 represent predefined coefficients as r_1 and r_2 represent random numbers in [0,1]. w is the inertia factor that can be equal to 1. rep_H is the position vector of the representative solution H which has the lowest CD, selected from the solution repository. We follow the equation presented in Haeri and Tavakkoli-Moghaddam (2012) to update the particle positions as follows.

$$x_{i+1} = x_i + v_{i+1} \quad (19)$$

MOPSO continues until its termination condition is reached. Normally, the number of iterations is considered as the termination condition. During the parameter tuning process, it is necessary to set up the values for the parameters. It is recommended that the number of particles be in the range 20 to 80 and the number of iterations is set between 80 and 120 (Coello Coello & Lechuga, 2002). We ran the optimization model for 50 selected manuscripts and its accuracy and MRR were computed. We kept unchanged

DOI: 10.1002/asi

FIG. 8. A sample of visual research CV. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

one parameter when computing the other. The model generated optimal solutions when the number of particles was equal to 50, the number of iterations was equal to 120. c_1 and c_2 are equal to 2 and the capacity of the repository is 20. The achieved precision was 0.76 and MRR was 0.5978.

Implementation and Evaluation

Journal Recommender System

A prototype system that implements the proposed approach has been developed using the ScholarMate platform. ScholarMate (http://www.scholarmate.com) was developed by the authors' team and is a professional research social network that connects people to research with the aim of "innovating smarter." It offers research social network services that help researchers find suitable funding opportunities, suitable journals, and potential research collaborators. On the one hand, researchers can use ScholarMate to manage their research outcomes and research in progress, including research proposal preparation. On the other hand, transparency in information sharing among scholars in ScholarMate will open up opportunities for researchers to participate timely in relevant scholarly activities, such as becoming potential reviewers.

The journal recommender system when initiated first presents an interface to collect information on the manuscript including the title, user defined keywords, and abstract. Once a user logs into the system it automatically collects the user's publication history, h-index, and citations. ScholarMate has a search tool to help researchers extract their publications from existing bibliographic databases (e.g., ISI, Scopus) directly, along with citations of the paper and impact factor of the journal. One way of viewing the extracted and analyzed information is visualizing the research CV. A sample research CV is shown in Figure 8. The left-hand side of the research CV looks exactly the same

DOI: 10.1002/asi

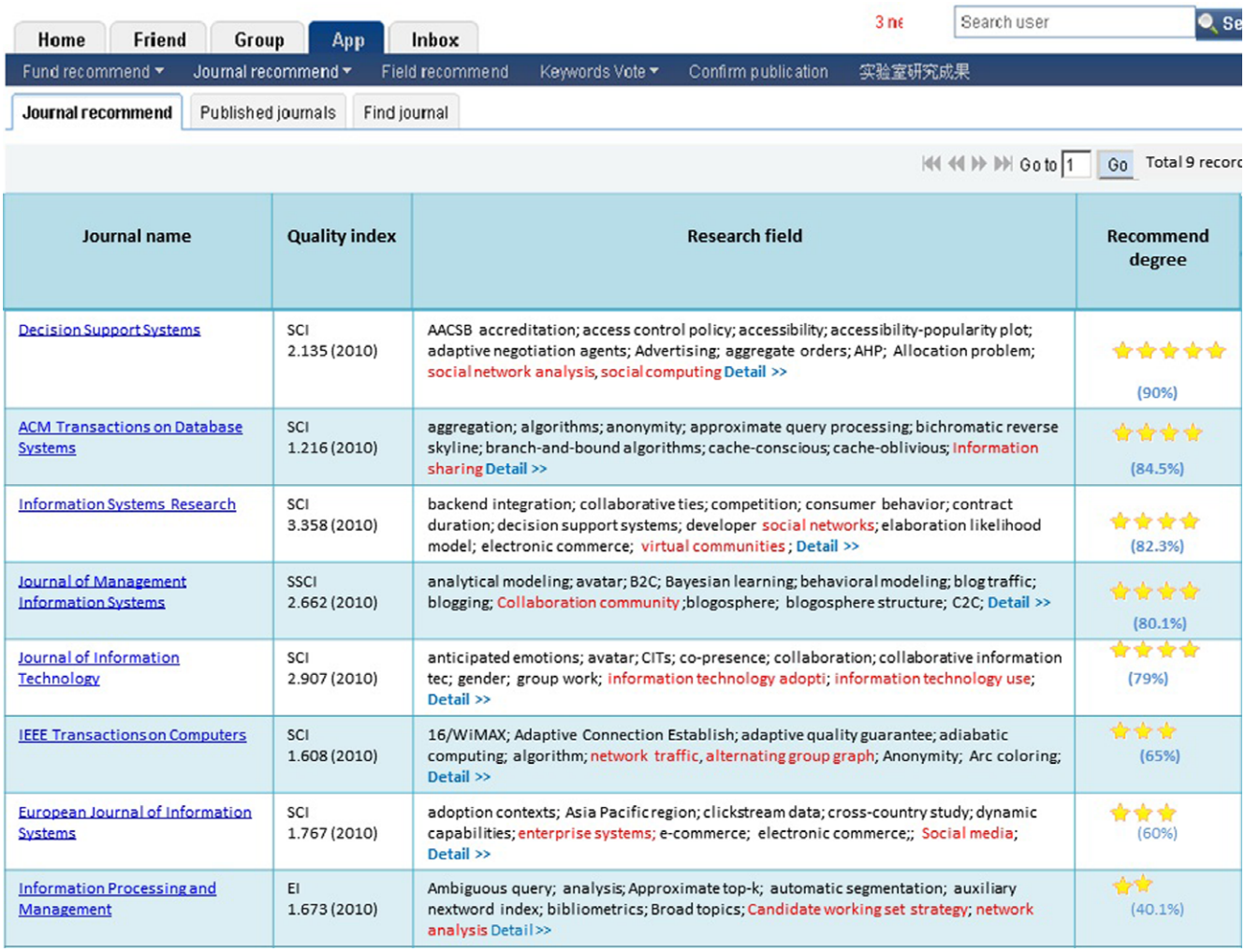
as any standard CV. On the right-hand side of the CV, information about h-index, citations, and number of publications is visualized. These services greatly increase the effectiveness of profile generation.

Journal profiles are constructed beforehand to improve the computational efficiency of the system. During journal profile construction, information about the subject categories that they belong to, keywords describing scope of journals, and other keywords covered by published articles are collected by using preconfigured web crawlers.

Once the system completes gathering the required information, corresponding profiles of the manuscript and journals are created and the degree to which the profile match in terms of relevance, productivity, and connectivity measurements is calculated. Figure 9 presents the system generated list of recommended journals with their appropriate degree of recommendation for the research article titled "Information Exchange in Virtual Communities Under Extreme Disaster Conditions." As shown in the figure, the system also provides additional details about journals such as journal name, quality index, and corresponding research areas. In the figure, common keywords that appear in the journal and the manuscript profiles are marked in red. The relevance score is calculated and displayed in the last column as rating stars.

System Evaluation

The evaluation segment aims to demonstrate the effectiveness of the proposed approach compared to the benchmarked methods in terms of accuracy and quality of the recommended results. Further, it demonstrates the usefulness of the developed system. We adopted two types of evaluation approaches: comparative assessment and online survey following pervious research on recommender system evaluation (Shani & Gunawardana, 2011). The comparative evaluation focuses on measuring the effectiveness of the recommended results of the proposed RAF approach compared to other benchmarked approaches. We used an offline experiment and user-based experiment to compare the



Journal name	Quality index	Research field	Recommend degree
Decision Support Systems	SCI 2.135 (2010)	AACSB accreditation; access control policy; accessibility; accessibility-popularity plot; adaptive negotiation agents; Advertising; aggregate orders; AHP; Allocation problem; social network analysis, social computing Detail >>	(90%)
ACM Transactions on Database Systems	SCI 1.216 (2010)	aggregation; algorithms; anonymity; approximate query processing; bichromatic reverse skyline; branch-and-bound algorithms; cache-conscious; cache-oblivious; Information sharing Detail >>	(84.5%)
Information Systems Research	SCI 3.358 (2010)	backend integration; collaborative ties; competition; consumer behavior; contract duration; decision support systems; developer social networks; elaboration likelihood model; electronic commerce; virtual communities ; Detail >>	(82.3%)
Journal of Management Information Systems	SSCI 2.662 (2010)	analytical modeling; avatar; B2C; Bayesian learning; behavioral modeling; blog traffic; blogging; Collaboration community ;blogosphere; blogosphere structure; C2C; Detail >>	(80.1%)
Journal of Information Technology	SCI 2.907 (2010)	anticipated emotions; avatar; CITs; co-presence; collaboration; collaborative information tec; gender; group work; information technology adopti; information technology use; Detail >>	(79%)
IEEE Transactions on Computers	SCI 1.608 (2010)	16/WiMAX; Adaptive Connection Establish; adaptive quality guarantee; adiabatic computing; algorithm; network traffic, alternating group graph; Anonymity; Arc coloring; Detail >>	(65%)
European Journal of Information Systems	SCI 1.767 (2010)	adoption contexts; Asia Pacific region; clickstream data; cross-country study; dynamic capabilities; enterprise systems; e-commerce; electronic commerce;; Social media; Detail >>	(60%)
Information Processing and Management	EI 1.673 (2010)	Ambiguous query; analysis; Approximate top-k; automatic segmentation; auxiliary nextword index; bibliometrics; Broad topics; Candidate working set strategy; network analysis Detail >>	(40.1%)

FIG. 9. Recommended list of journals for an article titled "Information Exchange in Virtual Communities Under Extreme Disaster Conditions." [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

DOI: 10.1002/asi

TABLE 3. Comparison statistics of the evaluation data set.

Subject category	#articles	#journals
Computer science- AI	6	6
Computer science-information systems	50	16
Computer science-software engineering	14	6
Computer science-theory and methods	7	8
Mathematics-applied	27	4
Business	47	19
Information science	24	8
Social sciences- mathematical methods	13	4
Management	59	16
Engineering-industrial	16	10
Engineering-multidisciplinary	16	8
Engineering-electrical & electronics	21	11

effectiveness of the recommended results. An online survey was used to measure the quality of the recommended results and usefulness of the system.

Data. We randomly selected 300 published articles from four disciplines, management, engineering, computer science, and information sciences, and used them as manuscripts in our experiments. The published journals of the selected articles were recorded and those were selected as the gold standard. For the user-based experiment and online survey, 60 subjects who are registered users of ScholarMate were selected. To get more realistic comments from the subjects and increase the subjects' familiarity with the articles, the authors of those articles were selected as the subjects of this experiment. Among them were 30 research students, 22 research assistants, six assistant professors, and two professors. Altogether, there were 49 distinct journals from 12 different subject categories in the experimental data set. The data statistics are presented in Table 3. The subject categories that the articles belonged to were derived based on the ISI subject category list. By nature, journals are multidisciplinary, and thus one journal can belong to one or more subject categories. To make the evaluation less complex we assigned each article to one main subject category.

Experimental setup. In the offline and user-based effectiveness-measuring experiments, prediction accuracies and precision values of the proposed RAF approach were compared with those that use traditional content-based as well as collaborative filtering approaches. The benchmark models used in this experiment are listed as follows:

1. TF-IDF based recommender approach (CB-TF) (Bellogín, Cantador, & Castells, 2010). It uses TF-IDF weighting scheme (Baeza-Yates & Ribeiro-Neto, 1999) to determine the similar items.
2. Semantic-expansion Content Filtering Method (SeCF) (Liang et al., 2008). SeCF is the enhanced version of TF-IDF method which uses precomputed keyword similarity matrix.
3. Profiling, Jaccard, Cosine (PJC). This method is a part of the proposed approach used to determine content level similarity of the journal and manuscript. It can be considered one of the content-based approaches.
4. Item-based CF recommender (Bellogín et al., 2010). This technique was developed to suggest scientific articles based on authors' similarity. Based on the same argument, we apply item-based CF recommender to recommend journals in which co-authors have published in.
5. Co-author Network Analysis (CNA). This method is used in the proposed approach to identify widely accepted journals by similar researchers. It has been presented in the Connectivity Index section.

For the offline experiment, we compared the prediction accuracy (i.e., precision) and MRR of the proposed RAF approach against the five benchmark models presented above. In all, 150 articles and their corresponding information were used as the training data and the rest were used as the testing data. Sixty articles from the testing data set were used for the offline experiment and the remainder were used in the user-based experiment. For the user-based experiment and online survey, we assigned five different selected manuscripts to each subject and asked them to use the system to find suitable publication outlets. As mentioned previously, before the experiment began information about authors of the articles, including their h-index, professional titles, and citations as well as information about the journals including their rank, disciplines, and impact factor, was fed to the system by using multiple services in ScholarMate. The result lists generated for a given manuscript is the combination of results of six approaches (our proposed approach [RAF], CB-TF, SeCF, PJC, item-based CF, and CNA). Subjects were asked to judge the relevance of the generated results to the manuscript and rank the results based on their relevance to the manuscript. For this task, subjects were prompted with an interface that enables them to mark each item as either relevant or not relevant.

An online survey was carried out at the end of the experiment to collect users' feedback. The online survey measures the quality of the recommended results as well as end user satisfaction. The quality of the recommended results was measured in terms of their relatedness to the journal (i.e., correlation of manuscript and journal profiles) and the quality of the resulted journals.

Evaluation matrices. To measure the accuracy of the proposed approach and to compare results, Precision@K (Prec@K), Mean Average Precision (MAP) (Croft et al., 2010), and MRR were used.

$$Prec@K = \frac{N_{relevent}}{K} \tag{20}$$

$$MAP = \frac{1}{U}\sum_{q=1}^{u}\frac{1}{|m_q|}\sum_{k=1}^{n}P(R_{qk}) \tag{21}$$

DOI: 10.1002/asi

$$MRR = \frac{1}{|U|}\sum_{j=1}^{|U|}\frac{1}{rankF_j} \quad (22)$$

Where K is the number of recommended journals, and K can either be 5, 7, or 10; $N_{relevant}$ is the number of relevant journals in the resulted list; U denotes the number of test manuscripts; m_q is the number of relevant journals for manuscript q; n is the number of retrieved results; $P(R_{qk})$ represents the precision of the retrieved results from the top result until you get to journal k; $rankF_j$ is the rank (position) of the journals in which manuscript q has been published.

Experimental Results and Analysis

Offline and user-based experiment results and analysis. The comparison of MAP, MRR, and precision (prec@5, prec@7, prec@10) of the offline and user-based experiment results are shown in Tables 4 and 5, respectively. According to the results, the proposed RAF approach outperforms all other benchmark models in both experiments. This indicates that journals in which similar researchers have published and the quality of the journals are more influential factors when deciding suitable publication outlets. The item-based collaborator recommender technique achieved the lowest performance, showing that selecting only journals published by similar researchers while ignoring content level similarity will not be an ideal publication outlet selection method.

Furthermore, the proposed content level matching technique outperformed the traditional TF-IDF-based and content-based approaches in terms of precision and MRR. One possible explanation of this behavior is that the proposed approach computes comprehensive profiles while aggregating subjective, objective, and social information, which could help us discover new keywords/keyphrases or missing ones. Hence, the proposed one could solve the serendipity issue as well as the sparsity issue better than the TF + IDF method. When comparing the performance of collaborative filtering approaches such as CAN and item-based CF, the proposed CAN outperforms item-based CF. One possible explanation for this behavior is that we identify similar researchers by analyzing their co-author network in a more narrowed way using two-mode networks. But item-based CF follows ratings to identify similar experts, which is affected by the serendipity issue. This allows us to identify publication outlets of researchers who have similar expertise as manuscript authors and to achieve higher accuracy.

We conducted pairwise t tests on overall prediction accuracy and precision. The t tests compared the performance of our approach at different prec@k, MAP, and MRR against the five baseline recommender approaches in the two experiments. This resulted in 60 comparisons in total for each of our five evaluation metrics. The t test results are shown in Tables 6 and 7, respectively. Only p values less than .001 were considered statistically significant at $\alpha = 0.001$. Thus, we can claim that RAF significantly outperformed all other approaches in terms of overall prediction accuracy and MRR in both experiments.

TABLE 4. Comparison of prec@k, MAP, and MRR of offline experiment.

	prec@5	prec@7	prec@10	MAP	MRR
CB-TF	0.41	0.4077	0.3677	0.4531	0.7295
SeCF	0.4196	0.4089	0.3701	0.4558	0.7445
PJC	0.524	0.5198	0.5268	0.5478	0.7974
Item-based CF	0.2799	0.2736	0.3005	0.3412	0.7343
CNA	0.5142	0.4963	0.494	0.4987	0.7752
Our approach (RAF)	0.5957	0.6311	0.6624	0.7053	0.8067

TABLE 5. Comparison of prec@k, MAP, and MRR of user-based experiment.

	prec@5	prec@7	prec@10	MAP	MRR
CB-TF	0.48	0.4777	0.4377	0.5231	0.8695
SeCF	0.4896	0.4789	0.4401	0.5258	0.8145
PJC	0.512	0.4981	0.4725	0.5587	0.8564
Item-based CF	0.3449	0.3436	0.3565	0.4112	0.7943
CNA	0.489	0.4978	0.4586	0.4638	0.8379
Our approach (RAF)	0.6857	0.6981	0.7321	0.7753	0.8967

Survey Results and Analysis

All 60 participants completed the survey but there were two surveys with incomplete information. Thus, only 58 valid responses were analyzed to evaluate the system from the end user perspective. The questionnaire was developed based on the evaluation framework for recommender systems

TABLE 6. P values of paired t test corresponds to the offline experiment.

	RAF approach				
P values of paired t-test	prec@5	prec@7	prec@10	MAP	MRR
CB—TF	0.00024*	0.00050*	0.000124*	0.00026*	0.00012*
SeCF	0.00014*	0.00017*	0.00037*	0.00038*	0.00053*
PJC	0.00078*	0.00014*	0.00062*	0.00048*	0.00051*
Item-based CF	0.00062*	0.00012*	0.00058*	0.00069*	0.00060*
CNA	0.00054*	0.00011*	0.00045*	0.0000*	0.00019*

*P values are significant at $\alpha = .001$.

DOI: 10.1002/asi

TABLE 7. *P* values of paired *t* test correspond to the user-based experiment.

	RAF approach				
P values of paired *t* test	prec@5	prec@7	prec@10	MAP	MRR
CB—TF	0.00012*	0.00001*	0.00027*	0.000*	0.0002*
SeCF	0.00054*	0.000024*	0.00034*	0.000*	0.0003*
PJC	0.00015	0.00048	0.00014	0.0009	0.00021
Item-based CF	0.00079*	0.00002*	0.00000*	0.000*	0.000*
CNA	0.00014*	0.00023*	0.00017*	0.000*	0.000*

**P* values are significant at $\alpha = .001$.

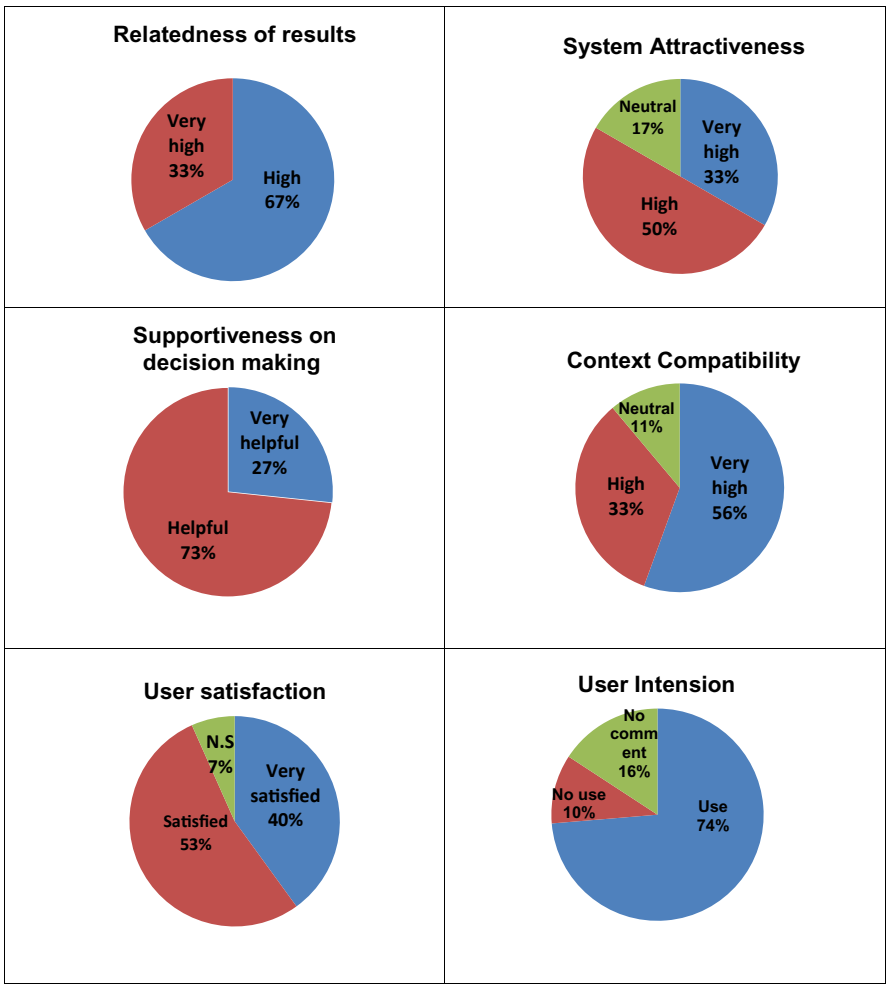


FIG. 10. Survey results. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

proposed by Pu, Chen, and Hu (2011). It covers all the essential aspects of the system and consists of 21 questions. Among the questions are essential system accuracy, that is, whether the system generates highly related journals for a given manuscript (see the first chart in Figure 10); system attractiveness, whether the system produces attractive and convincing outcomes (see the second chart in Figure 10); easy support for decision making, that is, whether the system assists for effective journal selection (see the third chart in figure 10); context compatibility, whether the system provides personalized recommendation results or a general set of results (see the fourth chart in Figure 10). User satisfaction and user intention to use (see the last two charts in Figure 10) are the most significant. Statistical results on the questionnaire survey indicated that the user comments are positive.

DOI: 10.1002/asi

Summary and Conclusion

Our main theoretical contribution is the development of a novel three-dimensional profile-boosted research analytics framework that integrates relevance, productivity, and connectivity by using different technologies such as business intelligence, bibliometric analysis, and social network analysis for effective journal recommendation. Our approach exhibited good predictive accuracy compared to other benchmarked methods. The survey results show that the system is useful, especially for young researchers who are either PhD students or faculty members seeking tenure promotions. In summary, we constructed profiles of different research entities; manuscript and journals from three aspects, that is, relevance, productivity, and connectivity via integrating multiple scientific databases, for example, ISI, Scopus, CNKI, and so on. Building on comprehensive profiles of journals and manuscripts, a unique matching algorithm based on relevance, connectivity, and productivity indices was developed. We model the personalized journal recommendation problem as a multiobjective optimization problem.

This approach can be generalized for any type of recommendation in the research social network environment. Some other potential applications include recommending funding opportunities, recommending scientific articles for researchers, and recommending potential research collaborators. For example, the system may recommend researchers who work in the same research areas to each other within and across different research communities. Based on a researcher's profile, the research social network may also recommend research articles which contain key topics. All of these functions are useful in promoting timely distribution and target dissemination of research work.

There are a number of limitations and possible future research directions. First, journal selection for manuscript submission is highly subjective. We are aware that there are many more influential factors such as review quality, personal relationships, and regional factors which affect the decision-making process in journal selection. Here, we do not model the author's personal beliefs or the rewards that the author is expecting after publishing a manuscript.

Second, the power of ScholarMate is its ability to extract and aggregate information from multiple sources. We need to continuously improve the search tool to meet the increasing needs of users. Moreover, standardization of the keyword dictionary can greatly help the phrase pattern recognition. While we keep evaluating and updating the keyword dictionary based on the feedback of the performance of the algorithm, we are aware that a social vote is another efficient approach in identifying relevant keywords and removing the less meaningful ones.

Third, the current approach uses a collaboration network to evaluate connectivity measurements. It may be difficult to construct the network if a user has had no previous collaborations. Thus, in the future we will incorporate various types of networks including citation network, friends' network, user-research supervisor network, and social-group network to construct a comprehensive network and to achieve high performance in journal recommendation. Moreover, the developed "Smart CV" function in ScholarMate and activities in ScholarMate could greatly help us to identify information about novel users in future.

Acknowledgments

The authors thank the Editor-in-Chief and anonymous reviewers for their valuable comments and suggestions. This research is partially funded by Project Nos.: CityU 148012, CityU 119611 of General Research Fund of Hong Kong; 6000201 of CityU Teaching Development Grant; 71171172, 71371164 of National Natural Science Foundation of China.

References

Adomavicius, G., Tuzhilin, A., & Zheng, R. (2011). REQUEST: A query language for customizing recommendations. Information Systems Research, 22(1), 99–117.

Albers, C.A., Floyd, R.G., Fuhrmann, M.J., & Martínez, R.S. (2011). Publication criteria and recommended areas of improvement within school psychology journals as reported by editors, journal board members, and manuscript authors. Journal of School Psychology, 49(6), 669–689.

Andonie, R., Dzitac, I., Agora, C.D., & Tineretului, P. (2010). How to write a good paper in computer science and how will it be measured by ISI web of knowledge. International Journal of Computers, Communications and Control, 5(4), 432–446.

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

Bellogín, A., Cantador, I., & Castells, P. (2010). A study of heterogeneity in recommendations for a social music service. In _Proceedings of the 1st International Workshop on Information Heterogeneity and Fusion in Recommender Systems (pp. 1–10). Barcelona, Spain.

Biswas, H.K., & Hasan, M. (2007). Using publications and domain knowledge to build research profiles: An application in automatic reviewer assignment. In Proceedings of the International Conference on Information and Communication Technology, ICICT′07 (pp. 82–86). Bangladesh, Dhaka.

Bjork, B.C., & Holmstrom, J. (2006). Benchmarking scientific journals from the submitting author's viewpoint. Learned Publishing, 19(2), 147–155.

Brochner, J., & Bjork, B.C. (2008). Where to submit? Journal choice by construction management authors. Construction Management and Economics, 26(7), 739–749.

Caulkins, J.P., Ding, W., Duncan, G., Krishnan, R., & Nyberg, E. (2006). A method for managing access to web pages: Filtering by statistical classification (FSC) applied to text. Decision Support Systems, 42(1), 144–161.

Chen, C.C., Chen, M.C., & Sun, Y. (2001). A web document personalization user model and system. In Proceedings of the Workshop on Machine Learning, Information Retrieval and User Modeling, Sonthofen, Germany, 2001.

Coello Coello, C.A., & Lechuga, M.S. (2002, May). MOPSO: A proposal for multiple objective particle swarm optimization. In Proceedings of the Congress on Evolutionary Computation (CEC '2002), Vol. 1 (pp. 1051–1056). Honolulu, HI.

Croft, W.B., Metzler, D., & Strohman, T. (2010). Search engines: Information retrieval in practice. Addison-Wesley.

DOI: 10.1002/asi

Deb, K., Pratap, A., Agarwal, S., & Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Transactions on Evolutionary Computation, 6(2), 182–197.

Dong, Y., Sun, Z., & Jia, H. (2006). A cosine similarity-based negative selection algorithm for time series novelty detection. Mechanical Systems and Signal Processing, 20(6), 1461–1472.

Fan, W., Gordon, M.D., & Pathak, P. (2005). Effective profiling of consumer information retrieval needs: A unified framework and empirical comparison. Decision Support Systems, 40(2), 213–233.

Faria, J.R. (2005). The game academics play: Editors versus authors. Bulletin of Economic Research, 57(1), 1–12.

Haeri, A., & Tavakkoli-Moghaddam, R. (2012). Developing a hybrid data mining approach based on multi-objective particle swarm optimization for solving a traveling salesman problem. Journal of Business Economics and Management, 13(5), 951–967.

Hidenao, A., & Shusaku, T. (2010). Comparing a clustering density criteria of temporal patterns of terms obtained by different feature sets. Rough sets and knowledge technology (pp. 248–257): Springer.

He, Q., Kifer, D., Pei, J., Mitra, P., & Giles, C.L. (2011). Citation recommendation without author supervision. Paper presented at the Proceedings of the Fourth ACM International Conference on Web Search and Data Mining (pp. 755–764). Hong Kong, China.

Heintzelman, M., & Nocetti, D. (2009). Where should we submit our manuscript? An analysis of journal submission strategies. The BE Journal of Economic Analysis & Policy, 9(1), Article 39.

Hettich, S., & Pazzani, M.J. (2006). Mining for proposal reviewers: lessons learned at the national science foundation. In Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 862–871). Philadelphia, PA.

Hwang, S., & Chuang, S. (2004). Combining article content and Web usage for literature recommendation in digital libraries. Online Information Review, 28(4), 260–272.

Im, I., & Hars, A. (2007). Does a one-size recommendation system fit all? The effectiveness of collaborative filtering based recommendation systems across different domains and search modes. ACM Transactions on Information Systems (TOIS), 26(1), Article 4, 1–30.

Joachims, T. (2001). A statistical learning learning model of text classification for support vector machines. In Proceedings SIGIR-01, 24th ACM International Conference on Research and Development in Information Retrieval (pp. 128–136).

Kennedy, J. (2006). Swarm intelligence. Handbook of nature-inspired and innovative computing. Springer.

Kostoff, R.N., Braun, T., Schubert, A., Toothman, D.R., & Humenik, J.A. (2000). Fullerene data mining using bibliometrics and database tomography. Journal of Chemical Information and Computer Sciences, 40(1), 19–39.

Kostoff, R.N., del Rio, J.A., Humenik, J.A., Garcia, E.O., & Ramirez, A.M. (2001). Citation mining: Integrating text mining and bibliometrics for research user profiling. Journal of the American Society for Information Science and Technology, 52(13), 1148–1156.

Leydesdorff, L., & Rafols, I. (2008). A global map of science based on the ISI subject categories. Journal of the American Society for Information Science and Technology, 60(2), 348–362.

Li, Y., Zhang, C., & Swan, JR. (2000). An information filtering model on the Web and its application in JobAgent. Knowledge-Based Systems, 13(5), 285–296.

Li, Y., Zhou, X., Bruza, P., Xu, Y., & Lau, R.Y.K. (2012). A two-stage decision model for information filtering. Decision Support Systems, 52(3),706–716.

Liang, T.P., Yang, Y.F., Chen, D.N., & Ku, Y.C. (2008) A semantic-expansion approach to personalized knowledge recommendation. Decision Support Systems, 45, 401–412.

Liu, B., Wang, L., Jin, Y., & Huang, D. (2006). An effective PSO-based memetic algorithm for TSP. Intelligent Computing in Signal Processing and Pattern Recognition, 345, 1151–1156.

Manning, C.D., Raghavan, P., & Schutze, H. (2008). Introduction to information retrieval. Cambridge: Cambridge University Press.

McNee, S. (2006). Meeting user information needs in recommender systems. PhD thesis, University of Minnesota.

McNee, S.M., Albert, I., Cosley, D., Gopalkrishnan, P., Lam, S.K., Rashid, A.M., Konstan, J.A., & Riedl, J. (2002). *On the recommending of citations for research papers*. Proceedings of the 2002 ACM conference on Computer supported cooperative work, New Orleans, Louisiana, USA.

Mitchell, T.M. (1997). Machine learning (McGraw-Hill Series in Computer Science). New York: McGraw-Hill Higher Education.

Mostafa, J., & Lam, W. (2000). Automatic classification using supervised learning in a medical document filtering application. Information Processing & Management, 36(3), 415–444.

Newman, M.E.J. (2001). The structure of scientific collaboration networks. Proceedings of the National Academy of Sciences, 98(2), 404–409.

Nihalani, P.K., & Mayrath, M.C. (2008). Educational psychology journal editors' comments on publishing. Educational Research Review, 20(1), 29–39.

Pu, P., Chen, L., & Hu, R. (2011). A user-centric evaluation framework for recommender systems. Proceeding of the ACM RecSys 2010 Workshop on User-Centric Evaluation of Recommender Systems and Their Interfaces (pp. 14–21). Barcelona, Spain.

Ratnayaka, R.K.T., Wang, Z.J., Anamalamudi, S., & Cheng, S. (2012). Enhanced greedy optimization algorithm with data warehousing for automated nurse scheduling system. E-Health Telecommunication Systems and Networks, 1(4), 43–48.

Reyes-Sierra, M., & Coello, C.A.C. (2006). Multi-objective particle swarm optimizers: A survey of the state-of-the-art. International Journal of Computational Intelligence Research, 2(3), 287–308.

Robertson, S., & Soboroff, I. (2002). The TREC 2002 filtering track report. In Proceedings of the Tenth Text REtrieval Conference (TREC 2001; pp. 26–37), Gaithersburg, MD.

Shani, G., & Gunawardana, A. (2011). Evaluating recommendation systems. Recommender Systems Handbook, 257–297.

Strzalkowski, T. (1994). Robust text processing in automated information retrieval. In Proceedings of ACL-Sponsored Workshop on Very Large Corpora. Columbus: Ohio State University.

Sun, Y.H., Ma, J., Fan, Z.P., & Wang, J. (2008). A group decision support approach to evaluate experts for R&D project selection. IEEE Transactions on Engineering Management, 55(1), 158–170.

Turban, E., Zhou, D., & Ma, J. (2004). A group decision support approach to evaluating journals. Information & Management, 42(1), 31–44.

VandenBos, G.R., Appelbaum, M., Comas-Diaz, L., Cooper, H., Light, L., Ornstein, P., & Tetrick, L. (2010). Publication manual of the American Psychological Association. Washington, DC: American Psychological Association.

Vellino, A. (2010, October). A comparison between usage-based and citation-based methods for recommending scholarly research articles. Proceedings of ASIS&T 2010 (pp. 1–2). Pittsburgh, PA.

Vivacqua, A.S., Oliveira, J., & De Souza, J.M. (2009). i-ProSE: Inferring user profiles in a scientific context. The Computer Journal, 52(7), 789–798.

Wang, F., Shi, N., & Chen, B. (2010). A comprehensive survey of the reviewer assignment problem. International Journal of Information Technology & Decision Making, 9(4), 645–668.

Watanabe, S., Ito, T., Ozono, T., & Shintani, T. (2005). A paper recommendation mechanism for the research support system papits. In Proceedings of the International Workshop on Data Engineering Issues in ECommerce (pp. 71–80).

Yang, Y., & Pedersen, J.O. (1997). A comparative study on feature selection in text categorization. In Proceedings of ICML-97, 14th International Conference on Machine Learning (pp. 412–420). Nashville, TN.

Zheng, Z., Chen, K., Sun, G., & Zha, H. (2007). A regression framework for learning ranking functions using relative relevance judgments. In Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 287–294).