Information Processing and Management 59 (2022) 103028

Contents lists available at ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/ipm

Reviewer assignment algorithms for peer review automation: A survey☆

Xiquan Zhao, Yangsen Zhang *

Beijing Information Science and Technology University, NO.12 Xiaoying East Road, Qinghe, Haidian District, Beijing, 100192, China

ARTICLE INFO

Keywords:
Matching degree
Information retrieval
Reviewer assignment problem
Optimization algorithm
Natural language processing
Peer review

ABSTRACT

Assigning paper to suitable reviewers is of great significance to ensure the accuracy and fairness of peer review results. In the past three decades, many researchers have made a wealth of achievements on the reviewer assignment problem (RAP). In this survey, we provide a comprehensive review of the primary research achievements on reviewer assignment algorithm from 1992 to 2022. Specially, this survey first discusses the background and necessity of automatic reviewer assignment, and then systematically summarize the existing research work from three aspects, i.e., construction of candidate reviewer database, computation of matching degree between reviewers and papers, and reviewer assignment optimization algorithm, with objective comments on the advantages and disadvantages of the current algorithms. Afterwards, the evaluation metrics and datasets of reviewer assignment algorithm are summarized. To conclude, we prospect the potential research directions of RAP. Since there are few comprehensive survey papers on reviewer assignment algorithm in the past ten years, this survey can serve as a valuable reference for the related researchers and peer review organizers.

1. Introduction

Peer review is an activity in which some experts in one or several fields jointly evaluate the scientific research proposals or achievements. In modern society, peer review is firmly embedded in the practice of scientific research, it is the cornerstone of today's academic research landscape spanning submissions to journals, academic conferences and funding bodies across numerous disciplines (Price & Flach, 2017). Nowadays, peer review is regarded favorably by a significant majority of researchers, and it is the gateway process and quality control mechanism to both improve the quality of published research and validate the legitimacy of research publications (Fiez, Shah, & Ratliff, 2020; Rogers & Augenstein, 2020; Xu, Zhao, Shi, & Shah, 2019).

While part of the academic peer review process has been streamlined in the last few decades, there are still several challenges that arise in peer review relating to the integrity of the review process, one of them is how to automatically assign a number of reviewers to the well-suited papers at the same time, which is the key step to improve the quality of peer review. With the development of science and technology, the number of papers submitted to academic conferences or journals has also been growing rapidly (Bornmann & Mutz, 2015; Tabah, 1999), which make it increasingly difficult to assign reviewers to papers manually. As shown in Fig. 1, the number of manuscripts received by NeurIPS and AAAI in 2020 is 5.6 and 5.5 times respectively than that in 2014. The tremendous growth in the number of submissions further increase the necessity and urgency of peer review automation. Therefore, improving the automation level of peer review has gained increasing attention from researchers from different disciplines in recent decades (Shah, 2021).

☆ This work was supported by National Natural Science Foundation of China (Grant No. 62176023).

* Corresponding author.
E-mail addresses: zhaoxiquan@bistu.edu.cn (X. Zhao), zhangyangsen@163.com (Y. Zhang).

https://doi.org/10.1016/j.ipm.2022.103028
Received 9 March 2022; Received in revised form 4 July 2022; Accepted 11 July 2022
Available online 3 August 2022

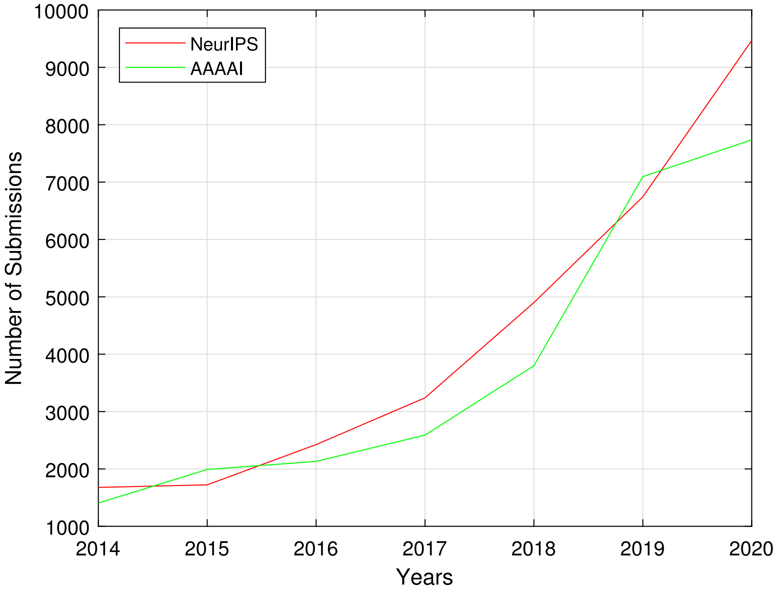

Fig. 1. The number of submissions to NeurIPS and AAAI from 2014 to 2020.

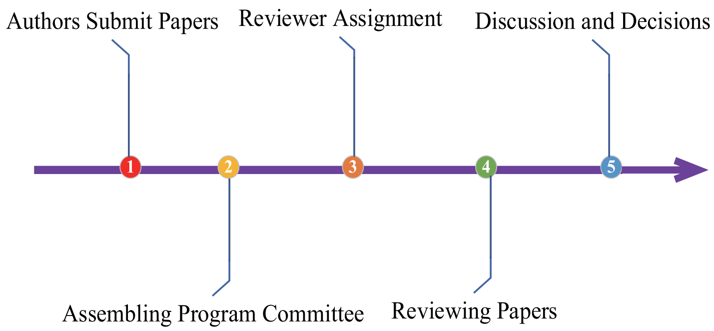

Fig. 2. A chronological summary of the academic conference paper review process.

To illustrate the importance of reviewer assignment problem, we take the review process of academic conference paper as an example. As shown in Fig. 2, the whole process of academic paper review can be divided into the following stages:

(1) Authors submit papers before the deadline.

(2) The Program Chair assembles a program committee. This is an expert finding problem, i.e., finding enough competent experts for the papers to be reviewed.

(3) The Program Chair assigns papers to the review experts in the program committee.

(4) Review experts evaluate the papers and return review comments.

(5) Based on the review comments, the Program Chair finally decides whether to accept each paper or not respectively.

As can be inferred from Fig. 2, the assignment result of reviewers to papers determines the expertise of the reviewer who will review the paper, hence it is the crucial step in the whole peer review process. As stated by Rodriguez, Bollen, and Van de Sompel (2007): “one of the first and potentially most important stage is the one that attempts to distribute submitted manuscripts to competent referees”. Bence and Oppenheim analyzed and compared all the stages of traditional peer review and the web-based peer review process, they believed that the most important stage was to assign the submissions to reviewers with sufficient expertise, which directly determined the final peer review quality (Bence & Oppenheim, 2004). Therefore, there is a huge demand for efficient reviewer assignment algorithms to realize automatic and accurate assignment of reviewers to papers, which can not only significantly raise the efficiency of the whole peer review process, but also can improve the peer review quality greatly with the help of the computer techniques such as natural language processing. Over the past few decades, many reviewer assignment systems have been proposed, e.g., the Toronto Paper Matching System (TPMS) (Charlin & Zemel, 2013), SubSift (Flach et al., 2009), Global Review Assignment Processing Engine (GRAPE) (Di Mauro, Basile, & Ferilli, 2005), Erie (Li & Hou, 2016), Reviewer Assignment

System (RAS) (Kou et al., 2015), decision support system (Hoang, Nguyen, & Hwang, 2019), OpenReview[1] and Microsoft Conference Management Toolkit (CMT).[2] These systems are completely automated and have been used in the review process of many academic conferences such as NIPS, ICML, and ACL.

Automatically assigning reviewers to papers is an evolving inter-disciplinary problem, which has attracted a large number of researchers from different disciplines such as computer science, artificial intelligence and operations research. Natural language processing is the first driving force for automatic reviewer assignment. With the invention of Latent Semantic Indexing (LSI) (Deerwester, Dumais, Furnas, Landauer, & Harshman, 1990; Dumais & Nielsen, 1992) explored the automatic reviewer assignment problem. In different literature, researchers adopted some other terms to refer to this problem, such as Paper–Reviewer Assignment (PRA) (Long, Wong, Peng, & Ye, 2013), Reviewer Assignment Problem (RAP) (Nguyen, Sánchez-Hernández, Agell, Rovira, & Angulo, 2018; Wang, Shi and Chen, 2010; Wang, Zhou, & Shi, 2013), Committee Review Assignment (CRA) (Karimzadehgan & Zhai, 2012) and Conference Paper Assignment Problem (CPAP) (Goldsmith & Sloan, 2007; Lian, Mattei, Noble, & Walsh, 2018). For the sake of consistency, this paper adopts the most popular Reviewer Assignment Problem (RAP) to represent the problem. For the same reason, the submissions (e.g., academic papers, scientific research project documents, and research funding proposals) are represented by papers, and experts in various peer review scenarios are represented by reviewers uniformly.

Several survey papers on reviewer assignment algorithm have been published, but they cannot reflect the latest research progress in this field. First of all, several surveys were published more than ten years ago, thus they cannot cover the latest reviewer assignment algorithms (Goldsmith & Sloan, 2007; Kalmukov & Rachev, 2010; Kolasa & Krol, 2011; Wang, Shi et al., 2010). Second, some surveys focus on only one or two aspects of reviewer assignment algorithm, which is not comprehensive to present overall research perspective (Goldsmith & Sloan, 2007; Patil & Mahalle, 2020). To fill the research gap, this paper aims to provide a comprehensive review of the primary research achievements on reviewer assignment algorithm from 1992 to 2022, and makes the following contributions:

- We conducted the most comprehensive survey on reviewer assignment algorithm in the past ten years.
- We systematically summarized the datasets and metrics for reviewer assignment algorithm evaluation.
- The future directions of reviewer assignment algorithm were prospected in this survey.

The remainder of this paper is structured as follows. Section 2 introduces our survey methodology in detail. Section 3 divides reviewer assignment algorithms into three stages, then presents the research progress of all the stages respectively. Section 4 introduces the performance evaluation methods of reviewer assignment. Future research directions of RAP are prospected in Section 5. We discuss the findings and implications of this survey in Section 6, and Section 7 concludes the paper.

2. Survey methodology

We focus on the algorithms that were designed to address RAP from all perspectives, and we adopted a systematic approach to conduct literature survey in this study. The approach included two strategies, i.e., publication search strategy and publication selection strategy, whose details will be introduced in the following two subsections. With this methodology, we are confident that there is very low risk of introducing researcher's bias into the survey.

2.1. Publication search strategy

There are two issues to determine before conducting publication search, that is, selecting literature sources and designing search keywords, both of them are crucial for accuracy and comprehensiveness of literature search results.

We first chose four well-known online literature sources indexing publications of Computer Science, Artificial Intelligence and Operational Research, i.e., IEEE Xplore, ACM Digital Library, SpringerLink and Elsevier ScienceDirect. In addition, we used the popular Google Scholar search engine to find out more relevant publications included in other databases.

With regard to the search keywords, we chose several synonyms of RAP to find the relevant publications from the literature sources, including "Reviewer Assignment Problem", "Paper–Reviewer Assignment", "Committee Review Assignment" and "Conference Paper Assignment Problem". Note that no publication time span was set during this search process.

2.2. Publication selection strategy

In order to ensure the quality and credibility of the selected papers, several inclusion and exclusion criteria were considered to decide whether a publication is suitable for our survey or not. The inclusion criteria are as follows:

- Publications should relate directly to the RAP topic. The relevance degree of the publications can be ensured by readings their titles, abstracts and even the full-text.
- Publications should be published online from 1992 to 2022.

Likewise, the following exclusion criteria were designed to help us exclude the inappropriate literatures:

- Publications that are not available in English.
- Duplicated publications.

[1] https://openreview.net/
[2] https://cmt3.research.microsoft.com/

Table 1
Summary of reviewer database construction methods.

Methods	Description	Related works
Online-recruiting method	Expert information in database was entered online by experts themselves	Campbell, Maglio, Cozzi, and Dom (2003) and Davenport and Prusak (1998)
Experience-based method	Expert information was collected according to the experience or social relationships of peer review organizers	Kautz, Selman, and Shah (1997) and Zhang, Tang, and Li (2007)
Dynamic reviewer database construction method	Expert information can be updated dynamically with multi-source data from internet based on web crawler	Charlin and Zemel (2013) and Tang (2016)

• Publications that have not been peer reviewed, such as preprints, working papers or technical reports.

To reduce the risk of omitting relevant papers, the famous snowballing technique (Wohlin, 2014) was also used in our study. Snowballing is a time-consuming but extremely effective technique, especially the iteration of backward and forward procedure, which assisted us in finding a number of important papers omitted by keywords searching and thus guaranteed the credibility of our study.

3. Key techniques of reviewer assignment algorithms

According to the timeline of RAP, the reviewer assignment algorithm can be divided into the following three stages:

(1) Constructing the candidate reviewer database.

(2) Computing the matching degree between every paper–reviewer pair.

(3) Based on the matching degree matrix, the reviewer assignment scheme is generated.

Since the candidate reviewer database is the basis of subsequent stage of reviewer assignment algorithm, we regard its construction procedure as the first stage of RAP in this survey. More specifically, the candidate reviewers do not have to be the most competent experts for every paper, but must have sufficient matching degree with part of the papers.

In stage 2, based on the database of candidate reviewer, some criteria are needed to select suitable reviewers, and the most widely used one is the matching degree between each paper–reviewer pair. In different literature, matching degree is also represented by some other terms, such as similarity, affinity, relevance, fitness, confidence degree and suitability score, and herein we will use these terms interchangeably throughout this survey according to its expression method in the source literature. As will be discussed in Section 2, matching degree can be calculated by several algorithms in natural language processing.

In stage 3, the reviewer assignment scheme is generated based on the matching degree and some other constraints in different review scenarios, such as the workload balance between reviewers and coverage of the papers. Therefore, RAP should be modeled according to different constraints and is solved by different assignment optimization algorithms.

As an important principle in the peer review process, conflict of interest (COI) between reviewer and the authors of papers should be identified and avoided. However, we found that different researchers may address the COI issue in different stages, that is, it may be examined in stage 1 or stage 2, or some algorithms may even address it in stage 3. For this particular issue, we will discuss the COI issue separately in the last part of this section.

3.1. Construction of candidate reviewer database

Constructing a comprehensive and accurate candidate reviewer database is the basis of reviewer assignment algorithm, it belongs to the research domain of expert finding (Deng, King, & Lyu, 2008; Yimam-Seid & Kobsa, 2003). However, most of the literature on RAP usually assumes the database is pre-existing, thus they do not discuss the reviewer database constructing process. For example, most of the conference management software do not mention how to construct their candidate reviewer database (i.e., the program committee) from scratch. Instead, they just focus on the review management workflow such as sending and receiving review invitations.

In the real world, many reviewer databases are mainly constructed based on static information, e.g., the information manually entered by candidate reviewers, there is still much room for improvement in the accuracy and comprehensiveness of the reviewer databases. In this section, we will mainly discuss three kinds of reviewer database construction methods, as we summarized in the Table 1.

With respect to some expert recommendation systems, the candidate reviewers are generally recruited online by peer review organizer (Campbell et al., 2003; Davenport & Prusak, 1998). Usually, many experts may receive the invitation email, but only part of them will respond to the email and finally enter their personal information into the recruiting system. This kind of expert finding method has the following disadvantages. First, the information entered by experts themselves is relatively subjective, hence its completeness and accuracy may be not good enough. Second, since the information is manually entered into the expert database, it cannot be updated in time and reflect the latest research fields of experts. Third, the method of collecting information manually does not make full use of the expert information on the Internet, resulting in the relative lack of information. In a word, the online recruiting method may generate an expert database that is not accurate and comprehensive enough, which will lead to unsatisfying peer review result.

In addition, some peer review organizers tend to find candidate reviewers according to their own experience or social relationships (Kautz et al., 1997; Richards, Taylor, & Busch, 2008; Zhang et al., 2007). In some peer review scenarios, the authors who have published papers in a journal (or academic conference) will be automatically added into the corresponding expert database, and some projects or funding management agencies prefer treating the applicants who was funded as the candidate reviewers. For example, based on the idea that "those eligible to apply to NSF are eligible to review", National Science Foundation (NSF) automatically adds researchers who were once funded by NSF to candidate reviewer database (Hettich & Pazzani, 2006). In this way, although the competency of candidate reviewers can be guaranteed, it may lead to a shortage of experts in some cases.

Given the limitations of candidate reviewer database construction methods mentioned above, the construction of dynamic reviewer database based on multi-source heterogeneous data has gained increasing attention from many researchers, its key techniques include the crawling of candidate reviewer information, and disambiguation and alignment of reviewers' names.

3.1.1. Candidate reviewer information crawling

There is a huge amount of expert information on the Internet. Data from some popular academic websites (e.g., academia.edu, Google Scholar, Microsoft Academic Search and Elsevier's Reviewer Finder) is public and easy to obtain, hence these websites mentioned above are usually called "shallow" websites. In contrast, some academic websites (e.g., PubChem and Science.gov) contain approximately hundreds of times expert information (Bergman, 2001). However, there are two difficulties in collecting data from these "deep" websites. First, data from different web pages may have different data formats (e.g., web page text, social media text, Word documents, PDF documents, or PPT documents). Second, some websites are not visible to search engines, and part of them are even configured with anti-crawling strategy. Therefore, data crawling technology for multi-source data is indispensable for collecting expert information from the deep websites, and it has been developing rapidly in recent years (Desai, Devulapalli, Agrawal, Kathiria, & Patel, 2017; Shkapenyuk & Suel, 2002).

Hoang, Nguyen, Collins, and Hwang (2021) used Scrapy crawler to collect data from various sources to generate a database of scientist profiles, and the academic data included information about authors, coauthors, articles, citations, academic networks and so on. Also, the multi-source reviewer information crawling technology has been applied in some commercial conference management software. For instance, the TPMS, which has been integrated with widely used Microsoft CMS and adopted in several top-tier academic conferences in the field of machine learning and computer vision (e.g., NeurIPS, ICML, CVPR, ICCV), collects lots of information of reviewers automatically. The expert finding method of TPMS is very representative, it constructs the expert database based on both the static and dynamic data. For the static data, experts who received the review invitation log in TPMS to upload their representative publications, or provide corresponding URL identifications, TPMS will crawl the URL to obtain the documents. Besides, TPMS also dynamically crawl the relative websites (e.g., Google Scholar) to update the information of candidate reviewers. Based on these multi-source heterogeneous data, TPMS can finally construct the database with rich and accurate information, and the database can also be reused by other academic conferences in the similar research field.

3.1.2. Name disambiguation and alignment

During the construction process of candidate reviewer database, there may be cases where some experts' names are ambiguous. The ambiguity of expert names is generally caused by two reasons. First, one expert entity has multiple names, e.g., besides his (or her) real name, the expert may also has alias or pen name, or his (or her) real name may have different spellings in foreign languages. Second, one name may refer to multiple expert entities in the real world. For instance, different experts have the same name. Therefore, name disambiguation and alignment are both the core steps in the candidate reviewer database construction.

There is already plenty of research work on name disambiguation and alignment. AMiner (Tang, 2016) is a famous academic search engine and social network mining platform, formerly known as ArnetMiner (Tang et al., 2008). AMiner establishes millions of multi-dimensional and high-precision expert portraits, which contain experts' educational experience, expertise, published papers, granted patents, gender, and nationality (Zhang, Zhang, Yao, & Tang, 2018). To construct the database, AMiner utilized a name disambiguation algorithm based on the probability framework of Hidden Markov Random Fields (HMRF) (Basu, Bilenko, & Mooney, 2004). In this way, AMiner was not only used to recommend academic resources such as papers and patents, but also play an important role in the process of talent introduction and expert recommendation tasks for some grant agencies (Pradhan, Sahoo, Singh, & Pal, 2021).

3.2. Computation of matching degree between papers and reviewers

Matching degree can indicate the relevance between paper and reviewer's expertise, and it is very suitable to evaluate the competence level of the assigned reviewers. The higher the matching degree, the more competent the reviewer is to evaluate the paper, thus the review quality will be better. Matching degree is a crucial factor to solve RAP, it includes two core standards (Tang, Tang, & Tan, 2010). One standard is its literal meaning, that is, the similarity between reviewer's knowledge and the content of paper, it will be finally transformed to the calculation of text similarity between them. Another standard is the COI between reviewer and the author of the paper. If there is any cooperative or competitive relationship between the two parties, their matching degree should be set to 0. COI will be discussed separately in a subsequent section, thus we will not discuss it here.

There are mainly three considerations for the matching degree computation stages in RAP. First of all, some researchers believed that the subjective ratings of experts were credible and they took expert's review willingness score as the matching degree between papers and experts. Secondly, some other researchers thought that the text content of papers and the publications of experts were more objective and should be used to calculate the matching degree between them. Therefore, they used various NLP techniques

Table 2
Summary of matching degree computation methods.

Methods	Description	Related works
Subjective scoring method	Matching degree is evaluated by the review willingness such as bidding score	Cabanac and Preuss (2013), Fiez et al. (2020), Rodriguez et al. (2007) and Shah, Tabibian, Muandet, Guyon, and Von Luxburg (2018)
Method based on text information	Matching degree is calculated according to the information mined from the text content, e.g., the word frequency information, topic information or other deep semantic information	Charlin and Zemel (2013), Conry, Koren, and Ramakrishnan (2009), Dumais and Nielsen (1992), Ferilli, Mauro, Basile, Esposito, and Biba (2006), Hettich and Pazzani (2006), Karimzadehgan, Zhai, and Belford (2008), Kou, Leong Hou, Mamoulis and Gong (2015), Li and Hou (2016), Mimno and McCallum (2007), Tang and Zhang (2008), Yarowsky and Florian (1999), Zhao et al. (2018) and Zhang et al. (2020)
Method based on multiple information	Besides text information of papers and reviewer's publications, other types of information was also utilized synthetically	Di Mauro et al. (2005), Liu, Suel, and Memon (2014) and Medakene, Bouanane, and Eddoud (2019)

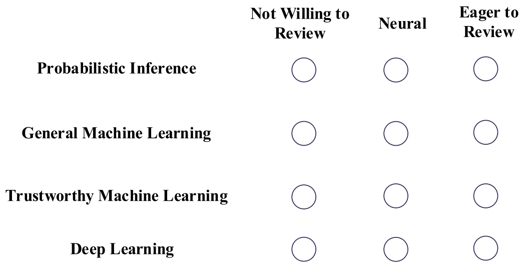

Fig. 3. A toy example of graphical interface of bidding for reviewers.

to unearth word or semantic information from the text of both sides, and then matching degree was calculated according to those information. Finally, some researchers thought that it was not accurate enough to determine whether an expert was suitable for reviewing a paper just based on the text information, and the more accurate matching degree could be obtained when considering some information outside the text comprehensively, such as the authority or the character (stringent or lenient) of experts. Therefore, based on these different considerations, the methods of matching degree computation between papers and reviewers can be divided into three categories, that is, subjective scoring method, method based on text information and method based on multiple information. A summary of the matching degree computation methods is given in Table 2.

3.2.1. Subjective scoring method

For subjective scoring method, reviewers are required to browse the list of papers (or abstracts), and score each paper separately with a discrete value, which can be used to represent their review willingness. One simple example is 0 to indicate the reviewer is unwilling to review the paper, and 1 to indicate the opposite. There are also some more sophisticated representations. A typical application scenario is the bidding process in the review process of academic conference papers (Beverly & Allman, 2012; Cabanac & Preuss, 2013; Fiez et al., 2020; Meir, Lang, Lesca, Mattei, & Kaminsky, 2021; Rodriguez et al., 2007; Shah et al., 2018). Bidding is also knows as collecting reviewers' preference, or rating papers, it is adopted by many conference management software and grant agencies. For example, to determine the expertise level of reviewers in the project review process of National Natural Science Foundation of China (NSFC), reviewers are required to fill in a form to declare their expertise level in some disciplines, with some publications as the supporting evidence. The expertise level of the reviewers is measured with the level score 1 3, where level 3 signifies that a reviewer is very familiar with the corresponding research area, level 2 familiar and level 1 less familiar. In this way, the matching degree between a certain reviewer and a proposal can be represented by the expertise level illustrated above (Xu et al., 2010).

An example sample graphical interface of bidding is shown in Fig. 3. Experts express their review willingness by selecting the options in the graphical interface. Each option in Fig. 3 corresponds to an integer, a bigger integer value means a stronger review willingness. If a paper is not rated, then the integer number referring to "Neural" option will be assigned as the default value.

The bidding method has an obvious advantage. For the conscientious and responsible experts, they know best whether they are competent to review the papers or not, thus the credibility of their bidding scores is relatively high. However, bidding method has some inherent drawbacks. First, the bidding process is too tedious and time-consuming for the reviewers. Due to the large number of papers, it takes lots of time to score all the papers. Second, if too many papers are not rated, the reviewer–paper bidding result will be a sparse matrix, which will significantly reduce the accuracy of reviewer assignment scheme. Third, some reviewers frequently conflate expertise and interest, or ignore expertise altogether. They may bid based on their current research interest rather than their expertise, which will introduce a certain amount of noise in the bidding score, and finally lead to inaccurate review assignment results. To handle this problem, some efforts have been made to improve the accuracy of the bidding scores. For instance, through an in-depth analysis of the incentives of reviewers during the bidding phase, Meir et al. (2021) proposed a market-inspired bidding scheme to guide reviewers to bid sincerely, which could be used to improve both the bid distribution and the resulting assignment.

3.2.2. Method based on text information

In this section, we will discuss some NLP models that were used to calculate matching degree between papers and reviewers. According to the different types of information that was extracted, the NLP models can be divided into three categories, namely word information models, topic models and the neural network models.

3.2.2.1 Word information models In some early literature, researchers first generated the vector representations of reviewers and papers based the word information in the papers and publications of experts, and then calculated their matching degree according to the vector representations. The following three methods has been applied in RAP field, namely keyword matching, Bag of Words (BOW) model, and Language Model (LM).

In terms of keyword matching, Tang and Zhang (2008) constructed a keyword network, then used cosine similarity to calculate the distance between reviewer and paper's keyword vector, which was used to represent their matching degree. Protasiewicz (2014) extracted keywords based on expert knowledge and get the keyword classified, then calculated the weight in combination with paper's publication time, and also leveraged the cosine similarity to calculate the matching degree.

With respect to the BOW model, Yarowsky and Florian (1999) leveraged the TF-IDF algorithm for the first time, and they measured the similarity between reviewer and the paper by the cosine distance. Hettich and Pazzani (2006) adopted TF-IDF algorithm to get the key information of project document and reviewer's publications, then the matching degree was calculated by the key information. Abduljaleel, Naser, and Al-mamory (2021) first divided each paper of authors and reviewers into five sections (i.e. Title, Abstract, Keywords, References, and the rest text in the paper), then they adopted TF-IDF and cosine similarity techniques to calculate the matching degree between each reviewer-author pair.

LM also has some applications in the expert finding field. Balog, Azzopardi, and de Rijke (2006) utilized a n probabilistic language model to evaluate the suitability of experts from the document repositories. A typical example is in the famous paper–reviewer assignment system TPMS, Charlin and Zemel (2013) proposed a word-LM method to get the matching degree between reviewers and papers.

Though the methods based on word information can characterize word-level frequency characteristics of the papers, it cannot portray the semantic information, which is abstract and more important for the RAP. For example, variants or synonyms with the same semantics cannot be matched based on the word information. Therefore, the matching degree calculated by these models may have large deviation in some cases.

3.2.2.2 Topic models Since RAP was raised, several topic models have been used to characterize papers and reviewers. There are several semantic models that has been used more often to solve RAP, i.e., LSI, Probabilistic Latent Semantics Analysis (PLSA) (Hofmann, 1999) and Latent Dirichlet Allocation (LDA) (Blei, Ng, & Jordan, 2003).

LSI is the first semantic model applied in matching degree computation in RAP. Dumais and Nielsen (1992) were the first to use LSI to extract topic information of papers and reviewer's publications, then leveraged cosine similarity to evaluate their matching degree. Ferilli et al. (2006) also used LSI to extract topics from the title and abstract of submissions, and obtain the expertise of the reviewers from the titles of their publications. Finally, the topics associated to papers and reviewers were fed to GRAPE to perform assignments. LSI has also been widely used in the paper review process in academic conference. Erie, the reviewer assignment system of IEEE INFOCOM 2015 and INFOCOM 2016, adopted LSI to calculate the suitability score between the papers and the candidate reviewers from Program Committee (Li & Hou, 2016). The computation process of LSI is very simple, it only takes one Singular Value Decomposition (SVD) (Kalman, 1996) operation to obtain the topic information. However, LSI has some disadvantages. First, SVD is computational intensive and the running speed is relatively slow. Second, the number of topics is difficult to determine. Third, LSI is not a probabilistic model, thus the topics obtained by it are always difficult to understand intuitively. As a result, LSI has been rarely used in the latest literature on RAP.

Compared with LSI, PLSA leverages a generative model and has better interpretability. PLSA assumes that each document contains a series of possible potential topics, and every word in the document is generated with a certain probability based on the distribution of the topics. Karimzadehgan et al. (2008) were the first to leverage PLSA to extract topics from papers and reviewer's publications, then the similarity of every reviewer–paper pair could be measured by their topic distance. Conry et al. (2009) also used PLSA to obtain the topic of the papers, and further integrated the information based on multiple sources to learn the reviewer–paper similarity. However, PLSA model also has two disadvantages. First, it cannot generate new unknown documents. Second, it may generate an overfitting model when the number of documents is large, hence leading to an inaccurate similarity. These two factors both limit the further application of PLSA model in the RAP field.

After PLSA model, the LDA model and some variants were proposed and promoted the research of RAP (Xu & Wang, 2011). For LDA model, each document has a distribution over topics. To generate a word, a topic should be chosen first from the document's topic distribution, then the word will be selected from the topic's distribution over the vocabulary. In the popular TPMS, the publications of reviewers were concatenated into a single document to create reviewer's archive, then LDA model was used to extract abstract features from both reviewer's archive and the submissions, finally their matching degree can be measured by the metrics such as dot product and KL-divergence (Charlin & Zemel, 2013). Besides the direct application of LDA model (Kim & Lee, 2018; Nguyen et al., 2018; Xu & Zuo, 2016; Yang, Liu, Yi, Chen, & Niu, 2020), some variants of LDA were also widely used to carry out the paper–reviewer matching degree calculation, the Author Topic Model (ATM) (Rosen-Zvi, Griffiths, Steyvers, & Smyth,

2004) was a representative example. Jin, Geng, Zhao, and Zhang (2017) considered three dimensions, namely the relevance between reviewers and submissions, the interest trend of reviewers, and the authority of reviewers, then leveraged Expectation Maximization (EM) algorithm and ATM to evaluate the topic distributions of reviewer' publications and the papers. Kou, Leong Hou, Mamoulis et al. (2015) also modeled the topic distributions of papers and reviewers' publications by ATM, a group of experts were eventually recommended according to their topic weights. Meanwhile, various topic models have extended the ATM by introducing different kinds of information, and some of them were utilized to solve RAP. On the basis of ATM, Mimno and McCallum (2007) proposed the Author-Persona-Topic (APT) model to match papers with the expertise of reviewers. In the APT model, each author can write under one or more "personas", and each persona is represented as independent distributions over hidden topics calculated by the LDA model. Inspired by the document class layer in AIT (Author-Interest-Topic) model (Kawamae, 2010; Mou, Geng, Jin, & Chen, 2015) proposed the Author–Subject–Topic (AST) model, which introduced the 'Subject' layer into the ATM and was used to infer an author's interest with external metadata, then the interest can be used to calculate the matching degree between papers and reviewers. This work was then extended by Jin, Geng, Mou, and Chen (2019), the major differences from the original AST lie in the modeling method of the subject information of each document, the enhanced AST model was also used to address the difficulty in recommending reviewers for interdisciplinary papers.

Some researchers used both word information and topic information to calculate matching degree. Tan, Duan, Zhao, Chen, and Zhang (2021) combined language model and LDA topic model to calculate the matching degree between papers and reviewers. Peng, Hu, Wang, and Wang (2017) believed that the experts' publications closer to now reflect their current research directions better, so they proposed a time-aware and topic-based assignment model, which took statistical characteristics such as TF-IDF into consideration to model the matching degree better. Tan et al. (2021) proposed a word and semantic-based iterative model named WSIM to calculate matching degree. WSIM leveraged optimized language model and LDA model to extract both word and semantic features of reviewers and papers, thereby improving the accuracy of matching degree. Pradhan et al. (2021) leveraged LDA and term frequency-inverse reviewer frequency (TF–IRF) to model research interest of reviewers, then the topic network of the decision support system was constructed to recommend suitable reviewers.

Compared with the methods based on word information, topic models could unearth deeper semantic information of texts, thus the similarity will be more accurate. However, because of its inherent assumptions such as the independence of semantically related words, these above-mentioned semantic models cannot make full use of the implicit semantic association between different words belong to the same topic (Xie, Yang, & Xing, 2015). Moreover, probabilistic topic models such as LDA require a large corpus to accurately identify the topics and their distributions in each document, which may be problematic when applied to short documents such as abstracts (Anjum, Gong, Bhat, Hwu, & Xiong, 2019).

3.2.2.3 Neural network models In recent decades, deep learning has been booming in the field of natural language processing, various neural network models have been adopted to extract deep semantic features from text. With the neural network models, a set of low-dimensional dense vectors can be obtained as the semantic representation of text, then the similarity between different texts can be calculated by some methods such as cosine similarity. Currently, several neural network models have been used to compute text similarity, i.e., word2vec (Liu, Wang, & Zhu, 2022; Mikolov, Sutskever, Chen, Corrado, & Dean, 2013), doc2vec (Le & Mikolov, 2014), Skip-Thought vector model (Kiros et al., 2015), Recurrent Neural Network (RNN) (Socher, Karpathy, Le, Manning, & Ng, 2014), long and short-term memory (LSTM) (Tai, Socher, & Manning, 2015), and Convolutional Neural Networks (CNN) (Kalchbrenner, Grefenstette, & Blunsom, 2014).

Zhao et al. (2018) regarded RAP as a classification problem. First, they used word2vec to generate the word embedding of papers and reviewers' publications. Then they leveraged Word Mover's Distance (WMD) to calculate the distance between papers and reviewers. Finally, the assignment of reviewers and papers was conducted based on constructive coverage algorithm. Ogunleye, Ifebanjo, Abiodun, and Adebiyi (2017) proposed a framework based on word2vec model to derive the matching degree between submissions and reviewer's representative publications. Wu (2018) proposed a Distributed Memory model named Paragraph Vectors (PV-DM) based on the doc2vec algorithm. PV-DM model was trained by a large number of scientific and technological documents (e.g., academic papers, patent and scientific research project documents), and then was used to generate the word embedding of scientific documents and reviewers' publications, then each reviewer was scored based on the distance of embeddings. Finally, KNN algorithm was used to recommend suitable reviewer for each paper. Zhang et al. (2020) believed that unsupervised embedding features lack discriminative semantic information of the papers and reviewers, so they proposed the hierarchical and transparent representation (abbreviated as Hiepar), which was learned from a two-level bidirectional gated recurrent unit based neural network with two-level attention mechanism. Hiepar could capture hierarchical structure information and generate semantic expression for reviewers and papers, thus generating more accurate similarity.

By using neural network models, the representation of document text does not have to rely on traditional feature engineering methods, and the deep semantic representation (e.g., word embedding) of papers and reviewers with higher accuracy could be obtained. However, for the sake of fairness and confidentiality, real-world reviewer assignment data in the peer review process is generally not publicly accessible, leading to the lack of real training data that is needed in RAP. As a result, the neural network model are usually trained by general text data, which affects the effect of feature extraction to a certain extent.

3.2.3. Method based on multiple information

The methods discussed above just utilize the bidding scores or the text information of papers and reviewers' publications to calculate their matching degree. In the real world, multiple information has been used by more and more researchers, not only the text (e.g., word and semantic features), but also some other factors outside the text (e.g., the co-author relationships) have been

Table 3
Summary of reviewer assignment optimization algorithms.

Methods	Description	Related works
Retrieval-based approach	Every paper is used as a query to independently retrieve a certain number of reviewers who are most relevant to it	Basu, Hirsh, Cohen, and Nevill-Manning (1999), Biswas and Hasan (2007), Dumais and Nielsen (1992), Hettich and Pazzani (2006), Karimzadehgan et al. (2008) and Mimno and McCallum (2007)
Matching-based approach	All the papers as a batch are assigned to the reviewers simultaneously, and the reviewer assignment scheme should subject to certain constraints	Conry et al. (2009), Hartvigsen, Wei, and Czuchlewski (1999), Hettich and Pazzani (2006), Hoang et al. (2021, 2019), Hofmann (1999), Huang et al. (2016), Jecmen et al. (2020), Jin et al. (2019, 2017), Jin, Niu, Ji, and Geng (2020), Kalchbrenner et al. (2014), Kalman (1996), Kalmukov (2020a, 2020b), Kalmukov and Rachev (2010), Karimzadehgan and Zhai (2009), Kobren, Saha, and McCallum (2019), Kou, Leong Hou, Mamoulis et al. (2015), Long et al. (2013) and Sidiropoulos and Tsakonas (2015)

utilized to improve the accuracy of matching degree (Medakene et al., 2019). The methods based on multiple information can be categorized as follows:

• Rule-based method. Di Mauro et al. (2005) simultaneously considered several factors such as the expertise of reviewers, the topics of paper and the bidding scores, then the confidence degree (similar to matching degree) of the reviewer–paper pair were calculated based on some pre-set rules.

• Collaborative filtering method. Tong (2018) first collected all the bidding scores on the papers, then leveraged collaborative filtering algorithm to predict the missing scores. Based on the bidding scores and semantic features extracted by feedforward neural network, the expert recommendation system was constructed and finally generated the expert list for the scientific and technological projects.

• Graph-based method. Taking the relationships in the social network of experts into consideration, Xu (2010) evaluated the matching degree based on their cooperation relationship and communication record, which was a good supplement to the methods based on text information. Liu et al. (2014) proposed a graph-based model to capture the academic connection between the reviewers, which effectively improved the accuracy of matching degree between the reviewer–paper pair. Yong, Yao, and Zhao (2021) first adopted Word2Vec to extract core semantic keywords, then the rule engine was established by means of the knowledge graph to reason the relevancy between reviewers and papers.

3.3. Reviewer assignment optimization algorithms

The reviewer assignment algorithm consists of three stages in total. According to the different optimization methods used in stage 3, i.e., the reviewer assignment optimization stage, reviewer assignment algorithm can generally be divided into retrieval-based methods and matching-based methods (Long et al., 2013). It is worth noting that for a complete reviewer assignment algorithm, a retrieval-based method and a matching-based method may share the same matching degree computation algorithm in stage 2, such as the TF-IDF, LDA or ATM. In the same way, we classified the third stage of reviewer assignment algorithm into two categories, namely the retrieval-based reviewer assignment optimization approach and the matching-based reviewer assignment optimization approach. For the sake of conciseness, we abbreviated them to retrieval-based approach and matching-based approach respectively.

Generally, the retrieval-based approach is usually applied in the one-to-several reviewer assignment scenarios, i.e., selecting several reviewers from the candidate reviewers database for one paper at each time, a common example is the reviewer assignment process of papers submitted to the journals. By contrast, the matching-based approach is usually used in the many-to-many reviewer assignment scenarios such as the conference paper review process, in which many reviewers should be assigned to many papers simultaneously. A summary of reviewer assignment optimization algorithms is given in Table 3.

3.3.1. Retrieval-based approach

Dumais and Nielsen (1992) were the first to regard the reviewer assignment process as an information retrieval (IR) problem. With respect to this IR problem, each paper is treated as a query, and the knowledge of each candidate reviewer is represented by a set of documents (e.g., the self-declaration of expertise provided by reviewer, and the representative publications of reviewers). Therefore, the reviewer assignment for one paper turns into a query operation from the candidate reviewer database, and a fixed number of reviewers who are most competent to review the paper will be obtained. According to the idea of information retrieval, a series of retrieval-based reviewer assignment algorithms have been proposed (Biswas & Hasan, 2007; Fang & Zhai, 2007; Pradhan et al., 2021; Zablocki & Lee, 2012), and the difference between these algorithms mainly lies in the representation models that were discussed in previous sections, such as LSI, vector space models (Kalmukov, 2020b), topic model (Mimno & McCallum, 2007), language model (Karimzadehgan et al., 2008) and hybrid methods (Basu et al., 1999).

Generally, the retrieval-based reviewer assignment optimization algorithm has its own disadvantages. First, since every paper is retrieved independently, it takes many founds for the algorithm to assign reviewers to all the papers, which will be too time-consuming when the number of papers is large. Second, in the real world, some reviewer usually cover a wider range of research

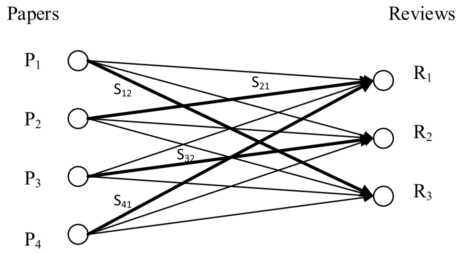

Fig. 4. A weighted bipartite graph constructed by the paper set and reviewer set.

fields and thus have relatively high expertise, these "popular" experts will be ranked relatively high in most of the rounds, which will lead to workload imbalance among reviewers. That is, the popular reviewers have too much papers to review, whereas the rest reviewers have fewer or even no papers. Third, the retrieved-based approach is beneficial to the papers retrieved earlier, the papers retrieved later may have to be assigned to the experts with lower expertise, which violates the fairness principle of peer review.

3.3.2. Matching-based approach

In recent years, a more appropriate formulation of the RAP would be to simultaneously optimize the assignments of all the papers to an entire committee of reviewers under certain constraints, more and more researchers regard RAP as a matching problem between the set of papers and reviewers (Charlin, Zemel, & Boutilier, 2011; Garg, Kavitha, Kumar, Mehlhorn, & Mestre, 2010; Hartvigsen et al., 1999; Karimzadehgan & Zhai, 2009; Merelo-Guervos & Castillo-Valdivieso, 2004; Mirzaei, Sander, & Stroulia, 2019; Tang et al., 2012). Instead of retrieving reviewers for each paper one by one, the matching-based reviewer assignment optimization algorithm first sets a global optimization objective, then it makes the matching all at once. Generally, the matching-based approach should satisfy the following constraints:

• Each paper must be reviewed by C_j reviewers. Here, the value of C_j can be set by the organizers of peer review.

• The number of papers assigned to each reviewer should not exceed U_i, and the value of U_i can also be set by the peer review organizers or reviewers.

• The number of papers assigned to each reviewer should be roughly the same, that is, the workload of the reviewers should be balanced.

• Each paper should be assigned to the reviewers with enough expertise in the research field of the paper.

• The reviewer is willing to review the paper assigned to him, and has no conflict of interest with it.

The matching-based approach can be illustrated as Fig. 4, where the nodes on the left correspond to papers and the nodes on the right refer to reviewers. Edges in bold denote a specified assignment scheme provided by the assignment algorithm. The weights associated with each edges indicate the strength of similarity between the papers and reviewers. Essentially, the matching-based RAP can be regarded as a Generalized Assignment Problem (GAP) of enhanced version, and the algorithms generally adopt the following paradigm (Kobren et al., 2019; Rodriguez & Bollen, 2008). First, a weighted bipartite graph is built, where the weight of the corresponding reviewer–paper edge is set to their matching degree, as shown in Fig. 4. Second, guided by the global optimization objective, different algorithms can be used to generate the optimal reviewer assignment scheme (Duan et al., 2019), such as genetic algorithm (Camara, Ortega, & de Toro, 2009; Ouni, Kula, & Inoue, 2016), ILP (Integer Linear Programming) (Charlin et al., 2011; Conry et al., 2009; Jin et al., 2020), simulated annealing-based stochastic approximation algorithm (Li, Cao, & Qu, 2017), network flow problem (Stelmakh, Shah, & Singh, 2021; Yan, Jin, Geng, Zhao, & Huang, 2017) and other heuristic algorithms (Kalmukov, 2020a; Pradhan, Chakraborty, Choudhary, & Nandi, 2020) . The matching-based approach is mainly carried out in accordance with the following three optimization objectives, namely maximizing total similarity, maximizing the topic coverage of papers, and optimizing fairness of reviewer assignment algorithm.

3.3.2.1. Maximizing total similarity As stated above, the similarity score (also known as matching degree) between each reviewer–paper pair can be calculated in a number of ways, such as bidding, BOW, LM and topic models. Some researchers thought that when the sum of the similarity in the reviewer assignment scheme reach the maximum value, it can be believed that every paper has been assigned the most suitable reviewer, that is, the reviewer assignment scheme is the optimal one. The idea of maximize total similarity has been applied in several representative conference management software, such as EasyChair, HotCRP and TPMS. Generally, there are mainly two constraints to be satisfied, namely the upper limit for each reviewer's workload and the minimum

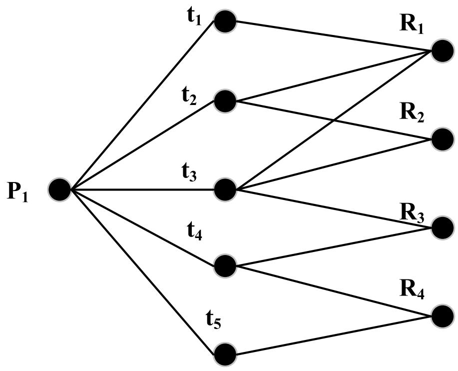

Fig. 5. A reviewer assignment example of maximizing topic coverage of papers.

number of experts required for each paper (Charlin et al., 2011). The RAP can be formalized as follows:

$$\begin{aligned} & max \sum_{i=1}^{|R|} \sum_{j=1}^{|P|} x_{ij} s_{ij} \\ subject \quad to \quad & \sum_{j=1}^{|P|} x_{ij} < U_i, \quad \forall i = 1, 2, \ldots, |R| \\ & \sum_{i=1}^{|R|} x_{ij} = C_j, \quad \forall j = 1, 2, \ldots, |P| \\ & x_{ij} \in \{0, 1\}, \quad \forall i, j \end{aligned} \tag{1}$$

Here, R represents the set of reviewers, P represents the set of papers to be reviewed, and s_{ij} represents the similarity between paper i and reviewer j. $\{U_i\}_{i=1}^{|R|}$ represents the upper bound of the reviewer's workload, $\{C_j\}_{j=1}^{|P|}$ represents the lower bound of the number of reviewers required for each paper. The matching result between reviewer r_i and paper p_j is represented by the variable x_{ij}, where $x_{ij} = 1$ indicates reviewer r_i is assigned to the paper p_j, otherwise $x_{ij} = 0$. Obviously, Formula (1) is a typical ILP problem, whose global optimization objective is to maximize the sum of the reviewer–paper similarity, and it can be solved by the standard linear programming solver such as commercial ILOG CPLEX 11.0 package.[3] (Karimzadehgan et al., 2008; Neshati, Beigy, & Hiemstra, 2014) Although the maximum-similarity matching is guaranteed to be found in this way, the time complexity of $O\left(n^3\right)$ or even higher makes them almost inapplicable for a large number of papers.

3.3.2.2. Maximizing the topic coverage of papers Although the idea of maximizing total similarity is very intuitive, sometimes it may leave part of paper's topics unevaluated by any reviewers, especially when the paper is interdisciplinary and contains multiple distinct-different topics.

As shown in Fig. 5, paper P_1 contains 5 topics ($t_1 - t_5$) and there are 4 reviewers ($R_1 - R_4$) in total. The edge between a paper-topic or reviewer–paper pair indicates that the topic is covered by the paper or reviewer's expertise. Supposing that a paper needs 2 reviewers and taking the number of shared topics between the paper–reviewer pair as their similarity. For instance, paper P1 and reviewer R_1 share 3 topics (i.e., t_1, t_2 and t_3), thus their similarity is 3. According to the approach of maximizing total similarity, the optimal assignment scheme will be $\left\{(P_1, R_1), (P_1, R_2)\right\}$, whose total similarity is 6. However, this scheme has an obvious disadvantage, that is, the two reviewers (R_1 and R_2) has no expertise to evaluate the two topics t_4 and t_5. A more reasonable assignment scheme would be $\left\{(P_1, R_1), (P_1, R_4)\right\}$. Although the total similarity slightly reduced from 6 to 5, the reviewer set (i.e., R_1 and R_4) could cover all the 5 topics of paper P_1, thus higher review quality could be guaranteed. Based on the above observation, some researchers argued that more attention should be paid to the topic coverage of papers, i.e., setting the global objective as maximizing the topic coverage of papers (Karimzadehgan & Zhai, 2012; Kou, Leong Hou, Nikos et al., 2015; Long et al., 2013; Mirzaei et al., 2019; Neshati et al., 2014; Sidiropoulos & Tsakonas, 2015). Long et al. (2013) believed that a paper was well-reviewed when the reviewers had the expertise to cover every single topic of the paper. They transformed the expertise of reviewers and the content of papers into two set of topics, and the quality of reviewer assignment scheme was assessed by the coverage ratio of paper topics. Therefore, RAP was modeled as a Maximum Topic Coverage Paper–Reviewer Assignment (MaxTC-PRA) problem. The MaxTC-PRA problem was proven to be NP-Hard (Long et al., 2013). To address this hardness, the authors designed a greedy algorithm, which could provide a $\frac{1}{3}$ approximation ratio.

Inspired by the work of MaxTC-PRA, Kou, Leong Hou, Mamoulis et al. (2015) further argued that the importance of different topics in a paper should vary. For instance, a paper could be related to many topics, but only one of them was the main subject

[3] https://www.ibm.com/products/ilog-cplex-optimization-studio

Table 4
A toy example paper–reviewer similarity matrix.

ID	Paper a	Paper b	Paper c
Reviewer 1	1	1	1
Reviewer 2	0	0	0.2
Reviewer 3	0.25	0.25	0.5

of the paper. Thus, different topics in a paper should be set with different weights. In this way, the RAP should be modeled as a Weighted-coverage Group-based group Reviewer Assignment Problem (WGRAP), which assessed the reviewer assignment quality by a group-based objective function under weighted coverage of topics. It was proven that the WGRAP was also a NP-Hard problem (Kou, Leong Hou, Mamoulis et al., 2015). Kou et al. designed an approximation algorithm, whose approximation ratio increased from $\frac{1}{3}$ to at least $\frac{1}{2}$, which was a significantly improvement compared to that of MaxTC-PRA.

3.3.2.3. Optimizing fairness of reviewer assignment algorithm Although the reviewer assignment approach that optimizes a global objective (total similarity or topic coverage) can get an accurate assignment scheme in some cases, it cannot guarantee the fairness between papers. For instance, some papers may be assigned to the right reviewers, but some papers may be assigned to reviewers with low similarity (or expertise). As a result, the review result may be inaccurate for the discriminated papers (Garg et al., 2010; Stelmakh et al., 2021). The inherent reason is that the optimal assignment may discriminate some papers in order to maximize the global objective. To see this issue, we consider a toy example with similarities listed in Table 4 (Stelmakh et al., 2021).

For this example, there are 3 reviewers and papers respectively, assuming the assignment should satisfy the following two constraints:

• One reviewer is required for every paper.
• Each reviewer will review at most one paper.

Paper c is easy to evaluate, because it has non-zero similarities with all the reviewers. Paper a and paper b are more specific, since the weak Reviewer 2 has no expertise to evaluate them. Reviewer 1 is an expert and is competent to evaluate all the three papers. According to the idea of maximizing total similarity, the algorithm will assign reviewers 1, 2, and 3 to papers a, b, and c respectively. For this assignment scheme, paper b is assigned a reviewer without sufficient expertise to evaluate it. By contrast, the alternative assignment scheme, which assigns reviewers 1, 2, and 3 to papers a, c, and b respectively, can ensure that every paper has a reviewer with similarity not less than 0.2. Obviously, this scheme is fairer and it does not discriminate against the disadvantaged paper b for improving the review quality of the already benefiting paper c.

Motivated by the above observation, Hartvigsen et al. (1999) necessitated every paper to have at least one reviewer with expertise higher than a certain threshold, and then maximized the threshold value. However, this strategy just partially avoided the discrimination against certain papers, i.e., although having one "strong" reviewer assigned to each paper, the algorithm may still discriminate against certain papers while assigning the remaining reviewers. For instance, some large academic conferences such as NIPS usually assign 4–6 reviewers to each paper, according to the "at least one strong reviewer for each paper" strategy, a careful assessment of the paper by the strong reviewer might be lost in the noise induced by the remaining weak reviewers.

Instead of guaranteeing high expertise for one reviewer, some researchers aimed at the assignment with high total expertise of all reviewers assigned to a paper. Inspired by the max–min fairness idea (Asadpour & Saberi, 2010; Hahne, 1991; Stelmakh et al., 2021) proposed the PR4 A algorithm, which optimized the fairness for the least satisfied paper. PR4 A algorithm could be formalized as a maximization over the minimum paper score with respect to an assignment, which was formalized as follows:

$$\Gamma^{S}(A) = \min_{j \in [m]} \sum_{i \in R_A(j)} s_{ij} \tag{2}$$

Here, A is a feasible assignment, S is the paper–reviewer similarity matrix, and $\Gamma^{S}(A)$ is the fairness optimization objective that maximizes the minimum sum similarity across all the papers. PR4 A iteratively solved maximum-flow through a sequence of specially constructed networks, and guaranteed to return a solution that is within a bounded multiplicative constant of the optimal solution with respect to their max–min objective. However, PR4 A did not enforce load equity among the reviewers, thus the workload of reviewers may be highly skewed.

To achieve both fairness and workload balance among all the reviewers simultaneously, Kobren et al. (2019) proposed a local fairness formulation of the RAP. The formulation was cast as the following integer linear program:

(1) Optimizing the global objective of maximizing total similarities, as shown in Formula (1).

(2) Introducing lower bound constraints that served to balance the reviewing workload among reviewers, as shown in Formula (3), where $\{L_i\}_{i=1}^{|R|}$ was the set of lower bounds on reviewer loads.

$$\sum_{j=1}^{|P|} x_{ij} \geq L_i, \qquad \forall i = 1, 2, \ldots, |R| \tag{3}$$

(3) Adding local fairness constraints, which ensured that each paper was assigned to a set of reviewers that collectively possess sufficient expertise. The local fairness constraints can be formalized as Formula (4), where they constrained the total similarity score at each paper to be no less than T.

$$\sum_{i=1}^{|R|} x_{ij} s_{ij} \geq T, \qquad \forall j = 1, 2, \ldots, |P| \tag{4}$$

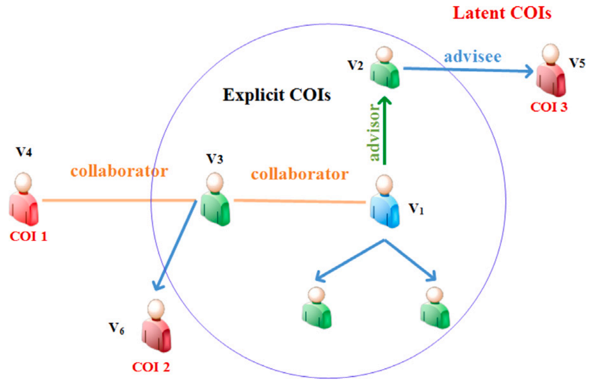

Fig. 6. An academic network including latent and explicit COIs.

The local fairness formulation was also proved to be NP-Hard (Vazirani, 2013). To address this hardness, Kobren et al. proposed two algorithms: FairIr and FairFlow. FairIr was best at optimizing the global matching while satisfying fairness constraints, and FairFlow was faster than PR4 A while achieving similar quality solutions on data from real conferences. Despite the lack of theoretical guarantees, FairFlow still could construct highly fair matchings.

The core of fair reviewer alignment algorithm lies in that the most "discriminated" papers can be assigned a group of reviewers with sufficient expertise. Considering the complexity with different fairness constraints, researches should balance accuracy and efficiency of the reviewer assignment algorithms, thereby improving its overall performance in different paper review scenarios.

3.4. Conflict of interest in RAP

Fairness is an eternal theme in peer review activity, and it is also one of the core requirements of reviewer assignment algorithm. Not only the max–min fairness idea in the reviewer assignment optimization process, but also the COI relationship has been study extensively to deal with the fairness dilemma in peer review process. Biased review results may deteriorate the career trajectory of some researchers, thus exacerbate the "Matthew Effect" of rich getting richer in academia. Fairness in peer review is influenced by several factors, such as the expertise of reviewers, the quality of review comments and the design of the review form, but the most important factor may be the relationships between authors and reviewers, also known as COIs, which has been explored by a number of literature on peer review (Aleman-Meza et al., 2006; Huang et al., 2016; Li et al., 2017; Pradhan et al., 2020; Sun, Barber, Gupta, Aggarwal and Han, 2011; Sun, Han, Yan, Yu and Wu, 2011; Wang et al., 2010; Wu, Leong Hou, Bhowmick, & Gatterbauer, 2017, 2018; Yan et al., 2017).

In various peer review scenarios, COIs can be categorized into two categories, namely Explicit COIs and Latent COIs. Explicit COIs refer to definite relationships that can be easily declared by the authors and reviewers guided by simple pre-defined COI rules, such as collaborator in the past five years, advisor–advisee relationship and colleague relationship. In contrast, Latent COIs refer to some potential relationships that are not required to be declared according to the pre-defined COI rules, thus they should be inferred from the Explicit COIs, such as academic sibling relationship or academic sibling's colleague. Fig. 6 shows a COI network composed of user nodes and three types of edges, including collaborator (orange lines), advisee (blue arrows) and advisor (green arrow). In this example, v_1 is an author node and all the red nodes are reviewers. As can be seen, the academic sibling relationship between v_1 and v_5 belongs to the Latent COI since it is too implicit to be covered by the pre-defined COI rules. The same conclusion can be drawn about the COI relationship between v_1 and v_4 (or v_6).

Benferhat and Lang were the first to raise the issue of COI, they pointed out that reviewers should avoid reviewing the papers with COI (Benferhat & Lang, 2001). Long et al. (2013) divided COI into four categories, namely co-author relationship, colleague relationship, advisor–advisee relationship, and competitive relationship. However, the method of checking COIs manually has two obvious disadvantages. On one hand, for the existing conference management systems such as EasyChair, although the explicit COIs can be collected by a set of declaration rules, some latent COIs may be omitted due to the imperfection of the rules. On the other hand, the COI list may include several hundreds of reviewers in some cases, even though program committee chairs or editors can check the suspicious COIs manually, the approach is too time consuming and error-prone. Therefore, there is an urgent need for efficient tools to facilitate the COI-detecting process.

With respect to latent COI detection, Tang et al. (2010) judged whether there were latent COIs between reviewers and authors based on the co-authoring relationship in the past five years and the institution where they worked. Yan et al. (2017) proposed the MinCOI-PRA algorithm, which divided COIs into two categories, namely the scholar-to-scholar and institution-to-institution relationship, then an academic network was constructed to detect the Latent COIs by calculating path distances between the corresponding nodes in the network. Wu et al. (2018) proposed a COI declaration and detection system named PISTIS, which took three steps to finish the COI-detection work. First, the authors and reviewers should make explicit COI declarations through the graphical interface of PISTIS. Second, PISTIS constructed a heterogeneous academic network of the COIs. Finally, PISTIS leveraged supervised learning algorithm (e.g., logistic regression) to detect the latent COIs automatically, which could enable program chairs to verify the reviewer assignment scheme and manually drop some suspicious assignment pairs.

On the whole, most of the literature on COI mainly focuses on the classification and automatic detection of COIs. Less attention is paid on the quantitative analysis on the relationship strength of different COIs, as well as the quantitative analysis of review avoidance strategy, which are both very important for the design of review avoidance strategy.

4. Performance evaluation methods

Performance evaluation is a key task for researchers when they propose novel algorithms or models. However, evaluating systems for reviewer assignment algorithm is difficult. First, the actual reviewer assignment information of the academic conference or journal is usually not available for the sake of fairness and confidentiality. Furthermore even if such data were available, it is not clear that the assignment results necessarily represent the best matching, or just a feasible compromise to certain difficult optimization. To tackle this dilemma, some researchers designed different datasets and metrics to evaluate the performance of their algorithms.

4.1. Datasets

Due to the paucity of real-world reviewer assignment data, researchers can only construct the dataset by themselves to evaluate the performance of reviewer assignment algorithms. According to whether the dataset contains annotation information, these datasets can be divided into two categories, namely dataset with ground truth data and dataset without ground truth data.

The NIPS dataset was collected by Mimno and McCallum (2007), it is the most popular benchmark dataset with ground truth and has been utilized by several other researchers (Anjum et al., 2019; Liu et al., 2014). The NIPS dataset contains 148 papers accepted in NIPS 2006 and abstracts from the publications of 364 reviewers, which were from the reviewer list of NIPS 2005. Several prominent researchers from the NIPS community were invited to provide a ground truth relevance judgment of a query paper and proposed reviewer pair. The ground truth consists of annotations from 9 annotators on the relevance of 650 reviewer–paper pairs. Each pair is rated by a four-level relevance scheme, where “0” means irrelevant, “1” means slightly relevant, “2” means relevant and “3” means very relevant. In addition, Karimzadehgan et al. (2008) created the SIGIR dataset as the golden standard for evaluating their approaches, and Tang et al. (2010) treated the real course-teacher assignment as the ground-truth. A few lucky researchers can get the actual reviewer assignment data, which was treated as the ground truth data. For instance, Kotak, Roy, Dasgupta, and Ghosal (2021) used an actual dataset to analyze the consistency of different NLP approaches for reviewer assignment algorithm. Specially, the dataset that is provided by The Technical Program Committee Chair of an academic conference comprises the following four parts: (1) all submitted papers, (2) the full list of reviewers with their affiliations, (3) track-wise list of reviewers, and (4) papers allocated to a set of reviewers (which is regarded as the ground-truth data). However, due to the particularity of RAP, dataset with ground truth is scarce and limited-sized, and thus the acceptance level is still not high enough.

Considering the scarcity and credibility of dataset with ground-truth data, many researchers tend to build or adopt datasets without ground truth to validate the effectiveness of their algorithms (Jecmen et al., 2020; Lian et al., 2018). A typical evaluation dataset generally consists of two parts, i.e., the papers to be reviewed and the reviewer list. A popular dataset collection method generally includes the following two steps. First, for one or several top-level academic conferences (e.g., CVPR, KDD and SIGIR), the publications and all the program committees (PCs) were selected as the papers and reviewers respectively. Then, for each reviewer, all his (or her) publications were collected from the academic database such as ACM Digital Library, DBLP (Digital Bibliography & Library Project) and Google scholar. This method can both ensure the quality of paper and expertise of reviewers, hence it was used by more of the researchers. However, due to the huge amount of optional academic data, the datasets constructed in different literature usually vary widely, which is not conducive to the generation of benchmark endorsed by the community.

To the best of our knowledge, currently there is no widely accepted dataset for reviewer assignment algorithm performance evaluation, and it will be a promising research direction that requires more efforts.

4.2. Metrics

As mentioned in Section 3, the whole reviewer assignment process includes three core stages based on different algorithms, i.e., construction of candidate reviewer database, computation of matching degree between papers and reviewers, and the reviewer assignment optimization algorithm. Theoretically, there are several different methods to conduct performance evaluation. One can either measure the overall performance of reviewer assignment algorithm, or just evaluate the certain stage they optimized, which can highlight the effectiveness of their work (Patil & Mahalle, 2020).

As to the stage 1 - candidate reviewer database construction, it can be easily measured by the specifications such as the number of reviewers contained in the database. Therefore, this stage is rarely evaluated individually in RAP literature. Most of the evaluation metrics are designed for the subsequent two stages, i.e., calculating matching degree between reviewers and papers, and assigning reviewers to papers.

In the numerous literatures of RAP, quite a few information retrieval techniques have been used to improve the accuracy of matching degree between reviewers and papers, and there are a number of evaluation metrics to measure the performance of various matching degree computation methods, which correspond to the stage 2 of reviewer assignment algorithm. According to whether the dataset with ground truth is available or not, the design methods of metrics can be divided into the following two categories. On

Table 5
Summary of representative metrics for reviewer assignment optimization algorithms.

Related works	Key ideas	Metrics	Comparison methods
Karimzadehgan et al. (2008)	The reviewers assigned should cover all the topical aspects of a paper	Coverage, confidence, average confidence	Language modeling with dirichlet smoothing
Tang et al. (2010)	The expertise matching problem should be handled considering domain-specific constraints such as topic coverage and load balance	Accumulative matching score, load variance and expertise variance	A greedy algorithm that assigns experts with highest matching score to each query
Long et al. (2013)	Maximize the coverage of the topics of the papers covered by the assigned reviewers	Topic coverage	An ILP method
Kou, Leong Hou, Mamoulis et al. (2015)	Different topics in a paper should be of different importance	Optimality ratio, superiority ratio and running time	Stable matching, ILP method, greedy algorithm
Kobren et al. (2019)	Guarantee fairness by the constraining the paper score at each paper to be no less than T	Running time, Paper score, and the load balance degree	PR4A, TPMS RAP
Jin et al. (2020)	Topical relevance, topical authority and research interest should be considered together and formulated as ILP to handle RAP	Distance@k for relevance, Interest@k for interest trend and Authority@k for authority	VSM (Vector Space Model), LM and ATM
Stelmakh et al. (2021)	Maximize the review quality of the most disadvantaged paper to achieve fairness and statistical accuracy	Approximation guarantee	An approximation algorithm proposed by Garg et al.

one hand, when the ground truth dataset was available, the metrics such as precision, recall and F-score could be used to measure the performance. Mimno and McCallum (2007) used the "Precision at relevance cutoff 2 (or 3) after retrieving n reviewers" to verify that their APT model could achieve better performance than other baseline algorithms (ATM, single-doc, max-doc model, etc.). Likewise, Anjum et al. (2019) adopted the metric P@k (Precision at k), which was defined to be the percentage of relevant reviewers in the top-k recommendations made by the model to a submission, and proved that their Common Topic Model had better accuracy than some other information retrieval techniques such as APT200, LDA and Hidden Topic Model. In addition, Zhang et al. (2020) transformed RAP into a multi-label classification issue through Hiepar algorithm, and they used Recall@k (Recall at top k) and NDCG@k (Normalized Discounted Cumulative Gain) metrics to prove its performance advantages over other algorithms (e.g., LDA, CNN, LSTM, Bi-LSTM, Word2vec, and BERT). One the other hand, some researchers argued that there was no gold standard data for RAP, which means that some evaluation metrics for information retrieval such as precision and recall were not appropriate. In this case, researchers proposed different evaluation metrics according to their own problem definition. For instance, when testing with the unlabeled CIKM dataset, Liu et al. (2014) designed several metrics, such as Expertise@k and Authority@k, to verify the performance advantages of their unification model. Jin et al. (2020) invented a group of persuadable evaluation metrics (e.g., Distance@k, Interest@k and Authority@k) about the relevance, the interest trend and the authority to conduct performance evaluation. Some other similar evaluation metrics can be also found in Tang et al. (2012, 2010).

Given the matching degree matrix, some researchers treated RAP as a matching problem between the reviewer set and paper set, thus some evaluation metrics were designed to validate the optimization objects of different reviewer assignment optimization algorithms. In this case, the evaluation metrics can mainly be categorized according to the following two considerations: goodness and fairness. For the goodness aspect, some researchers used topic coverage as the key evaluation metric (Kou, Leong Hou, Mamoulis et al., 2015; Long et al., 2013), which was consistent with their optimization objects. For the fairness aspect, researchers proposed some metrics to measure the fairness degree of their optimization algorithms. Kobren et al. (2019) used paper score as the metric to measure the fairness of every algorithm. In addition, the authors also used the running time to evaluate the speed of algorithms, which may be useful in the online paper–reviewer matching scenario (Tang et al., 2010). A detailed summary of representative metrics for reviewer assignment optimization algorithms is given in Table 5. Taking the metric "Topic coverage" as an example, it refers to the distinct topics covered by the assigned reviewers, which can be used to verify the advantages of MaxTC-PRA over the algorithms of maximizing total similarity (Long et al., 2013).

5. Future opportunities

With decades of continuous research, the RAP is always a hot research field where much progress has been made. However, there are still some problems to be solved.

• More state-of-the-art deep learning models should be applied to compute the matching degree between reviewers and papers. In recent years, deep learning technology has made breakthroughs in several fields such as natural language processing. For example, the BERT model (Devlin, Chang, Lee, & Toutanova, 2019), a pre-trained model proposed by Google, has set new standards in 11 NLP tasks (e.g., sentence pair classification, question answering, and single sentence tagging). State-of-the-art deep learning models have powerful feature extraction capabilities, which will be very effective to optimize the accuracy of matching degree between papers and reviewers, thus can further improve the overall assignment result of reviewer assignment algorithms.

• The reviewer avoidance strategy needs to be further refined based on the quantitative analysis on the relationship strength of different COIs. Reviewers are always in a complex social network, and there are different types of social relationships between every potential reviewer and reviewee (e.g., the authors of paper) pair, and the relationship strength of different social relationships may vary greatly. Existing literature tends to believe that the reviewer must avoid reviewing a paper if there is COI between the reviewer and reviewee, without considering its type and relationship strength. This one-size-fits-all way has several drawbacks, one of which is that it may lead to the lack of experts qualified for peer review. How to quantify the relationship strength of different COIs and design flexible avoidance strategies on this basis are still open problems, which needs more efforts to be made in the future.

• More widely-accepted annotated datasets need to be jointly designed by the researchers and peer review organizers. The world has entered the data-driven era, and data is playing an increasing important role in the decision-making process of many fields. However, the paucity of widely-accepted annotated datasets is limiting the conduction of performance comparison between different reviewer assignment algorithms, and thus hinders the further development of the research. Generally speaking, peer review organizes such as conference Program Chairs own plenty of reviewer assignment data, and the researchers in this field knows best how to design the best datasets with ground truth. Therefore, for researchers and peer review organizers, they need to cooperate to design and release more large-sized evaluation datasets, which are just like the widely-used ImageNet dataset and Netflix Prize data mining case in the computer vision field.

6. Further discussions and implications

Before this survey paper, Wang, Shi et al. (2010) conducted the most comprehensive survey of RAP. After the publication of this landmark survey, the research on RAP has made rapid progress and some new algorithms have emerged in every stage, which inspires us to write this paper to reflect the research progress systematically, and highlight the latest research achievements that the landmark survey did not cover. Through our analysis and summary, the following findings are obtained.

In the candidate reviewer database construction stage, with the development of NLP and Big Data technology, some previously forbidden areas of academic research have been explored, several large expert databases with accurate and comprehensive data have been put into practical use in some disciplines. Among the three reviewer database construction methods, the traditional online-recruiting and experience-based reviewer information collection methods have obvious shortcomings. Firstly, without taking advantage of the massive information in the Internet, the amount of reviewer information collected by these two methods are limited. Secondly, since the above two methods cannot conduct active data crawling, the reviewer information collected by them usually cannot be updated in time. The method of dynamic reviewer database construction based on massive Internet data can overcome the shortcomings mentioned above. In the future, with the growing frequency of peer review, the demand for more comprehensive reviewer database will increase accordingly, thus the dynamic reviewer database construction method will play a more important role. Meanwhile, considering the great efforts it takes to construct an accurate and comprehensive reviewer database, it may be very meaningful to promote the sharing and integration of these reviewer databases.

For the second stage, NLP is the biggest driving force for matching degree computation. Although the subjective scoring method is still the default operation in many peer review scenarios, it is easy to introduce reviewers' subjective deviation. Theoretically, the deep learning technology can unearth the deep semantic information of reviewers and papers, therefore, compared with the methods based on word information and topic models, it can generate more accurate matching degree results. However, the deep learning technology such as neural network models is usually limited by its "black box" property, which usually results in poor interpretability. Due to the lack of widely accepted ground truth data reviewer assignment, the advantages of deep learning model have not yet become the consensus of researchers in RAP field. Meanwhile, some researchers proposed a series of matching degree calculation methods based on multiple information, considering the factors outside the text content to calculate matching degree, such as authority, interest and recency of reviewer's publications, and achieved "desirable" assignment results on some private datasets. The idea is very natural and in line with the psychological expectations of many researchers. It can be predicted that more hybrid methods will be proposed and deployed in the real-world reviewer assignment systems in the future. Notably, how to set appropriate "weight" for every single factor is still an open question.

With respect to the third stage, there are mainly three considerations in total. Although the idea of maximizing the total similarity of reviewer assignment scheme is relatively natural, it may lead to an unfair assignment of the review resource (e.g., reviewer's expertise), or the topic coverage of the reviewed papers is highly skewed, thus it is difficult to generate a convincing assignment scheme in some cases. The idea of maximizing topic coverage could improve the review results in several cases. However, this approach also ignores the fairness of reviewer resource assignment. In addition, it can only work with certain topic extraction algorithms of the matching degree computation stage, resulting in poor synergy with some new technologies such as Neural Network models. The third consideration, that is, optimizing the fairness of reviewer assignment algorithm, both takes into account the accuracy and fairness of reviewer assignment algorithm, and it can theoretically generate better assignment scheme than the former two ideas, especially the algorithms proposed by literature (Stelmakh et al., 2021) and Kobren et al. (2019), they adopt different constraints to guarantee the review quality of the most disadvantaged papers. However, due to the complexity of constraints, it is difficult to get the optimal solution in some cases. Obviously, the assignment results of algorithms in stage 3 depends heavily on the accuracy of matching degree calculated in stage 2. In order to achieve better evaluation results, researchers may need to make more attempts in the combination of different matching degree computation and reviewer assignment optimization algorithms.

This survey paper has at least two implications for researchers in RAP field. Firstly, through our analysis, it is obvious that the research hotspots of RAP lie in the last two stages: matching degree computation and reviewer assignment optimization, while less attention has been paid to the first stage, i.e., construction of candidate reviewer database. As a matter of fact, the first stage is the basis of the whole reviewer assignment process, which directly determines the results of matching degree of stage 2 and finally affects the final reviewer assignment scheme. Therefore, for the RAP researchers, maybe they should adjust their research priorities and try to achieve more progress in this "blue ocean" stage. Secondly, most of RAP researchers just pay attention to one certain stage of reviewer assignment algorithm, but in fact the three stages of solving RAP are an interlocking and tightly coupled system, the output of a previous stage may strongly influence the results of the next stage. Therefore, instead of trying to obtain optimal result just in one certain stage, RAP researchers should perhaps start from the overall perspective and reconsider the design ideas of reviewer assignment algorithm systematically.

This survey can also provide some reference value for the peer review organizers. Due to the historical reasons, many peer review systems still adopt some "traditional" techniques, such as the online-recruiting method, the bidding procedure for matching degree computation, and the assignment optimization algorithms without considering the risk of unfairness. This survey provides a detailed analysis of the pros and cons for the algorithms used in every stage, which may inspire the peer review organizers to make some novel attempts on their current reviewer assignment systems. In addition, organizer of different peer review scenarios own a large number of actual reviewer assignment results, perhaps they can do some work on how to generate some ground truth datasets for RAP, which will be helpful to unify the evaluation method and greatly promote the development of RAP research.

7. Conclusion

Peer review is the gateway process and quality control mechanism for papers submitted to journals, conferences and granting agencies across a wide range of disciplines, its accuracy and fairness can significantly influence the career trajectory of many researchers. RAP is the core problem in peer review process, assigning competent reviewers to papers is the first and potentially most important stage in the peer review process. However, as the number of papers to be reviewed grows rapidly, it seems infeasible for the peer review organizer to manually assign the reviewers on their own. Therefore, how to automatically assign papers to well-suitable reviewers becomes the key problem of peer review.

To address this problem, plenty of researchers from several different disciplines proposed many reviewer assignment algorithms in the past few decades. In this paper, we provide a comprehensive survey on reviewer assignment algorithm, and find that natural language processing technology plays a key role in the development of the RAP field. Comparing with existing survey papers, this survey provides a more clear understanding of all the stages of reviewer assignment algorithm, and especially presents the latest research progress in the past decade. We hope this survey can serve as a valuable reference not only for the researchers who would like to make contributions in RAP field, but also for peer review organizers to adopt the most appropriate reviewer assignment algorithms.

CRediT authorship contribution statement

Xiquan Zhao: Conceptualization, Methodology, Investigation, Writing – original draft, Writing – review & editing. **Yangsen Zhang:** Supervision, Project administration, Funding acquisition, Resources.

References

Abduljaleel, A. Q., Naser, M. A. N., & Al-mamory, S. O. (2021). Reviewer assignment using weighted matching and hungarian algorithm. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, *12*(7), 619–627.

Aleman-Meza, B., Nagarajan, M., Ramakrishnan, C., Ding, L., Kolari, P., Sheth, A. P., et al. (2006). Semantic analytics on social networks: Experiences in addressing the problem of conflict of interest detection. In *15th international conference on world wide web, Edinburgh, Scotland* (pp. 407–416).

Anjum, O., Gong, H., Bhat, S., Hwu, W.-M., & Xiong, J. (2019). PaRe: A paper-reviewer matching approach using a common topic space. In *2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing, Hong Kong, China* (pp. 518–528).

Asadpour, A., & Saberi, A. (2010). An approximation algorithm for max-min fair allocation of indivisible goods. *SIAM Journal on Computing*, *39*(7), 2970–2989. http://dx.doi.org/10.1137/080723491.

Balog, K., Azzopardi, L., & de Rijke, M. (2006). Formal models for expert finding in enterprise corpora. In *SIGIR '06*, *Proceedings of the 29th annual international acm sigir conference on research and development in information retrieval* (pp. 43–50). New York, NY, USA: Association for Computing Machinery, http://dx.doi.org/10.1145/1148170.1148181.

Basu, S., Bilenko, M., & Mooney, R. J. (2004). A probabilistic framework for semi-supervised clustering. In *10th ACM SIGKDD international conference on knowledge discovery and data mining, Seattle, WA, USA* (pp. 59–68).

Basu, C., Hirsh, H., Cohen, W. W., & Nevill-Manning, C. (1999). Recommending papers by mining the web. In *IJCAI99 workshop on learning about users, Stockholm, Sweden*.

Bence, V., & Oppenheim, C. (2004). The influence of peer review on the research assessment exercise. *Journal of Information Science*, *30*(4), 347–368.

Benferhat, S., & Lang, J. M. (2001). Conference paper assignment. *International Journal of Intelligent Systems*, *16*(10), 1183–1192. http://dx.doi.org/10.1002/int.1055.

Bergman, M. K. (2001). White paper: the deep web: surfacing hidden value. *Journal of Electronic Publishing*, *7*(1).

Beverly, R., & Allman, M. (2012). Findings and implications from data mining the IMC review process. *ACM SIGCOMM Computer Communication Review*, *43*(1), 22–29. http://dx.doi.org/10.1145/2427036.2427040.

Biswas, H. K., & Hasan, M. M. (2007). Using publications and domain knowledge to build research profiles: An application in automatic reviewer assignment. In *International conference on information and communication technology* (pp. 82–86). Dhaka, Bangladesh: IEEE.

Blei, D. M., Ng, A., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, *3*, 993–1022. http://dx.doi.org/10.5555/944919.944937.

Bornmann, L., & Mutz, R. U. D. (2015). Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. *Journal of the Association for Information Science and Technology*, *66*(11), 2215–2222.

Cabanac, G., & Preuss, T. (2013). Capitalizing on order effects in the bids of peer-reviewed conferences to secure reviews by expert referees. *Journal of the Association for Information Science and Technology*, *64*(2), 405–415. http://dx.doi.org/10.1002/ASI.22747.

Camara, M., Ortega, J., & de Toro, F. (2009). A single front genetic algorithm for parallel multi-objective optimization in dynamic environments. *Neurocomputing*, *72*(16–18), 3570–3579. http://dx.doi.org/10.1016/j.neucom.2008.12.041.

Campbell, C. S., Maglio, P. P., Cozzi, A., & Dom, B. (2003). Expertise identification using email communications. In *12th international conference on information and knowledge management, New Orleans LA USA* (pp. 528–531).

Charlin, L., & Zemel, R. (2013). The toronto paper matching system: An automated paper-reviewer assignment system. In *ICML workshop on peer reviewing and publishing models, Atlanta, Georgia, USA*.

Charlin, L., Zemel, R. S., & Boutilier, C. (2011). A framework for optimizing paper matching. In *27th conference on uncertainty in artificial intelligence, Barcelona, Spain* (pp. 86–95).

Conry, D., Koren, Y., & Ramakrishnan, N. (2009). Recommender systems for the conference paper assignment problem. In *3rd ACM conference on recommender systems (RecSys), New York, New York, USA* (pp. 357–360).

Davenport, T. H., & Prusak, L. (1998). *Working knowledge: how organizations manage what they know*. Harvard Business Press.

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, *41*(6), 391–407.

Deng, H., King, I., & Lyu, M. R. (2008). Formal models for expert finding on dblp bibliography data. In *8th IEEE international conference on data mining* (pp. 163–172). Pisa, Italy: IEEE.

Desai, K., Devulapalli, V., Agrawal, S., Kathiria, P., & Patel, A. (2017). Web crawler: Review of different types of web crawler, its issues, applications and research opportunities. *International Journal of Advanced Research in Computer Science*, *8*(3), 1199–1202. http://dx.doi.org/10.26483/IJARCS.V8I3.3190.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *The proceedings of the 2019 conference of the north American chapter of the association for computational linguistics, Minneapolis, Minnesota* (pp. 4171–4186).

Di Mauro, N., Basile, T. M. A., & Ferilli, S. (2005). GRAPE: An expert review assignment component for scientific conference management systems. In *18th international conference on innovations in applied artificial intelligence, Bari, Italy* (pp. 789–798).

Duan, Z., Tan, S., Zhao, S., Wang, Q., Chen, J., & Zhang, Y. (2019). Reviewer assignment based on sentence pair modeling. *Neurocomputing*, *366*, 97–108. http://dx.doi.org/10.1016/j.neucom.2019.06.074.

Dumais, S. T., & Nielsen, J. (1992). Automating the assignment of submitted manuscripts to reviewers. In *15th international ACM SIGIR conference* (pp. 233–244).

Fang, H., & Zhai, C. (2007). Probabilistic models for expert finding. In *European conference on information retrieval* (pp. 418–430). Rome, Italy: Springer.

Ferilli, S., Mauro, N. D., Basile, T. M. A., Esposito, F., & Biba, M. (2006). Automatic topics identification for reviewer assignment. In *International conference on industrial and engineering applications of artificial intelligence and expert systems, Vol. 4031* (pp. 721–730).

Fiez, T., Shah, N., & Ratliff, L. (2020). A SUPER* algorithm to optimize paper bidding in peer review. In *Conference on uncertainty in artificial intelligence* (pp. 580–589). Toronto, Ontario, Canada: PMLR.

Flach, P. A., Spiegler, S., Gol E Nia, B., Price, S., Guiver, J., Herbrich, R., et al. (2009). Novel tools to streamline the conference review process: Experiences from SIGKDD'09. *ACM SIGKDD Explorations Newsletter*, *11*(2), 63–67. http://dx.doi.org/10.1145/1809400.1809413.

Garg, N., Kavitha, T., Kumar, A., Mehlhorn, K., & Mestre, J. (2010). Assigning papers to referees. *Algorithmica*, *58*(1), 119–136. http://dx.doi.org/10.1007/S00453-009-9386-0.

Goldsmith, J., & Sloan, R. H. (2007). The AI conference paper assignment problem. In *AAAI workshop on preference handling for artificial intelligence* (pp. 53–57).

Hahne, E. L. (1991). Round-robin scheduling for max-min fairness in data networks. *IEEE Journal on Selected Areas in Communications*, *9*(7), 1024–1039. http://dx.doi.org/10.1109/49.103550.

Hartvigsen, D., Wei, J. C., & Czuchlewski, R. (1999). The conference paper-reviewer assignment problem. *Decision Sciences*, *30*(3), 865–876. http://dx.doi.org/10.1111/j.1540-5915.1999.tb00910.x.

Hettich, S., & Pazzani, M. (2006). Mining for proposal reviewers: lessons learned at the national science foundation. In *12th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 862–871). New York: ACM.

Hoang, D. T., Nguyen, N. T., Collins, B., & Hwang, D. (2021). Decision support system for solving reviewer assignment problem. *Cybernetics and Systems*, *52*(5), 379–397. http://dx.doi.org/10.1080/01969722.2020.1871227.

Hoang, D. T., Nguyen, N. T., & Hwang, D. (2019). Decision support system for assignment of conference papers to reviewers. In *International conference on computational collective intelligence* (pp. 441–450). Hendaye, France: Springer.

Hofmann, T. (1999). Probabilistic latent semantic analysis. In *15th conference on uncertainty in artificial intelligence, San Francisco, CA, USA* (pp. 289–296).

Huang, Z., Zheng, Y., Cheng, R., Sun, Y., Mamoulis, N., & Li, X. (2016). Meta structure: Computing relevance in large heterogeneous information networks. In *20th ACM SIGKDD conference on knowledge discovery and data mining, San Francisco, California, USA* (pp. 1595–1604).

Jecmen, S., Zhang, H., Liu, R., Shah, N. B., Conitzer, V., & Fang, F. (2020). Mitigating manipulation in peer review via randomized reviewer assignments. In *NIPS'20, Proceedings of the 34th international conference on neural information processing systems*. Curran Associates Inc..

Jin, J., Geng, Q., Mou, H., & Chen, C. (2019). Author–subject–topic model for reviewer recommendation. *Journal of Information Science*, *45*(4), 554–570.

Jin, J., Geng, Q., Zhao, Q., & Zhang, L. (2017). Integrating the trend of research interest for reviewer assignment. In *26th international conference on world wide web companion, Perth, Australia* (pp. 1233–1241).

Jin, J., Niu, B., Ji, P., & Geng, Q. (2020). An integer linear programming model of reviewer assignment with research interest considerations. *Annals of Operations Research*, *291*(1), 409–433. http://dx.doi.org/10.1007/s10479-018-2919-7.

Kalchbrenner, N., Grefenstette, E., & Blunsom, P. (2014). A convolutional neural network for modelling sentences. In *52nd annual meeting of the association for computational linguistics, Baltimore, Maryland* (pp. 655–665).

Kalman, D. (1996). A singularly valuable decomposition: The SVD of a matrix. *The College Mathematics Journal*, *27*(1), 2–23. http://dx.doi.org/10.1080/07468342.1996.11973744.

Kalmukov, Y. (2020a). An algorithm for automatic assignment of reviewers to papers. *Scientometrics*, *124*(3), 1811–1850. http://dx.doi.org/10.1007/s11192-020-03519-0.

Kalmukov, Y. (2020b). Automatic assignment of reviewers to papers based on vector space text analysis model. In *Proceedings of the 21st international conference on computer systems and technologies' 20* (pp. 229–235).

Kalmukov, Y., & Rachev, B. (2010). Comparative analysis of existing methods and algorithms for automatic assignment of reviewers to papers. *Information Technologies and Control*, *2*, 20–31.

Karimzadehgan, M., & Zhai, C. (2009). Constrained multi-aspect expertise matching for committee review assignment. In *18th ACM conference on information and knowledge management, Hong Kong, China* (pp. 1697–1700).

Karimzadehgan, M., & Zhai, C. (2012). Integer linear programming for constrained multi-aspect committee review assignment. *Information Processing & Management*, *48*(4), 725–740.

Karimzadehgan, M., Zhai, C., & Belford, G. (2008). Multi-aspect expertise matching for review assignment. In *17th ACM conference on information and knowledge management (CIKM), Napa Valley, California, USA* (pp. 1113–1122).

Kautz, H., Selman, B., & Shah, M. (1997). Referral web: combining social networks and collaborative filtering. *Communications of the ACM*, *40*(3), 63–65. http://dx.doi.org/10.1145/245108.245123.

Kawamae, N. (2010). Author interest topic model. In *Proceedings of the 33rd international ACM SIGIR conference on research and development in information retrieval* (pp. 887–888).

Kim, J., & Lee, E. (2018). Understanding review expertise of developers: A reviewer recommendation approach based on latent dirichlet allocation. *Symmetry*, *10*(4), 114.

Kiros, R., Zhu, Y., Salakhutdinov, R., Zemel, R. S., Torralba, A., Urtasun, R., et al. (2015). Skip-thought vectors. In *28th international conference on neural information processing systems, Montreal, Canada* (pp. 3294–3302).

Kobren, A., Saha, B., & McCallum, A. (2019). Paper matching with local fairness constraints. In *25th ACM SIGKDD international conference on knowledge discovery and data mining, Anchorage, AK, USA* (pp. 1247–1257).

Kolasa, T., & Krol, D. (2011). A survey of algorithms for paper-reviewer assignment problem. *IETE Technical Review*, *28*(2), 123. http://dx.doi.org/10.4103/0256-4602.78092.

Kotak, N., Roy, A. K., Dasgupta, S., & Ghosal, T. (2021). A consistency analysis of different NLP approaches for reviewer-manuscript matchmaking. In *International conference on Asian digital libraries* (pp. 277–287). Springer.

Kou, N. M., Leong Hou, U., Mamoulis, N., & Gong, Z. (2015). Weighted coverage based reviewer assignment. In *ACM SIGMOD international conference on management of data* (pp. 2031–2046). Melbourne, Victoria, Australia: ACM.

Kou, N. M., Leong Hou, U., Mamoulis, N., Li, Y., Li, Y., & Gong, Z. (2015). A topic-based reviewer assignment system. In *41st international conference on VLDB, Kohala Coast, Hawai'i, USA, Vol. 8* (pp. 1852–1855).

Le, Q., & Mikolov, T. (2014). Distributed representations of sentences and documents. In *31st international conference on machine learning, Beijing, China* (pp. 1188–1196).

Li, K., Cao, Z., & Qu, D. (2017). Fair reviewer assignment considering academic social network. In *Asia-Pacific web (APWeb) and web-age information management (WAIM) Joint Conference on Web and Big Data* (pp. 362–376). Springer.

Li, B., & Hou, Y. T. (2016). The new automated IEEE INFOCOM review assignment system. *IEEE Network*, *30*(5), 18–24. http://dx.doi.org/10.1109/MNET.2016.7579022.

Lian, J. W., Mattei, N., Noble, R., & Walsh, T. (2018). The conference paper assignment problem: Using order weighted averages to assign indivisible goods. In *32nd AAAI conference on artificial intelligence, New Orleans, Louisiana, USA*.

Liu, X., Suel, T., & Memon, N. (2014). A robust model for paper reviewer assignment. In *8th ACM conference on recommender systems* (pp. 25–32). Foster City, Silicon Valley California, USA: ACM.

Liu, X., Wang, X., & Zhu, D. (2022). Reviewer recommendation method for scientific research proposals: a case for NSFC. *Scientometrics*, 1–24.

Long, C., Wong, R. C., Peng, Y., & Ye, L. (2013). On good and fair paper-reviewer assignment. In *IEEE 13th international conference on data mining, Dallas, TX, USA* (pp. 1145–1150).

Medakene, A. N., Bouanane, K., & Eddoud, M. A. (2019). A new approach for computing the matching degree in the paper-to-reviewer assignment problem. In *2019 international conference on theoretical and applicative aspects of computer science (ICTAACS), Skikda, Algeria, Vol. 1* (pp. 1–8).

Meir, R., Lang, J., Lesca, J., Mattei, N., & Kaminsky, N. (2021). A market-inspired bidding scheme for peer review paper assignment. In *35th AAAI conference on artificial intelligence*.

Merelo-Guervos, J. J., & Castillo-Valdivieso, P. (2004). Conference paper assignment using a combined greedy/evolutionary algorithm. In *International conference on parallel problem solving from nature* (pp. 602–611). Birmingham, UK: Springer.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *26th international conference on neural information processing systems, Lake Tahoe, Nevada, USA, Vol. 2* (pp. 3111–3119).

Mimno, D., & McCallum, A. (2007). Expertise modeling for matching papers with reviewers. In *13th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 500–509). San Jose, California, USA: ACM.

Mirzaei, M., Sander, J., & Stroulia, E. (2019). Multi-aspect review-team assignment using latent research areas. *Information Processing & Management*, *56*(3), 858–878.

Mou, H., Geng, Q., Jin, J., & Chen, C. (2015). An author subject topic model for expert recommendation. In *AIRS* (pp. 83–95). Springer.

Neshati, M., Beigy, H., & Hiemstra, D. (2014). Expert group formation using facility location analysis. *Information Processing & Management*, *50*(2), 361–383.

Nguyen, J., Sánchez-Hernández, G., Agell, N., Rovira, X., & Angulo, C. (2018). A decision support tool using order weighted averaging for conference review assignment. *Pattern Recognition Letters*, *105*, 114–120. http://dx.doi.org/10.1016/j.patrec.2017.09.020.

Ogunleye, O., Ifebanjo, T., Abiodun, T., & Adebiyi, A. A. (2017). Proposed framework for a paper-reviewer assignment system using word2vec. In *Covenant university conference on e-governance in Nigeria, Ota, Ogun State, Nigeria*.

Ouni, A., Kula, R. G., & Inoue, K. (2016). Search-based peer reviewers recommendation in modern code review. In *2016 IEEE international conference on software maintenance and evolution* (pp. 367–377). Raleigh, NC, USA: IEEE.

Patil, A. H., & Mahalle, P. N. (2020). Trends and challenges in measuring performance of reviewer paper assignment. *Procedia Computer Science*, *171*, 709–718. http://dx.doi.org/10.1016/j.procs.2020.04.077.

Peng, H., Hu, H., Wang, K., & Wang, X. (2017). Time-aware and topic-based reviewer assignment. In *International conference on database systems for advanced applications, Suzhou, China* (pp. 145–157).

Pradhan, D. K., Chakraborty, J., Choudhary, P., & Nandi, S. (2020). An automated conflict of interest based greedy approach for conference paper assignment system. *Journal of Informetrics*, *14*(2), Article 101022.

Pradhan, T., Sahoo, S., Singh, U., & Pal, S. (2021). A proactive decision support system for reviewer recommendation in academia. *Expert Systems with Applications*, *169*, Article 114331. http://dx.doi.org/10.1016/j.eswa.2020.114331.

Price, S., & Flach, P. A. (2017). Computational support for academic peer review. *Communications of the ACM*, *60*(3), 70–79. http://dx.doi.org/10.1145/2979672.

Protasiewicz, J. L. A. (2014). A support system for selection of reviewers. In *2014 IEEE international conference on systems, man, and cybernetics* (pp. 3062–3065). San Diego, CA, USA: IEEE.

Richards, D., Taylor, M., & Busch, P. (2008). Expertise recommendation: A two-way knowledge communication channel. In *4th international conference on autonomic and autonomous systems* (pp. 35–40). Gosier, France: IEEE.

Rodriguez, M. A., & Bollen, J. (2008). An algorithm to determine peer-reviewers. In *17th ACM conference on information and knowledge management, Napa Valley, California, USA* (pp. 319–328).

Rodriguez, M. A., Bollen, J., & Van de Sompel, H. (2007). Mapping the bid behavior of conference referees. *Journal of Informetrics*, *1*(1), 68–82. http://dx.doi.org/10.1016/j.joi.2006.09.006.

Rogers, A., & Augenstein, I. (2020). What can we do to improve peer review in NLP? In *Findings of the association for computational linguistics: EMNLP 2020* (pp. 1256–1262).

Rosen-Zvi, M., Griffiths, T., Steyvers, M., & Smyth, P. (2004). The author-topic model for authors and documents. In *20th conference on uncertainty in artificial intelligence (UAI), Banff, Canada* (pp. 487–494).

Shah, N. B. (2021). WSDM 2021 tutorial on systematic challenges and computational solutions on bias and unfairness in peer review. In *The 14th ACM international conference on web search and data mining, New York, NY, USA* (pp. 1131–1133).

Shah, N. B., Tabibian, B., Muandet, K., Guyon, I., & Von Luxburg, U. (2018). Design and analysis of the NIPS 2016 review process. *Journal of Machine Learning Research*.

Shkapenyuk, V., & Suel, T. (2002). Design and implementation of a high-performance distributed web crawler. In *18th international conference on data engineering* (pp. 357–368). IEEE.

Sidiropoulos, N. D., & Tsakonas, E. (2015). Signal processing and optimization tools for conference review and session assignment. *IEEE Signal Processing Magazine*, *32*(3), 141–155. http://dx.doi.org/10.1109/MSP.2014.2359230.

Socher, R., Karpathy, A., Le, Q. V., Manning, C. D., & Ng, A. Y. (2014). Grounded compositional semantics for finding and describing images with sentences. *Transactions of the Association for Computational Linguistics*, *2*, 207–218. http://dx.doi.org/10.1162/tacl_a_00177.

Stelmakh, I., Shah, N. B., & Singh, A. (2021). PeerReview4All: Fair and accurate reviewer assignment in peer review. *Journal of Machine Learning Research*, 1–66.

Sun, Y., Barber, R., Gupta, M., Aggarwal, C. C., & Han, J. (2011). Co-author relationship prediction in heterogeneous bibliographic networks. In *2011 international conference on advances in social networks analysis and mining* (pp. 121–128).

Sun, Y., Han, J., Yan, X., Yu, P. S., & Wu, T. (2011). PathSim: Meta path-based top-k similarity search in heterogeneous information networks. *VLDB Endowment*, *4*(11), 992–1003. http://dx.doi.org/10.14778/3402707.3402736.

Tabah, A. N. (1999). Literature dynamics: Studies on growth, diffusion, and epidemics. *Annual Review of Information Science and Technology*, *34*, 249–286.

Tai, K. S., Socher, R., & Manning, C. D. (2015). Improved semantic representations from tree-structured long short-term memory networks. In *53rd annual meeting of the association for computational linguistics and the 7th international joint conference on natural language processing, Beijing, China* (pp. 1556–1566).

Tan, S., Duan, Z., Zhao, S., Chen, J., & Zhang, Y. (2021). Improved reviewer assignment based on both word and semantic features. *Information Retrieval Journal*, *24*(3), 175–204.

Tang, J. (2016). AMiner: Toward understanding big scholar data. In *9th ACM international conference on WSDM, San Francisco, California, USA* (p. 467).

Tang, W., Tang, J., Lei, T., Tan, C., Gao, B., & Li, T. (2012). On optimization of expertise matching with various constraints. *Neurocomputing*, *76*(1), 71–83. http://dx.doi.org/10.1016/j.neucom.2011.04.039.

Tang, W., Tang, J., & Tan, C. (2010). Expertise matching via constraint-based optimization. In *2010 IEEE/WIC/ACM international conference on web intelligence and intelligent agent technology, Toronto, ON, Canada* (pp. 34–41).

Tang, X., & Zhang, Z. (2008). Paper review assignment based on human-knowledge network. In *IEEE international conference on systems, man and cybernetics* (pp. 102–107). Singapore: IEEE.

Tang, J., Zhang, J., Yao, L., Li, J., Zhang, L., & Su, Z. (2008). ArnetMiner: Extraction and mining of academic social networks. In *14th ACM SIGKDD international conference on knowledge discovery and data mining, New York, NY, USA* (pp. 990–998).

Tong, B. (2018). *Research on expert matching recommendation system based on collaborative deep learning* (Master's thesis), Kunming University of Science and Technology.

Vazirani, V. V. (2013). *Approximation algorithms*. Springer Science & Business Media.

Wang, C., Han, J., Jia, Y., Tang, J., Zhang, D., Yu, Y., et al. (2010). Mining advisor-advisee relationships from research publication networks. In *2010 ACM SIGKDD conference on knowledge discovery and data mining, Washington DC, USA*.

Wang, F., Shi, N., & Chen, B. (2010). A comprehensive survey of the reviewer assignment problem. *International Journal of Information Technology and Decision Making*, *9*(04), 645–668. http://dx.doi.org/10.1142/S0219622010003993.

Wang, F., Zhou, S., & Shi, N. (2013). Group-to-group reviewer assignment problem. *Computers & Operations Research*, *40*(5), 1351–1362. http://dx.doi.org/10.1016/j.cor.2012.08.005.

Wohlin, C. (2014). Guidelines for snowballing in systematic literature studies and a replication in software engineering. In *Proceedings of the 18th international conference on evaluation and assessment in software engineering* (pp. 1–10).

Wu, Y. (2018). *Research on technical expert recommendation system based on semantics* (Master's thesis), Hangzhou Dianzi University.

Wu, S., Leong Hou, U., Bhowmick, S. S., & Gatterbauer, W. (2017). Conflict of interest declaration and detection system in heterogeneous networks. In *2017 ACM conference on information and knowledge management* (pp. 2383–2386). Singapore, Singapore: ACM.

Wu, S., Leong Hou, U., Bhowmick, S. S., & Gatterbauer, W. (2018). PISTIS: A conflict of interest declaration and detection system for peer review management. In *2018 international conference on management of data* (pp. 1713–1716). Houston, TX, USA: ACM.

Xie, P., Yang, D., & Xing, E. (2015). Incorporating word correlation knowledge into topic modeling. In *2015 conference of the north American chapter of the association for computational linguistics: human language technologies, Denver, Colorado* (pp. 725–734).

Xu, Y.-H. (2010). *A network based approach for expertise recommendation in academic contexts* (Ph.D. thesis), University of Science and Technology of China.

Xu, Y., Ma, J., Sun, Y., Hao, G., Xu, W., & Zhao, D. (2010). A decision support approach for assigning reviewers to proposals. *Expert Systems with Applications*, *37*(10), 6948–6956. http://dx.doi.org/10.1016/j.eswa.2010.03.027.

Xu, G., & Wang, H. (2011). The development of topic models in natural language processing. *Chinese Journal of Computers*, *8*, 1423–1436. http://dx.doi.org/10.3724/SP.J.1016.2011.01423.

Xu, Y., Zhao, H., Shi, X., & Shah, N. B. (2019). On strategyproof conference peer review. In *28th international joint conference on artificial intelligence*.

Xu, Y., & Zuo, X. (2016). A LDA model based text-mining method to recommend reviewer for proposal of research project selection. In *2016 13th international conference on service systems and service management (ICSSSM)* (pp. 1–5). IEEE.

Yan, S., Jin, J., Geng, Q., Zhao, Y., & Huang, X. (2017). Utilizing academic-network-based conflict of interests for paper reviewer assignment. *International Journal of Knowledge Engineering*, *2*(3), 65–73. http://dx.doi.org/10.18178/ijke.2017.3.2.089.

Yang, C., Liu, T., Yi, W., Chen, X., & Niu, B. (2020). Identifying expertise through semantic modeling: A modified BBPSO algorithm for the reviewer assignment problem. *Applied Soft Computing*, *94*, Article 106483.

Yarowsky, D., & Florian, R. (1999). Taking the load off the conference chairs-towards a digital paper-routing assistant. In *Joint SIGDAT conference on empirical methods in natural language processing and very large corpora*.

Yimam-Seid, D., & Kobsa, A. (2003). Expert-finding systems for organizations: Problem and domain analysis and the DEMOIR approach. *Journal of Organizational Computing and Electronic Commerce*, *13*(1), 1–24. http://dx.doi.org/10.1207/S15327744JOCE1301_1.

Yong, Y., Yao, Z., & Zhao, Y. (2021). A framework for reviewer recommendation based on knowledge graph and rules matching. In *2021 IEEE international conference on information communication and software engineering (ICICSE)* (pp. 199–203). http://dx.doi.org/10.1109/ICICSE52190.2021.9404099.

Zablocki, J., & Lee, R. (2012). Auto-assign: An implementation to assign reviewers using topic comparison in START. In *International conference on e-learning, e-business, enterprise information systems, and e-government* (p. 1). The Steering Committee of The World Congress in Computer Science, Computer.

Zhang, J., Tang, J., & Li, J. (2007). Expert finding in a social network. In *International conference on database systems for advanced applications* (pp. 1066–1069). Bangkok, Thailand: Springer.

Zhang, Y., Zhang, F., Yao, P., & Tang, J. (2018). Name disambiguation in aminer: Clustering, maintenance, and human in the loop. In *24th ACM SIGKDD international conference on knowledge discovery and data mining, London, United Kingdom* (pp. 1002–1011).

Zhang, D., Zhao, S., Duan, Z., Chen, J., Zhang, Y., & Tang, J. (2020). A multi-label classification method using a hierarchical and transparent representation for paper-reviewer recommendation. *ACM Transactions on Information Systems*, *38*(1), 1–20. http://dx.doi.org/10.1145/3361719.

Zhao, S., Zhang, D., Duan, Z., Chen, J., Zhang, Y.-p., & Tang, J. (2018). A novel classification method for paper-reviewer recommendation. *Scientometrics*, *115*(3), 1293–1313. http://dx.doi.org/10.1007/s11192-018-2726-6.