Conference Paper Assignment Problem - A new System for Recommending and Assigning Reviewers to Scientific Articles

Ana Carolina Ribeiro[1,2]

[1]DEI/FEUP - Department of Informatics Engineering, Faculty of Engineering of the University of Porto, R. Dr. Roberto Frias, 4200-465 Porto, Portugal
[2]LIACC - Artificial Intelligence and Computer Science Laboratory, R. Dr. Roberto Frias, 4200-465 Porto, Portugal

Abstract

The recommendation and assignment of reviewers and articles is an emerging topic in the academic community and is known as the Reviewer Assignment Problem (RAP). This problem can be seen as a version of the well-known Generalized Assignment Problem (GAP), while the Conference Paper Assignment Problem (CPAP) is a specific case of RAP. The thesis proposal presented is centred on CPAP. The main objective of the doctoral thesis proposal is the development and validation of multi-information recommendation systems (extracting information from the content of articles and reviewers) capable of efficiently assigning expert reviewers to scientific articles submitted to a conference.

Keywords

Reviewer Assignment Problem, Conference Paper Assignment Problem, Peer-Review System, Recommendation System

1. Introduction

The recommendation and assignment of reviewers to articles, generally called in the literature a Reviewer Assignment Problem (RAP), has become an essential topic in the academic world. RAP can be seen as a version of the Generalized Assignment Problem (GAP), and Conference Paper Assignment Problem (CPAP) is a specific case of RAP [1].

In most cases, articles should be assigned to reviewers according to the following conditions: (i) Each manuscript should be assigned to a certain number of reviewers, *ai*, defined by the team responsible for assigning the reviewers; (ii) as far as possible, each article should be assigned to reviewers who are experts in the field. A specific limit T can be defined to identify the qualification of the reviewers; (iii) each reviewer should be assigned to, at most, a certain number of articles, *bj*, defined by the team responsible for assigning the reviewers; (iv) each reviewer should be assigned to approximately the same number of articles, to balance their workload.

Doctoral Consortium of the 45th German Conference on Artificial Intelligence (KI2022), September 20, 2022, Trier (Virtual), Germany
up201602840@edu.fe.up.pt (A. C. Ribeiro)
0000-0003-1557-654X (A. C. Ribeiro)

Given a set $P = \{1, ..., |P|\}$ of manuscripts and a set $R = \{1, ..., |R|\}$ of reviewers, c_{ij} denote the matching degree of manuscript *i* for reviewer *j*, where *i* belongs a *P* and *j* belongs a *R*. A binary variable x_{ij}, whose value is 1 if manuscript *i* is assigned to reviewer *j* and 0 otherwise, RAP is formulated by the Following Integer Programming formulation:

$$RAP : max \sum_{i \in P} \sum_{j \in R} c_{ij} x_{ij} \tag{1}$$

Subject to

$$\sum_{j \in R} x_{ij} = a_i. \tag{2}$$

$$\sum_{i \in P} x_{ij} \leq b_j. \tag{3}$$

$$x_{ij} \in \left\{ \left\lfloor \frac{c_{ij}}{T} \right\rfloor \right\}. \tag{4}$$

$$x_{ij} = 0, 1. \tag{5}$$

The objective function (equation 1) maximizes the total matching degree of the assignment. Constraints (equation 2) and (equation 3) ensure that conditions (i) and (ii) are satisfied, respectively. Constraint (equation 4) along with (equation 5) prevent a reviewer from being assigned to a manuscript whenever c_{ij} is smaller than the given threshold *T*. Constraint (equation 6)instead of (equation 4) is brought into the mathematical model to ensure that at least one reviewer whose matching degree for manuscript *i* is greater than or equal to *T*.

$$max_{j \in R} \{c_{ij} x_{ij}\} \geq T \tag{6}$$

.

The peer-review system is considered the main mechanism for quality control of scientific publications [2], with the potential to contribute to the rigour of the work published in the academic community [3]. The challenging task of the peer-review process is to recommend and assign suitable reviewers whose interests and research profiles fit appropriately in the submissions [4]. Often these tasks are still done manually by editors or conference organizers. However, there are obvious flaws in the method of selecting and manually assigning reviewers to articles. Firstly, this is a time-consuming process. The committee of conferences or journals needs to retrieve databases from expert researchers and find the most suitable ones for each submitted article.

Second, it is common to recommend expert researchers in the same research area as the article to be reviewed. However, there are many other indicators that should be considered when selecting reviewers, for example, publications, research projects, and patents. Also, teams that select reviewers use titles, rewards, and status to assess the quality of the experts. Third, the process of manually selecting and assigning reviewers to articles ignores possible relationships between reviewers and authors. Finally, since it is necessary to manage a large amount of

information about the reviewers and the articles submitted according to human and subjective criteria, bias may occur in the recommendation and assignment of researchers.

According to the problems presented above, there is a need to apply intelligent technologies capable of analyzing data, extracting valuable information from documents and unstructured texts, and thus automatically recommending and assigning the most appropriate researchers to scientific articles. This doctoral study aims to solve the CPAP through the development of a system capable of recommending the most appropriate and expert reviewers and efficiently assigning them to scientific articles, according to the established constraints.

2. Motivations

The development of automatic and efficient mechanisms for recommending experts has a high economic value. These mechanisms allow the increase in productivity, which transposes into finding experts in a valuable search time and increasing efficiency, ensuring the highest possible level of correspondence between their profile and task.

In the scientific world, these expert recommendation mechanisms provide the added value of inferring and identifying research teams and working groups, especially between researchers from different institutions; discovering emerging talents with low visibility; supporting students, allowing them to identify the best supervisors and co-supervisors of master or doctoral degrees.

Also, it is essential to keep the quality standard of science high. The review of scientific articles in scientific conferences and journals, and scientific projects is one of the best-known tasks for expert researchers. The recommendation and assignment of the most suitable experts to the articles are based on human and subjective criteria, which can generate wrong decisions, such as the rejection of excellent scientific work or a potentially successful project proposal. These decisions can have significant and adverse effects on the quality of the scientific standard, namely, the quality of published studies, researchers' careers, and the reputation of conferences and journals.

3. Objectives

The present PhD study aims to develop a Recommendation System (RS) based on constraint programming solvers applied to the CPAP case. A more detailed description of the objectives is as follows:

1. Development of a data collection system capable of collecting relevant input information from researchers who are candidates for reviewers and from scientific articles.
2. Development of an Information Extraction system combining deep neural networks and advanced NLP techniques in extracting the semantic representation of scientific articles' content and reviewers' information gathered.
3. Development of an expert RS capable of extracting the expertise level of re-viewers and finding the ranking of expert reviewers in a given research topic (s)/area(s).
4. Search, compare and select available open-source next-generation constraint programming technology solvers and other recent optimization approaches to deliver the best trade-off between efficiency and ease of integration and customization.

5. Integration of the constraint programming-based constraint-solving module in the developed RS.

4. Research Questions

According to the previous context, motivation and objectives, some research questions arise, namely:

1. How can the CPAP be formulated as a constraint programming problem?
2. What is the effect of deep neural networks combined with natural language processing techniques in extracting the semantic representation from the content of scientific articles?
3. Do CP solvers frameworks allow handling reviewers' constraints in the context of assigning articles more efficiently when compared to other current solving approaches?
4. Do CP solvers frameworks improve the balance between efficiency and ease of integration and customization compared to different methods?

5. Related Work

RAP has two main phases – 1) finding/recommending expert reviewers and 2) assigning reviewers to submitted manuscripts. These problems are different and therefore require different approaches. In phase 1), the main objective is to compute the article-reviewer similarity factors depending on the method (implicit or explicit) chosen to describe the articles and the competencies of the reviewers. While in phase 2) the main objective is the effective assignment/allocation of expert reviewers to scientific articles.

According to the literature, different approaches can be applied in phase one, namely: decision support systems [5], [6] recommendation systems [7], [8] and machine learning-oriented approaches [9], [10]. In phase two, we can find studies developed with optimization approaches based on heuristics and meta-heuristics [11], [12], fuzzy approaches [13], answer set programming [14], and others.

Finding expert reviewers (phase 1) generally requires the development of relevant tasks such as collecting data on the reviewers and, sometimes, on the authors of submitted articles; the construction of the profiles of reviewers and articles submit-ted and, finally, the computation of the similarity between reviewers and authors. The approaches based on DSSs and RSs have several similarities in their processes. For example, the development of the reviewers' profile and articles submitted through specific measures (quality, relevance, authority, diversity, among others) have been shown to achieve good results in recommending reviewers [5], [15]; The information retrieval process is also common in the approaches, relevant and includes the classification of publications, the author's disambiguation (for example, rule-based algorithms, clustering-based algorithms), the extraction of relevant information (for example, LDA algorithm, Doc2Vec model, TF-IDF), among others; finally, in the similarity calculation between reviewers and articles, the similarity of cosine ordered weighted averaging aggregation function and Kullback-Leibler di-vergence are the most commonly used techniques and with better results.

The problem of assigning/allocating reviewers to submitted articles (phase 2) presents several types of approaches tested by researchers. The approaches based on heuristic and metaheuristic algorithms have a strong presence in researchers' studies [11], [16]. Also, greedy and genetic algorithms are the best known and selected by researchers to try to solve the attribution problem. Other approaches, such as the fuzzy approach [17], ASP [14] or integer linear programming, also aroused the interest of researchers and showed good results.

6. Approach

The defined approach proposes the construction of knowledge about the expertise of researchers and the recommendation and assignment of the most appropriate expert reviewers for scientific papers. The proposed system comprises four modules:

1. Data Collection Module – The main purpose of this module is to collect data from multiple data sources (open databases and Web) to create a database with information about researchers (affiliations, co-authors, citations, number of publications, awards, among others) and scientific articles (metadata).
2. Information Extraction Module – The main objective of this module is to extract relevant information from the data, transforming it into more significant representations of its semantic content and, consequently, easier to analyse. This module combines advanced NLP techniques and deep neural networks for the extraction of semantic and contextual information from the collected data.
3. RS of Experts Module – This module focuses on the recommendation of reviewers' experts in a given research topic/area. The relevant information extracted in the previous module allows the construction of reviewers' profiles (based on quality, reputation, and expertise) and scientific articles (based on metadata and key insights). Subsequently, the objectives of RS are to extract the behavioural model of experts, compute the expertise level of each reviewer and, finally, present the ranking of experts.
4. Assignment Module – In this module, the main objectives are the definition of the constraints of the assignment of reviewers to scientific articles and the implementation of a sophisticated optimization approach (constraint programming) to enforce the adequacy of the recommended reviewers to the input constraints and requirements.

7. Evaluation

The system validation starts with the comparison of the results obtained in this study with the results found in the literature of approaches used by other researchers to solve the RAP. Furthermore, we intend to validate the system in a real environment, namely, at a scientific conference. At this stage, we want to compare the results of the recommendation and attribution of our system with the results of the conference system. In addition, we intend to assess the satisfaction of reviewers through a survey.

8. Contributions

The main expected contributions of this PhD study are:

1. Development of an RS to be integrated into a constraint resolution module based on sophisticated constraint programming and optimization approaches to handle constraints more efficiently when compared to other current constraint resolution approaches.
2. Study the effect of using Deep Neural Networks in extracting semantic representation of scientific papers' content and researchers' contents for the recommendation task.

9. Acknowledgments

This work was financially supported by: Base Funding -UIDB/00027/2020 of the Artificial Intelligence and Computer Science Laboratory – LIACC - funded by national funds through the FCT/MCTES (PIDDAC). The first author is supported by FCT (Portuguese Foundation for Science and Technology) under grant PD/BD/2020.04698.BD.

References

[1] X. Zhao, Y. Zhang, Reviewer assignment algorithms for peer review automation: A survey, Information Processing & Management 59 (2022) 103028. URL: https://www.sciencedirect.com/science/article/pii/S0306457322001388. doi:https://doi.org/10.1016/j.ipm.2022.103028.

[2] L. Bornmann, Scientific peer review, Annual Rev. Info. Sci & Technol. 45 (2011) 197–245.

[3] R. H. Fletcher, S. W. Fletcher, 4 : The effectiveness of journal peer review, 2003.

[4] J. Goldsmith, R. H. Sloan, The ai conference paper assignment problem, in: AAAI Conference on Artificial Intelligence, 2007.

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

[6] A proactive decision support system for reviewer recommendation in academia, Expert Systems with Applications 169 (2021) 114331. doi:https://doi.org/10.1016/j.eswa.2020.114331.

[7] C. K. Kreutz, R. Schenkel, Revaside: Assignment of suitable reviewer sets for publications from fixed candidate pools, iiWAS2021, Association for Computing Machinery, New York, NY, USA, 2022, p. 57–68. URL: https://doi.org/10.1145/3487664.3487673. doi:10.1145/3487664.3487673.

[8] Y. Yong, Z. Yao, Y. Zhao, Beyond accuracy: A feature crossing method for chinese thesis reviewer recommendation, in: 2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2021, pp. 1151–1158. doi:10.1109/SMC52423.2021.9658668.

[9] L. Cagliero, P. Garza, A. Pasini, E. Baralis, Additional reviewer assignment by means of weighted association rules, IEEE Transactions on Emerging Topics in Computing 9 (2021) 329–341. doi:10.1109/TETC.2018.2861214.

[10] D. Zhang, S. Zhao, Z. Duan, J. Chen, Y. Zhang, J. Tang, A multi-label classification method using a hierarchical and transparent representation for paper-reviewer recommendation, ACM Trans. Inf. Syst. 38 (2020). URL: https://doi.org/10.1145/3361719. doi:10.1145/3361719.

[11] C. Yang, T. Liu, W. Yi, X. Chen, B. Niu, Identifying expertise through semantic modeling: A modified bbpso algorithm for the reviewer assignment problem, Applied Soft Computing 94 (2020) 106483. URL: https://www.sciencedirect.com/science/article/pii/S1568494620304221. doi:https://doi.org/10.1016/j.asoc.2020.106483.

[12] D. K. Pradhan, J. Chakraborty, P. Choudhary, S. Nandi, An automated conflict of interest based greedy approach for conference paper assignment system, Journal of Informetrics 14 (2020) 101022. URL: https://www.sciencedirect.com/science/article/pii/S1751157719301373. doi:https://doi.org/10.1016/j.joi.2020.101022.

[13] K. Mittal, A. Jain, K. S. Vaisla, D. K. Tayal, A novel method for reviewer assignment problem based on reviewers' profile and fuzzy graph connectivity measure, in: 2020 International Conference on Intelligent Engineering and Management (ICIEM), 2020, pp. 386–391. doi:10.1109/ICIEM48762.2020.9160042.

[14] G. Amendola, C. Dodaro, N. Leone, F. Ricca, On the application of answer set programming to the conference paper assignment problem, in: G. Adorni, S. Cagnoni, M. Gori, M. Maratea (Eds.), AI*IA 2016 Advances in Artificial Intelligence, Springer International Publishing, Cham, 2016, pp. 164–178. doi:https://doi.org/10.1007/978-3-319-49130-1_13.

[15] M. Maleszka, B. Maleszka, D. Król, M. Hernes, D. M. L. Martins, L. Homann, G. Vossen, A modular diversity based reviewer recommendation system, in: P. Sitek, M. Pietranik, M. Krótkiewicz, C. Srinilta (Eds.), Intelligent Information and Database Systems, Springer Singapore, Singapore, 2020, pp. 550–561. doi:https://doi.org/10.1007/978-981-15-3380-8_48.

[16] Y. Kalmukov, An algorithm for automatic assignment of reviewers to papers, Scientometrics 124 (2020) 1811–1850. URL: https://doi.org/10.1007/s11192-020-03519-0. doi:10.1007/s11192-020-03519-0.

[17] K. Mittal, A. Jain, K. S. Vaisla, Understanding reviewer assignment problem and its issues and challenges, in: 2019 4th International Conference on Internet of Things: Smart Innovation and Usages (IoT-SIU), 2019, pp. 1–6. doi:10.1109/IoT-SIU.2019.8777727.