Knowledge-Based Systems 106 (2016) 164–178

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

A recommender system of reviewers and experts in reviewing problems

Jarosław Protasiewicz[a,*], Witold Pedrycz[b], Marek Kozłowski[a], Sławomir Dadas[a], Tomasz Stanisławek[a], Agata Kopacz[a], Małgorzata Gałężewska[a]

[a] National Information Processing Institute, Warsaw, Poland
[b] Department of Electrical & Computer Engineering, University of Alberta, Edmonton T6R 2V4 AB Canada & Department of Electrical and Computer Engineering, Faculty of Engineering, King Abdulaziz University, Jeddah, 21589, Saudi Arabia and Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland

ARTICLE INFO

Article history:
Received 9 February 2016
Revised 18 May 2016
Accepted 21 May 2016
Available online 24 May 2016

Keywords:
Reviewer assignment problem
Recommender system
Data acquisition
Information retrieval
Content-based filtering

ABSTRACT

In this study, we propose the architecture of a content-based recommender system aimed at the selection of reviewers (experts) to evaluate research proposals or articles. We introduce a comprehensive algorithmic framework supported by various techniques of information retrieval. We propose a well-rounded methodology that explores concepts of data, information, knowledge, and relations between them to support a formation of a suitable recommendation. In particular, the developed system helps collecting data characterizing potential reviewers, retrieving information from relational and unstructured data, and formulating a set of recommendations. The designed system architecture is modular from the functional perspective and hierarchical from the technical point of view. Each essential part of the system is treated as a separate module, whereas each layer supports a certain functionality of the system. The modularity of the architecture facilitates its maintainability. The process of information retrieval includes classification of publications, author disambiguation, keywords extraction, and full-text indexing, whereas recommendations are based on the combination of a cosine similarity between keywords and a full-text index. The proposed system has been verified through a case study run at the National Center for Research and Development, Warsaw, Poland.

1. Introduction

When people think about a reviewing process they usually focus on articles because they are looking for suitable journals as a way to disseminate high quality and timely knowledge. Not only publications are important in the science world. Research and development projects are crucial to the improvement of security and prosperity of countries, organizations and companies. As innovative activities are highly risky, therefore, these projects are predominantly financed by public funds, either sponsors' or philanthropists' donations, or private companies searching for new technologies. Forasmuch as granting bodies have a finite amount of money to redistribute they need to prioritize applications in accordance with social needs, economic reasons, scientific goals, and quality of proposals. Usually, evaluation of project proposals and distribution of available funding are based on reviews prepared by academic reviewers and professional experts. On the other hand, already completed projects need an assessment to check whether their objectives and targets were achieved. Moreover, companies look for professionals to assess investment plans, realize complicated projects, and alike. Finally, these demands imply a need for coping with the recommendation issue of reviewers, experts, and professionals suitable for a specified problem, which could be a project proposal, an article, a completed project, or just a demand for professionals.

It should be emphasized that a lack of conflict of interest, independence, and competencies of people producing reviews, recommendations, and opinions are crucial to the quality of evaluations. However, we should also realize, that those people have limited knowledge, experience, and perspective of looking at the works of other people. These limitations can cause a misinterpretation of the author's viewpoint and can lead to the rejection of an excellent scientific work, or a potentially successful project proposal. In order to understand the nature of the issues mentioned above, we should consider the psychological and social aspects of the review process. More specifically, there are three heuristics of cognitive distortions: availability, anchoring, and representativeness [1]. What is meant by this is that some biases may occur while

* Corresponding author.
E-mail addresses: jaroslaw.protasiewicz@opi.org.pl, jaroslaw.protasiewicz@gmail.com (J. Protasiewicz), wpedrycz@ualberta.ca (W. Pedrycz).

http://dx.doi.org/10.1016/j.knosys.2016.05.041
0950-7051/

choosing reviewers, experts, and professionals manually. Thus, this process should be supported by disinterested and automatic methods. Moreover, granting bodies expect that selection of reviewers will be unbiased and the time of processing applications should be as short as possible. To fulfill the requirements, we realized that we have to provide a possible large dataset of scientists as well as efficient methods for selecting them. These are the reasons why we decided to work on a recommender system of reviewers and experts.

We may view the assignment of people to problems as an extended version of the problem of generalized assignment. There are a number of sophisticated solutions to this problem, which are well documented in the literature (for details, see Section 2). However, some approaches contain only theoretical propositions, or the existing experimental evidence may be insufficient to be considered reliable, e.g. they are tested on data originating from a particular conference. Despite these concerns, some developments of practical relevance are worth noting. For instance, Chien and Chen [2] presented an empirical study in the semiconductor industry, Rodriguez and Bollen [3] proposed a fully automatic solution tested on data coming from a selected conference, whereas Wang et al. [4] used real data but only for algorithms checking. Decision support systems proposed by Tian et al., Fan et al., Sun et al. and Xu et al. [5–8] are the most interesting solutions as they have been reported as practically used indicating their level of maturity and usefulness. However, these tools require human assistance, for instance, they need a manual commitment in classifying reviewers and proposals, and eventually preparing knowledge rules. It should be noted that Fan et al. [6] introduced automatic proposals grouping process, which brought some improvement to this area. Another problem is that algorithms take advantage of human experience expressed in a structured way, however, in reality, such data may be nonstructural. A good example here is keywords describing people's experience and objects like a manuscript, a project, etc. It could be difficult to match efficiently people and objects using, for example, cosine measure between keywords because different terms can have similar meaning, so a strict fit is impossible. Tayal et al. [9] proposed fuzzy logic to circumvent this problem. A more promising way to follow would be the use of unstructured data or semantic relations between terms coming from the Computer Science Bibliography (DBLP) [10], manuscript references [3], and home pages [11] or the application of context-aware systems [12].

The observations made above underline a genuine need for further research and development of new approaches to the selection of reviewers and experts. In this study, we put forward a proposal along with associated experimental studies aimed at capturing experience, intuition, and informal observations. Firstly, expertise of people may be described not only in terms structured data, e.g. keywords, but also unstructured data could be useful with this regard. There is some rationale assumption in the present literature, and this is coincident with our observations. Secondly, we believe that areas of expertise declared by experts might be inconsistent with those inferred from their professional track record. Thirdly, the previous research has proposed algorithms that indeed need some human assistance. However, we are convinced that it is possible to choose automatically reviewers or experts using a recommender system that works autonomously without any manual adjustment.

In this study, having these assumptions in mind, we propose a methodology that explores concepts of data, information, knowledge, and relations between them. The methodology supports a formation of a recommender system, which collects data concerning researchers coming from various sources including public databases and the Internet. Next, information is retrieved from relational and unstructured data to build expertise profiles of researchers and professionals. Finally, it recommends reviewers and experts for a specified problem on a basis of knowledge concerning potential candidates. The similarity between a problem under discussion and peoples' expertise is quantified through the combination of cosine measure and a full-text index. It should be noted that Basu et al., Flach et al. and Ryabokon et al. [11,13,14] used a cosine measure for matching between expertise of people and a problem. However, our approach augments this as the similarity not only involves but combines it cosine similarity between keywords with a full-text index, therefore incorporating unstructured data into the final recommendation.

Recommender systems are mainly used in e-commerce [15]. We propose a new area for application such systems, which is the recommendation of people (experts and reviewers) who may be able to assess an article or a project. Among known types of filtering used in these systems, i.e. collaborative, demographic, content-based, and hybrid (for more, see Section 2), our approach utilizes content-based filtering. The collaborative recommendation systems make suggestions based on the past behavior of other users and the similarity between users and items. This kind of procedure demands historical data, i.e. users' ratings. In our case, there is the lack of historical data describing reviewer's assignments. Therefore, we decided to build the content-based recommendation engine, where we can avoid a cold-start problem. The key of the systems based on the content is that information coming from recommended objects is similar to the user's profile data. Such tools are mainly used to recommend documents, Web pages, publications, jokes, or news. Some examples are SYSKILL & WEBERT, which recommend Web pages or PTV, which recommends TV programs to the user [16]. We have not found any other content-based system used to recommend reviewers or experts for project proposals or articles. Our work seems to be the pioneer according to those two conditions, i.e. the reviewers' selection domain and the content-based type.

The main objective of this study is to design an architecture of a content-based recommender system along with a comprehensive algorithmic framework that supports a thorough information retrieval and offers a sound framework of ranking potential reviewers. The system should work autonomously without any or with a limited human input. One should stress, though, that the proposed recommender system is meant to support human and offering a decision-support environment.

The study elaborates on the overall architecture of the system and functionalities of each module. The main outcome of this study is the recommender system of reviewers and experts, which is the improved version of the decision support system for selection of reviewers [17]. Moreover, the recommender system was deployed in the National Centre for Research and Development of Poland (NCBR).

The paper is structured as follows. Section 3 presents an overview of the architecture of the proposed system, whereas Section 4 elaborates on the pertinent details showing the underlying algorithmic aspects of the system. Section 5 covers specific technical details as well as some experimental results. Finally, conclusions are presented.

2. Related studies

A reviewing of scientific works has been present since the Middle Ages. Its first mention appeared in IX century when Ishaq bin Ali al-Rahwi in his book *Ethics of the Physician* suggested that physician should rate methods used in treatment to improve their standards and quality [18]. In modern times (XVIII century), Henry Oldenburg established a review process of *Philosophical Transactions* of the *Royal Society* magazine, where a group of experts in a given field evaluated manuscripts to take a publishing decision.