Information Processing and Management 50 (2014) 361–383

Contents lists available at ScienceDirect

Information Processing and Management

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

Expert group formation using facility location analysis

Mahmood Neshati [a,*], Hamid Beigy [a], Djoerd Hiemstra [b]

[a] Department of Computer Engineering, Sharif University of Technology, Tehran, Iran
[b] Database Research Group, Electrical Engineering, Mathematics and Computer Science (EEMCS) Department, University of Twente, The Netherlands

ARTICLE INFO

Article history:
Received 23 February 2013
Received in revised form 30 September 2013
Accepted 8 October 2013
Available online 5 November 2013

Keywords:
Expert group formation
Expert finding
Topic model
Facility location analysis
Greedy approach
Linear programming

ABSTRACT

In this paper, we propose an optimization framework to retrieve an optimal group of experts to perform a multi-aspect task. While a diverse set of skills are needed to perform a multi-aspect task, the group of assigned experts should be able to collectively cover all these required skills. We consider three types of multi-aspect expert group formation problems and propose a unified framework to solve these problems accurately and efficiently. The first problem is concerned with finding the top k experts for a given task, while the required skills of the task are implicitly described. In the second problem, the required skills of the tasks are explicitly described using some keywords but each expert has a limited capacity to perform these tasks and therefore should be assigned to a limited number of them. Finally, the third problem is the combination of the first and the second problems. Our proposed optimization framework is based on the *Facility Location Analysis* which is a well known branch of the *Operation Research*. In our experiments, we compare the accuracy and efficiency of the proposed framework with the state-of-the-art approaches for the group formation problems. The experiment results show the effectiveness of our proposed methods in comparison with state-of-the-art approaches.

1. Introduction

The success of a project directly depends on the expertise of the people who are involved. Since the assignment of experts to a task/project must be based on both the required skills of the project and knowledge about the expertise of all candidate experts, it is not an easy task and it is challenging to optimize the assignment. As a result, expert group formation offers a new direction in research and also poses some brand new challenges. Here, the key problem is how to assign a group of experts to a given set of tasks/projects. The problem has attracted considerable interest from different domains. For example, several works have been made for conference paper-reviewer assignment (Karimzadehgan & Zhai, 2012; Karimzadehgan, Zhai, & Belford, 2008; Mimno & McCallum, 2007; Taylor, 2008), expert group formation in social networks (Kargar & An, 2011; Lappas, Liu, & Terzi, 2009) and optimal team formation in *Operation Research* problems (Wi, Oh, Mun, & Jung, 2009; Baykasoglu, Dereli, & Das, 2007).

In real scenarios, while several and sometimes diverse skills are needed to perform a task successfully and completely, these skills are implicitly expressed in the task descriptions. Besides the required skills of a task, in many cases, the relevant skills of experts are also implicitly reflected in their resume. The implicit notion of required skills/expertise aspects makes it difficult to assign an optimal group of experts to a given multi-aspect task. The main challenges of the expert group formation problem are:

* Corresponding author. Tel.: +98 2166166648.
E-mail addresses: neshati@mehr.sharif.edu (M. Neshati), beigy@sharif.edu (H. Beigy), d.hiemstra@utwente.nl (D. Hiemstra).

0306-4573/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.ipm.2013.10.001

Implicit Notion of Aspects: Textual description of projects can only implicitly express the required skills of them, thus a method is needed to transform the textual description of a project into the set of required skills of that project. Similarly, because of implicit notion of expertise, a method is needed to transform the expertise documents (e.g. resume, professional profile, etc.) of each expert into the set of his/her skills.
Coverage Condition: In an ideal expert matching, all required skills of a project should be covered by the union of the skills of the assigned members in a complementary manner.
Confidence Condition: In an ideal expert matching, besides covering all required skills of a project, it is preferable that each member of the group individually be able to cover as many as possible the required skills of that project.
Load Balancing: In many applications of expert group formation, there is a limit on the number of tasks to be done by each expert. To balance the load and conform to the capacity of each expert, it is necessary to set up the problem as to simultaneously assign tasks to all the available experts with consideration of task-load balancing.

As a case study of the expert group formation problem, we consider the problem of review assignment. Review assignment is a common task that many people such as conference organizers, journal editors, and grant administrators would have to do routinely. In this problem, top k relevant reviewers (i.e. a group of experts with k members) should be assigned to each paper (i.e. task) such that all above mentioned criteria are satisfied. Firstly, the required skills for reviewing a paper can be explicitly determined by some keywords or can be inferred from the abstract/body of the paper. Secondary, the related research areas/skills of each reviewer (i.e. expert) can be explicitly expressed by some keywords or can be inferred from his/here previous publications. Thirdly, in an ideal matching, the assigned group of reviewers for a paper should be able to cover all aspects of that paper in a complementary manner. Fourthly, it is preferable that each assigned reviewer of a paper be able to cover as many as possible aspects of the paper. Finally, each member of the program committee (i.e. each reviewer) can only be involved in the review process of a limited number of papers.

Considering expert group formation as a top k retrieval problem, it can be solved by using a standard retrieval model (e.g. the baseline language model proposed in Karimzadehgan et al. (2008)), which computes the relevance score of each expert *individually* and then returns the k experts with the highest relevance scores. However, as the inter-relationships among the relevant experts are ignored, the top k search results are often quite alike and cannot cover all required aspects of a given task. In order to resolve the query aspects and avoid the information redundancy, it is necessary to optimize the top k search results collectively.

In this paper, we formalize the expert matching problem within the unified framework of *Facility Location Analysis* (FLA) (Gonzalez, 2007) taken from *Operation Research*, as a way to account and optimize the expert assignment. In the facility location problem, given a set of customer "locations" D, we would like to find a subset $S \subset D$ to open k "facilities" there, so as to optimize a graph-theoretic objective function that is dependent on the cost of opening a facility at each location and also the distance between each pair of customer and facilities. Specifically, facilities such as warehouses, hospitals, and fire stations, should be placed as close as possible to their customers. Since a customer would just go to the closest facility, there is a competitive relationship among those k-facilities. Similarly, in our proposed framework, we consider the top-k reviewers of each paper as the desirable facilities to be placed as close as possible to their customers (i.e. aspects of papers).

We show that our proposed method can improve the performance of expert matching in comparison with the state-of-the-art techniques for multi aspect/ skill expert matching such as *Greedy Next Best* (Karimzadehgan et al., 2008) and *Integer linear programming* (Karimzadehgan & Zhai, 2012). According to different conditions of the expert matching problem, we define three problems that can be solved by the proposed frame work of FLA. These problems are modeled using the unified framework of facility location analysis. In these problems, given a set of N papers and M reviewers, each paper should be assigned to a group of exactly k reviewers. The above mentioned three problems are given below.

- *Problem 1 (Implicit Aspects – Unconstraint Matching):* In this problem, we assume the aspects (i.e. required skills) of each paper are implicitly represented in the abstract of the paper and the skills of each reviewer can be inferred from the expertise document of that specific reviewer. Generally, the expertise document of an expert can be his/her resume but in this paper, we consider the concatenation of one's publications as his/here expertise document.
  In this problem, while each paper should be assigned to a group of k reviewers such that the skill coverage and confidence of the assigned group be maximal, there is no limitation on the capacity of reviewers (i.e. arbitrary number of papers can be assigned to a reviewer).
- *Problem 2 (Explicit Aspects – Constraint Matching):* In this problem, we assume that the set of the required skills of each paper and also the set of the relevant skills of each reviewer are explicitly determined (for example by using the ACM Categories and Subject Descriptors[1] keywords for computer science related papers). Given a limited number of reviewers, each paper should be assigned to a group of k reviewers such that: firstly, in an ideal matching, all aspects of *all*papers should be covered by the skills of the assigned groups (i.e. maximal coverage). Secondly, in an ideal matching, each member of the assigned groups for a paper should be able to cover as many as possible required skills of that specific paper (i.e. maximal confidence) and finally each reviewer should only be involved in review process of a limited (predefined) number of papers (i.e. load balancing).

[1] http://www.acm.org/about/class/2012.