Knowledge-Based Systems 75 (2015) 161–172

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

Collaborator recommendation in interdisciplinary computer science using degrees of collaborative forces, temporal evolution of research interest, and comparative seniority status

Paweena Chaiwanarom *, Chidchanok Lursinsap

Advanced Virtual and Intelligent Computing Research Center, Department of Mathematics and Computer Science, Faculty of Science, Chulalongkorn University, 254 Phayathai Road, Pathumwan, Bangkok 10330, Thailand

ARTICLE INFO

Article history:
Received 4 June 2014
Received in revised form 17 October 2014
Accepted 27 November 2014
Available online 4 December 2014

Keywords:
Social network
Research collaboration
Collaborator recommendation
Knowledge discovery
Co-authorship network

ABSTRACT

Currently, the research in computer science has been exponentially expanded beyond its own fields into the other research fields such as medical science, business, and social science in forms of collaborative researches. This collaborative researches stimulate a new recommending algorithm for determining a potential research collaborator under the interdisciplinary environment. Unlike other research fields, the research problems in computer science can be transformed to other known and solvable problems. In this paper, a new hybrid algorithm based on dynamic collaboration over time was proposed for recommending an appropriate collaborator. Besides considering only three basic factors concerning social proximity, friendship, and complementarity skill as employed by others', three new additional factors related to research interest, up-to-date publication data, and seniority of researcher are involved in our analysis. A set of new measures for all six recommending factors were proposed. The experiments were conducted with real bibliographic data within six continuous years of publication and over six topics in computer science. Our results were significantly higher than the results of the other methods at 90% confidence level.

1. Introduction

Currently, the research in computer science has been exponentially expanded beyond its own fields into the other research fields such as medical science, business, and social science in forms of collaborative researches. The algorithms or techniques in computer science are rather versatile and capable of finding solutions to various difficult problems in scientific, engineering, medical science, or even social science. Furthermore, the concerned research problems in computer science can be transformed to another known and solvable problem. The solution to this known problem can be adapted to solve the concerned research problems within the same time and space complexities. The collaboration level in computer science papers is rather moderate with respect to other scientific fields [1]. One question usually encountered by a team leader is how to select the appropriate collaborators in a new research. Finding a potential research collaborator in computer science is a challenging problem for the following reasons. Computer science research in the aspects of research problems and algorithms is rapidly and temporally changed during the past decade [2]. The publications in the field are very heterogeneous and mostly interdisciplinary to many other topics [2] such as software engineering and data mining reported by Bird [3]. This interdisciplinary trend has stimulated a higher degree of research collaboration among the researchers in various fields of computer science. Computer science comprises various sub-fields requiring specialization and characteristic features [2,4]. However, it is noticeable that there are some of researchers published their papers in several related fields outside their main research fields. This is due to the nature of studied problems in computer science which can be transformable to another problem. Intuitively, a set of famous researchers in a specific research topic seems to be the best choice to collaborate in research. Unfortunately, most of these potential collaborators are typically overloaded with their own research activities. Thus, the problem of finding the best collaborator should be transformed to the problem of finding the appropriate collaborator instead.

The following factors [5,6] have been suggested and used by many researchers for analyzing and recommending potential collaborators to an inquiring researcher.

* Corresponding author.
E-mail addresses: cc.paweena@gmail.com (P. Chaiwanarom), lchidcha@gmail.com (C. Lursinsap).

http://dx.doi.org/10.1016/j.knosys.2014.11.029
0950-7051/

1. *Social proximity* covering cohesive publication force and possibility of publishing new paper. This factor was considered in many researches by analyzing the structure of a co-authorship network [7–13].
2. *Friendship* covering cohesive co-authorship force based on the list of their common co-authors. The degree of friendship is measured in terms of distance or number of hops between researchers as in Lee's work [11] for more than one hop distance, Cohen's work [12], CollabSeer system[1] [13], and our previous work [10] for three and four hops, respectively.
3. *Complementarity skills* covering the research background similarity [10–17]. Researchers with more background research experiences in a required topic are more likely to be selected.
4. *Research interest* governed by the probability to work in inquired research topic in the near future. The factor concerning the change or trend research interest in the future has not been studied before. This factor actually occurs among computer scientists due the nature of computer science research.
5. *Up-to-date information* regarding the publication [6,10,18]. This factor refers to the year of publication in social proximity measure.
6. *Seniority status* of researcher [5,6,19]. Senior researchers have more potential to be recommended than junior researchers.

The existing techniques in several research studies did not cover the essence of the six factors. The main disadvantage is the fresh information for supporting dynamic collaboration over time has been lost for measuring. Lack of time stamp in social proximity and research trend lead to mistake in selecting potential research collaborators. An example of unsatisfied output of researcher recommendation is the result obtained from CollabSeer system. This system used only the name of seeker for querying, but an inquired research topic was omitted. It was based on the assumption that the research topics in the past and the future are the same. This assumption is too rigid and it may create a pitfall if the inquired researcher wants to change his topic or find his collaborators across different topics based on interdisciplinary environment.

In this paper, we proposed a new algorithm to increase the relevance and accuracy of recommending a collaborator in computer science field. Any researcher who wants to find some research collaborators is called *inquiring researcher*. Not only a given inquiring researcher's name, but also an inquired topic are used for querying. The output is a ranked list of potential collaborators relating to both researcher's name and the inquired topic. The overall contributions mainly focusing on the above six factors are the following:

- Our proposed algorithm is a hybrid method consisting of a structural approach based on co-authorship network with social proximity and friendship factors. Moreover, it also uses semantic approach based on the content of papers covering the factor of complementarity skill and the temporal research interest.
- Unlike the previous other studies, time evolution is attended in our algorithm for taking into account with the up-to-date information factor. The year of publication is gradually integrated in both structural approach and semantic approach.
- Our algorithm pays attention to seniority factor by studying the seniority relationship among researchers in computer science sub-fields. The possible seniority relationship between a co-author pair are: (1) senior researcher collaborating with senior researcher; (2) senior researcher collaborating with junior researcher; and (3) junior researcher collaborating with junior researcher. The output is the proportions of each seniority relationship which can be used as the parameter in algorithm.

However, there are some sub-fields in computer science having a few researchers when compared with other sub-fields, i.e. pure theoretical computer science. This makes the analysis rather difficult and inaccurate. Therefore, only the topics with a large number of researchers involved are concentrated, i.e. Bio-informatics, Data Mining, Hardware, Neural Network, Software, and Algorithm and Theory. Our study focused on the problem of defining a measure for each determination factor. The following questions are the main concerns.

1. How to measure the cohesive publication force between two researchers who may be friends or who may only know each other via the publications?
2. How to measure the possibility of publishing new papers of the potential collaborator?
3. How to measure the cohesive co-authorship force between two researchers?
4. How to measure the similarity of research backgrounds between two researchers?
5. How to measure the probability of working in the same domain of the inquiring researcher and the potential collaborator?
6. How to measure the research trend of the potential collaborator?
7. How to measure the potential collaboration between two researchers?
8. How to determine the best collaborator based on these seven measures?

The rest of paper is organized into the following sections. Section 2 describes the proposed methodology to compute the relevance between researcher pairs. Section 3 describes the experiments and performance evaluation. Section 4 concludes the paper.

2. Proposed scoring measures and algorithms

The proximity and friendship are key factors of structure approach in co-authorship network. The complementarity skill and research interest were taken into account in semantic approach. Some relevant definitions are defined as follows.

Definition 1. Co-authorship network $G = (V, E, L)$ is an undirect multigraph consisting of a set of all researchers' names, $V = \{v_1, \ldots, v_i, \ldots, v_n\}$, appearing in the considered publication database, their relationships in terms of edges, $E = \{(v_a, v_b) | v_a, v_b \in V\}$, and a set of attributes, $L = \{l_{(v_a,v_b)} | (v_a, v_b) \in E\}$, attached to each edge. Each $l_{(v_a,v_b)}$ consists of (1) paper ID, (2) year of publication, and (3) number of authors in each paper.

Definition 2. The degree of separation between v_a and v_b, denoted as d_{v_a,v_b}, is the minimum number of edges forming a path from v_a to v_b.

Definition 3. A neighbor of any v_j is a vertex v_a such that $(v_j, v_a) \in E$ and $1 \leqslant d_{v_j,v_a} \leqslant 6$.

Definition 4. Researcher v_a is a *friend* researcher of v_i if v_a is a neighbor of v_i and they co-authored some papers. Let Q_{v_i} denote a set of friends of v_i.

Definition 5. Researcher v_a is a *non-friend* researcher of v_i if v_a is a neighbor of v_i but they never co-authored any papers. Let $\overline{Q}_{v_i}$ denote a set of non-friend researchers of v_i.

[1] http://collabseer.ist.psu.edu/.