Information Sciences 466 (2018) 258–269

Contents lists available at ScienceDirect

Information Sciences

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

Identifying advisor-advisee relationships from co-author networks via a novel deep model

Zhongying Zhao[a,b,*], Wenqiang Liu[a], Yuhua Qian[c], Liqiang Nie[d], Yilong Yin[e,*], Yong Zhang[b]

[a] College of Computer Science and Engineering, Shandong Province Key Laboratory of Wisdom Mine Information Technology, Shandong University of Science and Technology, Qingdao, China
[b] Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China
[c] Institute of Big Data Science and Industry, Shanxi University, Taiyuan, Shanxi, China
[d] School of Computer Science and Technology, Shandong University, Qingdao, China
[e] Shandong Province Key Laboratory of Software Engineering, Shandong University, Jinan, China

ARTICLE INFO

Article history:
Received 24 September 2017
Revised 4 April 2018
Accepted 25 July 2018
Available online 27 July 2018

Keywords:
Social network analysis
Relationship mining
Co-author network
Advisor-advisee prediction

ABSTRACT

Advisor-advisee is one of the most important relationships in research publication networks. Identifying it can benefit many interesting applications, such as double-blind peer review, academic circle mining, and scientific community analysis. However, the advisor-advisee relationships are often hidden in research publication network and vary over time, thus are difficult to detect. In this paper, we present a time-aware Advisor-advisee Relationship Mining Model (tARMM) to better identify such relationships. It is a deep model equipped with improved Refresh Gate Recurrent Units (RGRU). Extensive experiments over real-world DBLP data have well verified the effectiveness of our proposed model.

1. Introduction

Social media (e.g., Twitter and Wechat) are gaining momentum in these years and become indispensable to most of us. Various social relationships are built in such social media, such as friendships on Facebook and trust relations on Epinions. It is well recognized that different types of social relationships are essentially affecting different aspects of people's life. For example, a Ph.D. candidate's research areas are largely influenced by his/her advisor [6], while his/her hobbies or behaviors are often affected by his/her families or friends to a high extent [7,18].

A research publication network comes into being in the process of research. It contains rich information of authors, paper titles, publication year, publication venue and etc, which imply abundant knowledge about advisor-advisee relationships. Identifying such relationships can benefit many significant applications [14]. With such relationships at hand, we can easily discover how researchers form different communities [5] [22] [25], how research topics emerge and evolve over time [9] [19] [24], and how a researcher influences the academic research community [8]. It is thus important and interesting to identify the advisor-advisee relationships from research publication networks.

To clearly describe the problem, Fig. 1 illustrates an example of the advisor-advisee relationship analysis over a research publication network. In the left part, a research publication network contains the information of authors, papers, and author-paper relationships. The middle part shows the preprocessing of the left one. The edge between Bob and Adi indicates that

* Corresponding author.
E-mail addresses: zyzhao@sdust.edu.cn, zzysuin@163.com (Z. Zhao), lwqiang@163.com (W. Liu), jinchengqyh@126.com (Y. Qian), nieliqiang@sdu.edu.cn (L. Nie), ylyin@sdu.edu.cn (Y. Yin), zhangyong@siat.ac.cn (Y. Zhang).

https://doi.org/10.1016/j.ins.2018.07.064
0020-0255/

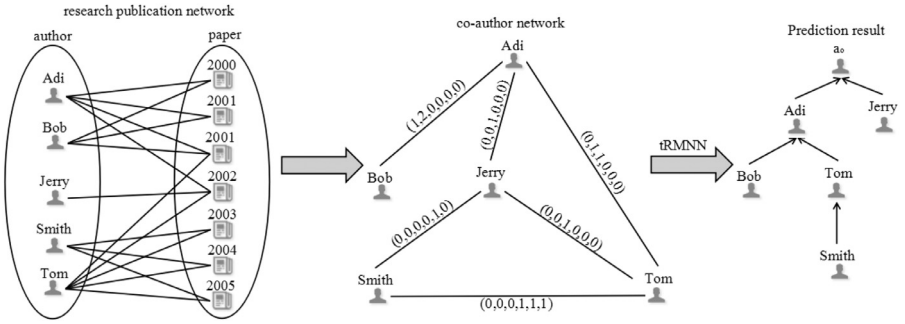

Fig. 1. Example of advising relationship analysis on the co-author network.

there is a co-author relationship between them. The vector (1,2,0,0,0,0) on this edge means that these two researchers have 1 co-authored paper in 2000, 2 co-authored papers in 2001, and have no collaboration since 2002. The right part shows the expected result by a visualized chronological hierarchy. The parent-child relation in the hierarchical structure is the advisor-advisee relationship. The advising path is from the root to the leaves.

However, in reality, identifying the advisor-advisee relationships faces some challenges. (1) Implicit. The advisor-advisee relationship is hidden in the research publication networks. The information that we can find from the network includes only authors, paper title, journal or conference name, and publication year. We have no direct or additional information or labels about such kind of relationship. (2) Time-dependent. The advisor-advisee relationship is highly time-dependent. When a postgraduate student joins his advisor's team, there is a strong advisor-advisee co-author relationship between them. However, after graduation, such relationship turns to be loosen gradually.

In this paper, we formulate the identification of the advisor-advisee relationship as a probabilistic ranking problem. An improved Refresh Gate Recurrent Unit (RGRU) is firstly presented. It contains only one gate unit, holding a simpler structure but performing better than Long Short-Term Memory (LSTM). Taking time dynamics into account, we propose a deep time-aware Advisor-advisee Relationship Mining Model(tARMM). The main contributions of this paper are summarized as follows.

- The RNN is reconstructed by devising a Refresh Gate Recurrent Unit (RGRU), inspired by the idea of variance Recurrent Neural Network (RNN) model like LSTM. RGRU is composed of only one gate unit, with a simpler structure but better performance than LSTM.
- Taking the time factor into account and combining RGRU and DNN together, we present tARMM to solve the problem of advisor-advisee relationship mining.
- In order to evaluate the performance of the proposed model, we conducted some extensive experiments on DBLP data. The experimental results show that tARMM achieves the best performance by comparing with several state-of-the-art models.

The remainder of the paper is organized as follows. Section 2 shows the related works. The problem definition is given in Section 3. Section 4 presents the time-aware Advisor-advisee Relationship Mining Model in detail. Extensive experimental results and discussions are presented in Section 5. Finally, Section 6 concludes this study and discusses the future work.

2. Related work

With the help of social media, people are more active than ever before with more diverse social activities, eg., making friends with each other, browsing and commenting others' posts, and forwarding interesting news [23]. Behind those social behaviors, there are many different types of social relationships and many research works have been done on the relationship mining from the social activities.

The relationships in social media are multi-dimensional since people always connect with each other in many different ways or for different reasons. To address this problem, Tang et al. [12] proposed a clustering based method to extract the latent social dimensions based on the network information. They then utilized them as features for discriminative learning which was capable of determining different social dimensions automatically. But they didn't study the semantic meanings related to the extracted dimensions. Wang et al. [17] proposed a novel collaborative filtering approach to predict the unobserved links in a network (or graph) by using the topological features. However, this approach was still not stable, and could not be extended to the large-scale networks effectively [26]. Factor graph was used to identify the social relationships in recent years. Tang et al. [13] proposed a partially labeled factor graph to predict the types of social relationships in large networks. They then developed a framework for classifying the types of social relationships by learning across heterogeneous networks [10]. Zhuang et al. [27] precisely defined the problem of inferring social ties and proposed a Partially Labeled Pairwise Factor Graph Model (PLP-FGM) to infer the types of social relationships. Tang et al. [11] developed a framework called TranFG to classify the types of social relationships by learning from the heterogeneous networks. The framework incorporated social theories into a factor graph model and thus effectively improved the accuracy of predicting the types of