Neurocomputing ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Neurocomputing

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

An estimation model for social relationship strength based on users' profiles, co-occurrence and interaction activities

Liyan Xiong [a], Yin Lei [a,*], Weichun Huang [b], Xiaohui Huang [a], Maosheng Zhong [a]

[a] School of Information Engineering, East China Jiaotong University, Nanchang 330013, China
[b] School of Software Engineering, East China Jiaotong University, Nanchang 330013, China

ARTICLE INFO

Article history:
Received 7 September 2015
Received in revised form
26 May 2016
Accepted 5 July 2016
Communicated by Jitao Sang

Keywords:
Social network
Social relationship strength
User's profiles
Co-occurrence of user names
Interaction activities

ABSTRACT

As the growing popularity of the Internet, the online social network has become an indispensable part and played a more and more important role in our daily life. People in the online social networks would link with their friends through chatting, e-mail, posting, commenting directly, which are built on the mature relationships in reality. However, this link-ship provides only a coarse representation of relationship rather than revealing the relationship strengths. Accurate measurement of social relationship could be applied to many applications, such as project or paper review. In this paper, we propose a probabilistic graphical model to measure the relationship strengths between different users in the social network by taking consideration of the similarities among users profiles, co-occurrence of user names and interaction activities in different activity fields.

1. Introduction

The social network service (SNS) has become a necessary contact way for humans to expression, communication and sharing information with their family, friends and colleagues, by messaging, publishing newsfeeds, uploading photos and posting comments. The current relationships of network users in Facebook, Twitter, Sina Weibo, are mostly simple binary status, such as Follow, Fans, Friends, which ignores their relationship strengths between different users, furthermore, much of the past work on social network relationship has focused on coarse relational ties (e.g. friends or not). However, people obviously prefer contacting with their friends to just acquaintances and more inclined to exchange emotion with their close friends. It seems more reasonable to use relationship strength to evaluate the closeness than just simple relationship status. Therefore, we will focus on designing an effective estimation method to measure the relationship strength accurately.

The existing works of estimating the relationship strengths consider only two cues: the users profiles and the interaction activities among different users. The users profiles including their age, location, hometown, hobbies, work experience etc, provide the basic attributes to help exhibit the similarities between users. It is believed that the higher similarity of users profiles is, the higher relationship strength will be. For instance, if the geographical positions between users are closer, or the users share more interests and friends, their relationships will be stronger. From the activity aspect, frequently interactions, such as sending messages, comment on friends posting, @ other users, could reveal their closeness in daily life, thus can be used to measure the relationship strengths between different users.

Most approaches of relationship strength estimation confused all the interaction activities together and didnt take into account of the fact that the relationship strengths between the same users will vary significantly in various activity fields. For example, user A has a frequent contact with his partner B in work time, that means, A and B have stronger relationship strength in work field compared to the scarce contact in entertainment or diet fields. Therefore, measuring the relationship strength on each specific activity field seems more reasonable and useful in real applications, rather than the overall relationship strengths between different users.

Inspired by this motivation, we proposed a general framework to measure the relationship strengths between different users on various activity fields, based on three information sources: similarities of users profiles, co-occurrence of user names and interaction activities. After data processing, our approach is composed of two primary sequential steps: (1) Classify the activity documents into their corresponding activity fields by a new classify algorithm SVM-KNN. The SVM-KNN is the mix-algorithm of Support Vector Machine and K Nearest Neighbor, which improves the accuracy of classification. (2) Measuring the relationship strengths

* Corresponding author.
E-mail address: rayin1992@163.com (Y. Lei).

http://dx.doi.org/10.1016/j.neucom.2016.07.022
0925-2312/

between different users on different fields. In this step, we proposed a probabilistic graphical model by taking consideration of the similarities of users profiles, co-occurrence of user names and interaction activities in different activity fields. By maximizing the joint probability of the data distribution, we obtain the single relationship strengths on each different field, and then calculate the overall relationship strength by the weighted sum of each ones.We adopt the normalized Discounted Cumulative Gain and Receiver Operating Characteristic to evaluate the performance of this work, and achieved a promising result compared with the previous typical methods.

To summarize, this work contributes on the follow aspects:

- We differentiate the relationship strength values on different fields of the same two users, and firstly use SVM-KNN to classify the users activities into corresponding fields.
- Distinguished from the existing methods which only consider users profiles and interaction activities, we add the co-occurrence of user names to build the probabilistic graphical model. And the experiment results prove that our new added information source significantly improves the accuracy of estimation.
- To verify the performance of our approach, we compare the result with two previous methods. One method only evaluates the overall relationship strength which neglects the activity fields, and the other one combines the users profiles with the interaction activities to measure the relationship strengths on different fields.

The rest part of this paper is organized as follows: Section 2 reviews the related works and Section 3 briefly generalizes the framework of our estimating model and detailed expresses the processing steps before measuring. The details of the proposed approach are elaborated in Section 4 and the experiment and result analysis are showed in Section 5 with the dataset from Sina Weibo. Finally, we conclude the paper and discuss the directions of the future works in Section 6.

2. Related work

Weak relationship was firstly proposed by Granovetter Mark [1] in 1974. Four indicators: interaction time, emotion intensity, intimacy and mutual benefit are used to divide the relationship into strong or weak ties, and he realized that weak ties are more valuable in information dissemination than strong ties. From then on, many researches focused on distinguishing the strong relationships from weak ones based on one or two information. For example Gilbertet al. [2] presented a predictive model for analyzing the relationship strengths together with their evolution based on various user activities. Zhang et al. [3] defined effective relationship strengths (ERS) to distinguish link importance by utilizing node activity, node attraction and link frequency. And Niladri and Sanasam [4] studied the effect of tie weight on link prediction and perform an analysis of the weak tie theory. Ricard [5] presented a BFF tool that could classify the friend types in communities and assign the relationship strength values to each other.

In addition to the above binary predicting methods, several achievements of representing the full spectrum of relationship strengths were made in succession. Horowitzand and Choudhury [6,7] traced the relationship on the basis of regular e-mail, instant message and personal profiles. Ehsan Khadangi et al. [8] utilized the profiles and interaction activities and Ye et al. [9] used the characteristics distribution of relations and interactions to present a method for quantifying links' tie strength. Especially, a latent variable model developed by Xiang et al. [10] estimated the relationship strengths based on user's profiles and interaction activities, along with a coordinate ascent optimization procedure for the inference.

However, these former works mixed all the activity fields together rather than considered that the relationship strength of a pair of users may be different in various activity fields. After that, Zhao et al. [11] proposed an improved method to estimate the different relationship strengths on different activity fields. They used the LDA clustering algorithm to cluster all the uses' activity documents, and then built a graphical model to estimate their relationship strengths on the basis of users' profiles and interaction activities. There exists another attempts focusing on different topics to investigate the strength of influence, which emphasized that the inter-user influence is essentially topic-sensitive. For example, Tang et al. [12] introduced topic-based social influence analysis and proposed a Topical Affinity Propagation (TAP) approach to describe the problem using a graphical probabilistic model. Liu et al. [13] moved one step further and not only formulated a probabilistic model to mine direct influence between nodes but also presented a topic-level influence propagation method to mine indirect and global influence. The experimental results showed that the learned influence model can greatly improve the accuracy of user's behavior prediction. Different from the former two works, Sang et al. [14] investigated the topic-sensitive influence in the multimedia context. Besides the explicit influence link and the textual annotation, they provided a multimodal framework simultaneously modeling visual content.

Ma Chao et al. [15] proposed a hierarchical entropy-based relationship measurement approach (HERMA) at the help of the tracking and positioning function of smart phone to analyze the user's activity trajectory. Even though the direction of this paper is off course with our research, it strongly demonstrated that the interaction frequency, the geography proximity and the duration of common position have the direct impact on relationship strength.

3. The system framework

As mentioned above, due to the distribution of users relationship strengths on different activity fields are different, we must firstly classify the users activity fields, and calculate the relationship strengths of different users in every single field on the basis of the similarity of users profiles, co-occurrence of user names and interaction activities, respectively. The estimation model will be described in detail in Section 4. Ultimately, the overall relationship strength is calculated from the weighted sum of all single relationship strengths on each activity field. The entire processing flow is as shown in Fig. 1, and the processing steps before estimating the relationship strengths are generally expressed as following.

3.1. Data acquirement and processing

The Open API of SinaWeibo [16] is utilized to obtain users' information data in a period time, and the downloaded data is processed by word correction, stop word removing, short text processing, feature extending etc. After that, we acquire the profiles and interaction activity documents of a user set $U = \{u_1, u_2, \ldots, u_k\}$, in which the user's profile $f_p = \{f_{p1}, f_{p2}, \ldots, f_{pn}\}$ is the essential attribute of user, composed of the location, personal description, job experience, education etc. The interaction activity documents $D = \{d_1, d_2, \ldots, d_n\}$ include newsfeed, received or post comment, forward or forwarded content etc. Each document is referred to one or several users. In our experiment, we will combine the newsfeed with its all comments and forwarding contents to a session $S = \{(d_1, d_3, \ldots, d_{n-3}), (d_2, d_5, \ldots, d_{n-2}), \ldots, (d_4, d_{11}, \ldots, d_n)\}$ and analyze the interaction strength on the basis of a session.

Please cite this article as: L. Xiong, et al., An estimation model for social relationship strength based on users' profiles, co-occurrence and interaction activities, Neurocomputing (2016), http://dx.doi.org/10.1016/j.neucom.2016.07.022