Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 129 (2018) 123–127

Procedia Computer Science

www.elsevier.com/locate/procedia

2017 International Conference on Identification, Information and Knowledge in the Internet of Things

A Dynamic Human Contacts Prediction Method in Mobile Social Networks

Baoling Wu[a], Feng Zeng[a,*], Wenjia Li[b]

[a]School of Software, Central South University, Changsha 410083, China
[b]Department of Computer Science, New York Institute of Technology, New York, NY 10023, USA

Abstract

Human contacts prediction is a challenging task in mobile social networks. In this paper, we extract some new features to predict human contacts and propose a new human contacts prediction method which is suitable for dynamic networks. The method first trains a classifier for each time period and assigns different weights to the classifiers of each time period. Then, an ensemble result of all the classifiers is used to predict human contacts. The experimental results show that our proposed method is efficient.

Selection and peer-review under responsibility of the scientific committee of the 2017 International Conference on Identification, Information and Knowledge in the Internet of Things (IIKI2017).

Keywords: Mobile Social Networks; Human Contacts Prediction; Features Extraction; Dynamic Networks.

1. Introduction

A Mobile social network is a dynamic network system, where persons and contacts between people in the networks emerge and disappear over time[1]. Human contacts prediction is an important research direction in mobile social networks, it can be used to help companies find potential partners or give assistance to the police for detecting terrorists. But, human contacts prediction is also a changing task. The first challenge is that we can only get a small amount of information about human contacts, and the second challenge is that the existing human contacts prediction methods are based on the static network structure, and directly applying these methods to dynamic network prediction is bound to reduce the prediction accuracy.

In the early stage of human contacts prediction study, many scholars have proposed several different methods of human contacts prediction. These methods provide us with some important features that is used in the human contacts prediction task. The early work of feature extraction makes the prediction accuracy of human contacts improved remarkably, but the performance of these researches are still not ideal and most of researchers do not consider the

* Corresponding author. Tel.: 13397645839.
E-mail address: fengzeng@csu.edu.cn

Selection and peer-review under responsibility of the scientific committee of the 2017 International Conference on Identification, Information and Knowledge in the Internet of Things (IIKI2017).
10.1016/j.procs.2018.03.058

issue that the structure of mobile social network is changing over time. In recent years, some link prediction methods based on time-aware method have achieved satisfied results in terms of predicting accuracy[2]. Motivated by the above research results, we propose a novel human contacts prediction method which is applicable to mobile social network.

In this paper, we regard human contacts prediction as a problem of graph reasoning, and view one person and a contact between people as a node and an edge, respectively. We extract some new features to predict human contacts and propose a dynamic human contacts prediction method based on naive bayes algorithm[3], named DHCP. DHCP predicts the contact of certain node pairs based on the historical contact records of all node pairs during all time periods in the past, rather than the contact records of all node pairs during a certain time period in the past.

2. Problem Definition

In this paper, we deal with the problem of human contacts prediction as a graph reasoning issue. We divide the experiment time into k_{max} time periods of τ seconds. The k^{th} time period can be denoted as $T_k = [t_0 + (k-1)\tau, t_0 + k\tau]$, where $1 \leq k \leq k_{max}$, t_0 is the start time of the experiment, and τ is the time interval. In the human contacts graph $G_k = (N^k, E^k)$, N^k is composed of internal nodes and external nodes during the k^{th} time period. N^k_{int} denotes a group of persons who are equipped with sensors called internal nodes. N^k_{ext} denotes the group of persons who are not equipped with sensors called external nodes. We denote the set of contacts between persons with $E^k = E^k_{konwn} \cup E^k_{unkonwn}$, and record repeat contacts between two people during T_k as multiple edges. The edges in $E^k_{konwn} \subset N_{int} \times (N_{int} \cup N_{ext})$ are known while edges in $E^k_{unkonwn} \subset N_{ext} \times N_{ext}$ are missing.

The human contacts prediction task is to predict the edges in $E_{unkonwn}$ exists or not at some time point in the future according to the historical contact information of all node pairs. The contact label of node u and v during T_k can be represented as $label^k_{(u,v)}$. If node u and v have a contact during T_k, $label^k_{(u,v)}$ is equal to 1, otherwise it is equal to 0.

3. Features extraction

3.1. Number of Common Neighbors(NCN)

The number of common neighbors is an important feature used in link prediction, so we choose the number of common neighbors as the first feature of human contacts prediction task. Let $|Ne^k_u|$ represents the degree of node u during T_k, the number of common neighbors between node u and v in T_k is defined as

$$NCN^k_{(u,v)} = |Ne^k_u \bigcap Ne^k_v|. \tag{1}$$

The number of common static neighbors derived from common neighbors and Adamic Adar index (AA)[4], Jaccards Coefficient (JC) [5] and Resource Allocation index (RA) [6] indicators of common neighbors have made great contribution to the link prediction task. And the product of the degrees of node pairs is also an important feature. In the experiments of this paper, we not only use the extracted new features to predict human contacts, but also add these five important features used in the link prediction.

3.2. Contact Time Ratio of Common Neighbors (CTRCN)

By further analysis of our friends network, we find the fact that if two people who do not know each other have a common friend and the two people have frequent contacts with the common friends, the two people are likely to contact in the future. On the contrary, if the contacts between the common friends and the two people are both not frequent, even if the two people have a common friend, they may not know each other. So, we choose contact time ratio of common neighbors for node u as the second feature of our prediction work, the third feature is contact time ratio of common neighbors for node v. The contact time ratio of common neighbors for node u is defined as

$$CTRCN^k_u = CT^k_{(u,CN)} / CT^k_{(u,Ne^k_u)}, \tag{2}$$