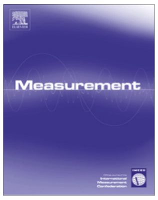
Measurement 56 (2014) 231–238

Contents lists available at ScienceDirect

Measurement

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



Supervised link prediction in symptom networks with evolving case

Buket Kaya [*], Mustafa Poyraz

Department of Electrical and Electronics Engineering, Fırat University, Elazığ, Turkey

ARTICLE INFO

Article history:
Received 24 May 2014
Received in revised form 26 June 2014
Accepted 17 July 2014
Available online 24 July 2014

Keywords:
Link prediction
Social networks
Future symptoms
Medical informatics

ABSTRACT

Medical care can improve the life quality since a patient can modify his habits and lifestyle in order to prevent the occurrence of probable correlated future symptoms causing to a disease. In this paper, we predict the onset of future symptoms on the base of the current health status of patients. The problem of predicting the relations between symptoms (abnormal parameters in this paper) which can be shown as the reason of a disease in the future is a really difficult and, at the same time, an important task. For this purpose, the present paper first constructs a weighted medical data network considering the relations between abnormal parameters. Then, we propose a link prediction method to identify the connections between parameters, building the evolving structure of medical data network with respect to patients' ages. To the best of our knowledge, this is the first attempt in predicting the connections between the results of laboratory tests. Experiments on a real network demonstrate that the proposed approach can reveal new abnormal parameter correlations accurately and perform well at capturing future disease risks.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

Social networks are social structures composed by entities and relationships among them. In this structure, new connections and entities appear or disappear overtime. This case makes them highly dynamic and complex systems. Analysis of social network includes broad field of research concerning method and strategies for the study of social networks [1]. Link prediction is an important task treated by social network analysis. This task is concerned to the problem of predicting the existence of a connection between two entities and the interactions already present in the network. In this meaning, link prediction can explain the reasons of the evolution in a social network.

So far, many approaches have been proposed in order to handle link prediction task. All of these methods are based on the measures indicating the proximity between nodes. These measures proposed in the literature are generally categorized into semantic or topological/structural measures [2]. In semantic measures, the content of the nodes is taken into account to evaluate proximity. For example, the similarity between topics discovered from titles can be used to predict future connections among the authors in a co-authorship network. As apart from the semantic ones, the topological measures consider the network structure to obtain the proximity values. Topological measures are more widely adopted because they are more general and do not require detailed information related to the content of node. Moreover, this content may not be present for every time at the social network considered. Topological measures are categorized into neighborhood-based or path-based measures [3]. The neighborhood-based measures consider the immediate neighbors of the nodes. According to this measure, if the neighbors of two nodes have a large overlap, they will probably form a link [2].

---

[*] Corresponding author.
E-mail address: bkaya@firat.edu.tr (B. Kaya).

http://dx.doi.org/10.1016/j.measurement.2014.07.008
0263-2241/© 2014 Elsevier Ltd. All rights reserved.

The most widely used measures among them are Common Neighbors [4], Preferential Attachment [4,5], Adamic-Adar Index [6] and Jaccard's Coefficient [7]. The path-based measures generate a proximity value considering the paths between the related nodes. According to this idea, it is probable that two nodes will form a link if short paths are present in between the related nodes [8].

Many studies have been done to overcome the link prediction problem [9–11]. However, in order to predict new relations at future time, many of studies are established on the applications of proximity measures to non-connected pairs of nodes in the network at present time. Proximity measures showing the similarity between pairs of nodes can be utilized either by unsupervised [9,10,12], or supervised link prediction [11,13,14]. In unsupervised methods, a proximity measure is selected and utilized to rank node pairs in the network. The node pairs having good score are predicted to be linked. In the supervised approach, the link prediction problem is handled as a classification task and attributes of network are used as predictor attributes by a classification method. A classifier employs these attributes to implement a binary classification to find whether the link will be occurred or not in the future. In all of the above mentioned studies, the proximity values are calculated without considering the evolution of the network. In other words, the proximity measures are calculated using all network data up to the current network state without taking into account when links are created. Therefore, hidden and potentially valuable source of information is not adequately considered for link prediction.

There are specific cases occurred between a pair of nodes in sequential time intervals. These cases deal with creation or removal of link between nodes. For instance, a new link can be formed between two nodes in the next interval, while they are not connected in a specific time interval. This situation represents the evolution of the related social network. Some studies, recently, have been done concerning the use of time intervals for link prediction. Tylenda et al. [15] proposed a network model in which link weight was showing the age of the most recent specific cases between the related nodes. Brigmann et al. [16] proposed a network with temporal information for discovering association rules explaining the network evolution. The other approaches proposed for link prediction handles the task as a time series forecasting problem [17–19]. In these studies, the time series models were utilized to improve prediction correctness of proximity values.

1.1. Motivation and contribution

For many different areas, link prediction can be used to expect future behavior or to recognize likely relationships that are hard or expensive to understand directly [20–22]. One of these areas is related to medical care research area. Medical care area needs to become more proactive than reactive in recognizing the onset of disease and risk [23,24]. Currently, physicians use laboratory results to further determine the patient's stage of health. However, there is some disadvantages such as generally to focus on only a few medical parameters (symptoms) and to linked by a particular doctor's experience, memory, and time. Therefore, current medical care is not proactive and is not enough treating or eliminating a disease at the earliest signs.

As a remedy to the above mentioned problems, in this paper, we propose a predictor to determine the risk of individuals to develop disease, and to undertake the correct actions at the earliest signs of illness. To this purpose, we first construct a weighted medical data network which indicates the relationships between abnormal parameters of disease. In this network, each node represents the abnormal parameters of patients; the edges connect these parameters appearing in the same patient. Then, we propose a link prediction method based on the evolution of the constructed medical data network in order to identify the relations between all the parameters which can cause any disease, gathering the results obtained at several laboratories. Finally, we test the proposed method on the medical data network constructed with laboratory results of patients more than 210,000. The experimental results demonstrate the accuracy of our method on supervised prediction and encourage us for further analysis.

The rest of the paper is organized as follows. Section 2 illustrates the construction of medical data network. Section 3 presents our link prediction method, proposing evolving structure of medical data network and evolving cases in this network. The conducted experiments and the results achieved for the medical data network are discussed in Section 4. Conclusions and future research directions are included in Section 5.

2. The construction of medical data network

In this paper, medical data contains the results of the laboratory tests. In these data, there are many medical parameters (symptoms) tested for a patient. If all the values of these parameters are within reference interval, it can be said that the person coming to the hospital does not have a disease. However if one of them or several parameters are not within reference interval, in other words, it exceeds the maximum value of the interval or it is lower than the minimum value of the interval, then this parameter exhibits an abnormal case and it is probable that the person having this laboratory result has a disease.

In the medical data network, nodes are the abnormal parameters which are not within reference interval. If the value of the parameter X is larger than the maximum value of reference interval then, this node is represented by $X.H$, where H stands for High. In contrast, if the value of the parameter X is lower than the minimum value of the reference interval then it is represented by $X.L$, where L stands for Low. If the parameter X has a categorical or binary value, the unexpected value is represented by $X.A$ where A stands for Abnormal. In the constructed medical data network, edges connect these pairs of abnormal parameters appearing in the same patient. Thus, a relationship between two abnormal parameters exists whenever they appear simultaneously in a patient. In this paper, we construct a medical data network and exploit the connections

Download English Version:

https://daneshyari.com/en/article/730200

Download Persian Version:

https://daneshyari.com/article/730200

Daneshyari.com