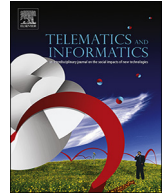


Contents lists available at [ScienceDirect](#)

Telematics and Informatics

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

Evaluating reliability of question-disease relations in online health forms: A link prediction approach

Buket Kaya^{a,*}, Esra Gündoğan^b^a Department of Electronics and Automation, Firat University, Elazığ, Turkey^b Department of Computer Engineering, Firat University, Elazığ, Turkey

ARTICLE INFO

Keywords:

Online health forums
Question-disease relations
Link prediction
Social network analysis

ABSTRACT

The Internet has become an indispensable part of human life in today. People can now easily find answers to questions they are curious about via the internet. The short, effortless and free way that the Internet provides is extremely attractive for people to have an idea in subjects they wonder related to their health. There are many online health forums where people can ask questions answered by health professionals. Every day, people ask thousands of questions on these sites and get answers about which diseases their complaints may be related to. The frequent use of online forum sites by people has led to the selection of these forums as data source for this study, and analysis of reliability. Firstly, in this study, link prediction in bipartite social networks, where intensive works have been done and it is applied on many areas nowadays, is tried to be carried out on question-disease bipartite network constructed with data obtained from analysis of online health forums whose use rate increase substantially. For this purpose, a novel link prediction method called as intensive link prediction is proposed, and prediction success of this method is compared with five of similarity-based link prediction methods. Better results have been obtained with the proposed method than the other methods. Then, the accuracy of the answers given to the users on online health forums which received intense interest are tested. The reliability of online health forums is measured by the accuracy analysis performed.

1. Introduction

Nowadays, the Internet is commonly used in the field of health as it is in every area. People can now easily find answers to questions they are curious about via the Internet. The Internet, through the short, effortless and free way it provides, has become a reference source for people to have ideas about their health problems. People use the Internet for many purposes in health care. They actively employ it to obtain an idea about which disease may be associated with their symptoms when people have certain disease symptoms. There are many online health forum sites where patients and doctors communicate on the Internet. Before people go to the doctor and do some tests, they write their complaints on online forum sites and get information from health professionals. Because of this tendency, forum sites have thousands of questions and answers. The frequent use of online health forums by people has led to the selection of these forums as data source for this study, and analysis of reliability.

We live together with many complex systems in our environment. The Internet, nervous system, protein networks, transport networks are some examples of complex networks we encounter. Complex networks are a structure used to model the complex systems, in which links are relationships, and nodes are persons, objects, etc. Complex networks make it easy to analyze the structure

* Corresponding author.

E-mail address: bkaya@firat.edu.tr (B. Kaya).

<https://doi.org/10.1016/j.tele.2018.05.009>

Received 21 February 2018; Received in revised form 24 April 2018; Accepted 22 May 2018

0736-5853/ © 2018 Elsevier Ltd. All rights reserved.

of complex systems, their development, and the relationships between the entities they represent. Therefore, the analysis of complex networks has become an important research area in many sciences.

Social network analysis is used in many areas for a variety of purposes. Link prediction among entities in the social network (Al Hasan and Zaki, 2011; Lü and Zhou, 2011), recommendation systems (Li and Chen, 2009), terrorist network analysis (Anil et al., 2015), and community discovery (Xu et al., 2013) are some of done studies on social network analysis. Analysis of enormous quantities of data on social networks provides good, useful information about individuals, communities, organizations (Wasserman and Faust, 1994). Valuable information can be discovered from social networks that will enable a better understanding of some situations and relations and allow improvements to be made in the current situation. Therefore, researchers from different fields have shown interest in social networks. But analyzing social networks is not an easy task. One of the difficulties in social network analysis is the dynamic nature of social networks. Over time, new nodes and links may be added to the network, or existing nodes and links may be lost. One of the most important social network analysis topics used to investigate dynamic structure of the network is link prediction. It estimates potential future connections by using the current state of the network (Liben-Nowell and Kleinberg, 2007).

Link prediction has important applications in many fields. It can be used for purposes such as finding co-authors for academic social networks (Pavlov and Ichise, 2007; Wohlfarth and Ichise, 2008), finding protein-protein interactions, exploring relationships that can occur between private persons in a network that shows the interactions of terrorists (Krebs, 2002), suggesting new friends for people (Aiello et al., 2012), recommending products to customers in online shopping (Akcora et al., 2011), realizing optimal orientation in sensor networks (Jia et al., 2013), finding relationships between diseases (Gül et al., 2016; Kaya and Poyraz, 2014; Kaya and Poyraz, 2015), estimating future citations (Jawed et al., 2015).

Many complex systems in our environment are bipartite network structure. The modeling of many complex systems in a bipartite network structure has increased the orientation to apply link prediction to such networks. Link prediction in networks modeled as bipartite has become an issue where intense work has been done so far. Different methods have been tried for link prediction in bipartite networks. Benchettara et al. (2010) have identified new topological features that can reflect the probability of a connection between two nodes and have used these features in a supervised machine learning approach to link prediction in bipartite networks. Xia et al. (2012) have made a study based on structural holes using weak ties between nodes. Nigam and Chawla (2016) have used a semi-bipartite network model to predict links in a bipartite network consisting of relations between users and health topics that interest them. Thus, it has been tried to be suggested health topics that will interest people by utilizing common features such as demographic characteristics, reading habits of people. Chang and Kao (2012) have formed a bipartite editor-article network with data obtained from Wikipedia. Some features have been extracted from the bipartite network created and these features were used in supervised machine learning algorithm. With this approach, link prediction has been made to suggest articles that authors can edit. Allali et al. (2011) have described a special type of link, called internal links, in bipartite networks. Link prediction has been made based on these links. Gao et al. (2017) have performed a projection-based link prediction. Link prediction is made based on assumption that the greater the sum of weights of patterns covered by a pair of candidate nodes, the higher the probability of a future connection between this candidate node pair. The concept of candidate node pair presented in the study is similar to the concept of internal links defined by Allali et al. (2011). However, there are some significant differences in this study. Computation time of this method has been reduced with applied changes and it has achieved higher speed.

In our previous study, we performed the prediction of symptom-disease links in online health forums (Gündoğan et al., 2017). In this study, we improve that method by considering question-disease links on a larger dataset. Also, as the second goal of this paper, we test the reliability of the information obtained from online health forums. For these goals, this study was carried out in two steps.

In first step,

- The questions on online health forums (<https://www.icliniq.com/tr/>, <https://www.drugs.com>) were analyzed and user complaints were extracted.
- Symptom information for each disease was obtained from <https://www.drugs.com/>.
- The symptoms of the diseases and the symptoms expressed in the questions were compared, and found which diseases the questions might be related to.
- A bipartite network consisting of diseases and questions was constructed.
- The intensive link prediction method was applied to bipartite network.
- Five of similarity-based link prediction methods were applied to the network for performance comparison.
- Six different methods were compared in terms of precision, recall and F-measure.
- It was seen that the proposed method has better performance and higher accuracy prediction rate.
- Also, real life data obtained by analyzing online health forums where are thousands of questions every day was used for link prediction.
- In second step,
- A disease-symptom bipartite network was constructed using the disease-symptom dataset created with data extracted from <https://www.drugs.com/>.
- User complaints in each question were identified.
- By analyzing the answers given by the doctors to every question analyzed, the diseases that the doctor thought the user had, were determined.
- Diseases corresponding to the symptoms of each question analyzed were found using constructed disease-symptom bipartite network.
- For each question, the diseases expressed by the doctor and the diseases found using the actual network were compared.

Download English Version:

<https://daneshyari.com/en/article/10225828>

Download Persian Version:

<https://daneshyari.com/article/10225828>

[Daneshyari.com](https://daneshyari.com)