Fuzzy Inf. Eng. (2016) 8: 113–126

Available online at www.sciencedirect.com
ScienceDirect
Fuzzy Information and Engineering
http://www.elsevier.com/locate/fiae

ORIGINAL ARTICLE

Performance Evaluation of Link Prediction Techniques Based on Fuzzy Soft Set and Markov Model

Y. Bhawsar · G.S. Thakur

Received: 21 January 2015/ Revised: 6 January 2016/
Accepted: 11 February 2016/

Abstract Link prediction in social networks represents a significant task in understanding the behavior and actions of users. There are various methods to it and some of them are Jaccard's Coefficient, Common Neighbor, and Sorenson etc. These methods predict the link correctly but sacrifice with efficiency. The reasons behind this are discussed in this paper with two new methods of link prediction to improve the efficiency. These methods are based on fuzzy soft set and Markov model. We analyze that the proposed work predicts more accurately links as compared to existing methods.

Keywords Social network · Link prediction · Fuzzy soft sets · Markov model

1. Introduction

Social media is continuously gaining attention by users and so by researchers. There are huge data associated with the social sites. Applying data mining in social media

Y. Bhawsar (✉)
Department of Computer Science, Government Women's Polytechnic College, Bhopal-462016, Madhya Pradesh, India
email: yachanabhawsar@gmail.com
G.S. Thakur
Department of Computer Application, Maulana Azad National Institute of Technology, Bhopal-462003, Madhya Pradesh, India
Peer review under responsibility of Fuzzy Information and Engineering Branch of the Operations Research Society of China.

http://dx.doi.org/10.1016/j.fiae.2016.03.007

data is an open research area. It includes link classification, link prediction, influence mining, viral marketing, and opinion mining. The problem of link prediction has been solved by defining proximity-based measures on the nodes in the underlying network [6, 15, 20], usefulness of different topological features [22], statistical relational models [1, 3, 13, 19, 21] and context of the classification problem [4, 11, 17]. These existing methods have some limitations. Methods which are based on topological features are not valid for new nodes. Classifier approaches fail in reliability of training data. Markov model is basically used for users' behaviour prediction on the web. This concept is applied to predict the relationship on social networks. The interest in the analysis of users' behaviour on the web has been increasing in recent years. Link prediction also utilizes the users' behaviour of using/generating different events such as message, comments, and follow, etc. It becomes easy to predict links with the help of the patterns of events followed by users for generating friendship event. Fuzzy soft sets and rough sets are used for dealing with uncertainty in many applications [16]. Fuzzy set theory was developed by Zadeh in 1965. Molodstov developed soft set theory in 1999. Fuzzy soft set is applied for community detection problem of social network analysis by many researchers. Fuzzy community detection algorithms quantify the strength of association between all pairs of nodes and communities. Uncertain data exist in a social network environment. Classical mathematics and traditional link prediction techniques are not sufficient to model the problem of uncertainty in social network data. Soft sets and fuzzy soft sets are the tools to deal with uncertainties associated with the data. The proposed approach uses fuzzy soft sets to predict links between nodes. The Jaccard's Coefficient gives better result as compared to other metrics, so it is adopted to combine with the fuzzy soft sets.

2. Preliminaries and Basic Definitions

In proposed work, objects are taken as link and parameters are the features or actions performed by the users. Communication between two nodes in terms of features is responsible for making a new link. Weights are applied on different features to represent the importance of that feature. The next subsection defines some terms and notations used for building the proposed models.

2.1. Some Definitions

Definition 2.1 *A social network user is defined as the set* $U = \{u_1, u_2, u_3, \cdots, u_n\}$ *in which* u_1,u_2,u_3, $\cdots$, u_n *are users or nodes. Each user has a user identification number, by which the user can login and uses the services of that network.*

Definition 2.2 *Different links are defined as the set* $L = \{l_1, l_2, l_3, \cdots, l_n\}$. *Node pairs set* $N_p = \{(u_1, u_2), (u_2, u_3), \cdots, (u_1, u_n)(u_2, u_n), \cdots, (u_{n-1}, u_n)\}$ *is the set of pairs of two users. A subset of* N_p *is also defined as a link.*

Definition 2.3 *Feature Vector is defined as the set* $F = \{f_1, f_2, f_3, \cdots, f_n\}$. *Features are different actions taken into account to predict link. Features are identified by a unique identification called feature number. For example, consider two users A and B of a network. If A communicates with B in terms of like, comment and message, then A has a feature vector* $F_{A-B} = \{like, comment, message\}$. *And if user B is not responding to user A, then B has a feature vector* $F_{B-A} = \{\}$.