A Novel Method to Dynamically Fix Threshold for Node Neighbourhood Based Link Prediction Techniques

Anand Kumar Gupta, Jaypee Institute of Information Technology, India
Neetu Sardana, Jaypee Institute of Information Technology, India

ABSTRACT

The objective of an online social network is to amplify the stream of information among the users. This goal can be accomplished by maximizing interconnectivity among users using link prediction techniques. Existing link prediction techniques uses varied heuristics such as similarity score to predict possible connections. Link prediction can be considered a binary classification problem where probable class outcomes are presence and absence of connections. One of the challenges in classification is to decide threshold value. Since the social network is exceptionally dynamic in nature and each user possess different features, it is difficult to choose a static, common threshold which decides whether two non-connected users will form interconnectivity. This article proposes a novel technique, FIXT, that dynamically decides the threshold value for predicting the possibility of new link formation. The article evaluates the performance of FIXT with six baseline techniques. The comparative results depict that FIXT achieves accuracy up to 93% and outperforms baseline techniques.

KEYWORDS
Classification, Ego Network, Link Prediction, Personalized Recommendation, Social Network Analysis, Statistical Approach, Threshold, Web Semantics

INTRODUCTION

In the current era, the usage of online social networks has grown exponentially. Nowadays, social networking websites offers various kind of services such as exchange of information, marketing of products or services, product review collection, opinion collection regarding an event, social or political awareness, etc. Services offered by online social networks can be efficiently utilized if the inter-connectivity among users in online social network is high. Since the online social network is continuously evolving, hence, to maximize the information flow in the existing network, it is essential to recognize the number of disconnected users in the present network system. In the present scenario, link prediction is one of the most fundamental problems in the broad domain of social network analysis (Zhou et al., 2018). Link prediction techniques are employed to calculate the possibility of any future link formation among the disconnected pair of users in current social network. Link

DOI: 10.4018/IJITWE.2020010102

Copyright © 2020, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.
prediction techniques generally uses common shared attributes of users (or nodes) to compute whether two users will get connected in the near future or not.

Link prediction problem in an online social network is usually considered as a binary classification problem (Hassan et al., 2011) (Lee & Seung, 2001). Generally, a trained classifier predicts if any association is possible among the two disconnected nodes in the future or not. The classifiers use several scoring functions such as Common Neighbours, Jaccard Index, Katz Index, etc., in order to measure the probability of link formation (Zhang & Chen, 2018). In the binary classification problems, one of the key challenges while classifying the objects is to decide the threshold value for object classification. The threshold value used for classification should neither be too high nor too low. A high threshold value might reduce the number of correct predictions (true positives) whereas a low threshold might cause the huge number of false predictions (false positives). For a given network, a link prediction technique calculates a similarity value among users based on varied attributes and then uses a fixed threshold value to determine if two users will get linked in the future or not. Two users will have high likelihood to be linked to each other in near future if the similarity value among them is greater than the given threshold value (calculated on the basis on some heuristic). Otherwise, if the calculated similarity score between a pair of nodes is lower than the given threshold value, it can be inferred that the possibility of a new association among the two nodes (or users) is minimal.

(Wang et al., 2015) presented an exhaustive survey of 130 research publications related to link prediction techniques in social networks. They classified the evaluation techniques used in analysing social networks under two categories: fixed threshold-based metrics and threshold curves. Fixed threshold-based metrics use precision and recall on Top-N predictions for evaluation of link prediction techniques. Receiver operating characteristic (ROC) curves are also widely-used measure for fixed threshold-based metrics. They further reported that precision and recall metrics suffers from a constraint that these metrics are based upon a fixed threshold value. They further stated that it is challenging to describe the link prediction evaluation approaches as compared to the normal classification procedures. They stressed upon the need of new evaluation methods or performance metrics for link prediction techniques. Hence, in order to improve the existing evaluation metrics dynamic thresholding techniques can be utilized. It is important to adopt a systematic methodology which can decide an appropriate threshold value to classify objects optimally as it will also enhance the prediction accuracy of the link prediction model. Setting the threshold value for any network is a challenging task as each user in the online social network is different and has its own connectivity pattern. Hence, it is important to set a threshold value for every user depending upon its individual connections in the considered network. This paper proposes a statistical method that uses the connectivity pattern of individual users and decides the threshold value based on which most probable connections can be predicted. Every user has its own social circles that forms its ego network (Mcauley & Leskovec, 2014). The threshold value is computed for every ego network that decides whether a link formation will take place or not. Alters (direct connections) of an ego network are considered as an independent ego network to extend the analysis over the whole network. The main research contributions in this work include:

1. This paper proposes a novel approach for setting the dynamic threshold value for node neighborhood-based link prediction techniques named FIXT (Fixed Threshold);
2. The paper presents the experimental results obtained while validating the proposed approach, FIXT, on a real-world dataset of Facebook collected from SNAP repository.

The rest of this paper is organized as follows. Section 2 briefly describes the necessary background details about various node neighborhood-based link prediction techniques. Section 3 presents the proposed standard deviation based statistical technique, FIXT for fixing the threshold value. Section 4
Related Content

Application of Data Mining in e-Commerce
www.igi-global.com/chapter/application-of-data-mining-in-e-commerce/137352?camid=4v1a

An Improved Task Scheduling Mechanism Using Multi-Criteria Decision Making in Cloud Computing
www.igi-global.com/article/an-improved-task-scheduling-mechanism-using-multi-criteria-decision-making-in-cloud-computing/222722?camid=4v1a

Evaluating E-Government
www.igi-global.com/chapter/evaluating-e-government/97020?camid=4v1a
Combining Expert Systems and Multiple Intelligences in an Adaptive and Intelligent Tutoring System


www.igi-global.com/article/combining-expert-systems-and-multiple-intelligences-in-an-adaptive-and-intelligent-tutoring-system/100050?camid=4v1a