International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2018 IJSRCSEIT | Volume 3 | Issue 1 | ISSN : 2456-3307

Survey on Link Prediction System

Kaustubh Mahajan, Prof. Mamta Bhamare

Department of Computer Engineering, MIT, Pune, Maharashtra, India

ABSTRACT

In the traditional link prediction problem, a snapshot of a social network is used as a starting point to predict, by means of graph-theoretic measures, the links that are likely to appear in the future. Link prediction in social networks has many potential applications such as recommending new items to users, friendship suggestion and discovering spurious connections. There are number of technique used for solving the link prediction problem. In this paper, discuss the survey on different techniques for link prediction problem. We also discuss cold start link problem. Also discussed the applications of machine learning to link prediction technique.

Keywords : Social Network, link prediction, machine learning, feature selection.

I. INTRODUCTION

A link is a connection between two nodes in a network. This simple concept can be used to represent extremely complex systems where a large number of elements interact among them. The proliferation of data that can be represented as networks has created new opportunities but also new challenges in the field of data mining. A large number of problems related to network mining are currently being studied, including community detection [8], structural network analysis [9], and network visualization. One of the most interesting network-related problems is link prediction, which consists of inferring the existence of new relationships or still unknown interactions between pairs of entities based on their properties and the currently observed links. The approaches and techniques designed to solve this problem enable the extraction of implicit information present in the network and the identification of spurious links, as well as modeling and evaluating network evolution mechanisms.

Link prediction methods have been successfully applied to biological networks in order to predict previously unknown interactions between proteins, significantly reducing the costs of empirical approaches. They have also been used to model highly dynamic systems, such as email or telephone call networks. Link prediction techniques are widely present in our daily lives, suggesting people we may know but we are not still connected to in our social networks or products we could be interested in in electronic commerce [Huang et al. 2005].

Networks have been extensively studied since the proposition of the first basic models to identify the laws that drive network formation and lead to their structural features. Some techniques that could be considered as link prediction methods were then proposed. However, it was not until specific link-prediction-oriented seminal works, which performed a comprehensive analysis of the problem, that this field came under the spotlight due to its applicability and usefulness in a great variety of contexts.

Link prediction is grounded in the empirical evidence that two entities are more likely to interact if they are similar. Similarity in networks must be understood as an abstract concept and could vary between networks. Understanding the domain that

the network represents is a crucial step to define the similarity between two nodes. In most domains, it has been observed that nodes tend to form highly connected communities. This has led to the common definition of similarity as the amount of relevant direct or indirect paths between nodes. One of the main difficulties in link prediction is achieving a good balance between the amount of information considered to perform the prediction and the algorithm complexity of the techniques needed to collect that information. Since actual networks are usually formed by hundreds of thousands or even millions of nodes, the techniques used to perform link prediction must be highly efficient. However, considering only local information could lead to poor predictions, especially in very sparse networks.

Link prediction in social networks

Traditionally, social networks have been modelled in a single layer, where all nodes and edges are from the same type. The link prediction problem considers the likelihood of existence of future link between two nodes in a network where there is no link between them in the current state [9]. The link prediction problem has a long history in modern information sciences; however, it was first systematically introduced for mining social networks in [10]. They formulated the problem for a given social network G as to predict list of edges not presented in $G[t_0, t'_0]$, but predicted to exist in $G[t_1, t'_1]$, where G[t, t'] denotes the subgraph of G at the time-stamp interval of [t, t']. $[t_0,t'_0]$ is the training interval and $[t_1, t'_1]$ is testing interval. Hasanet al.[11] extended this model and showed that one can obtain better results by employing not only the graph topology, but also contextual information on the nodes and edges. Song et al.[12] used matrix factorization to estimate similarity between pair of nodes in a real social network.

Mining heterogeneous networks is more challenging than those with the same type of nodes and links. Link prediction of heterogeneous networks has been studied in some works [13]. Hristova et al.[12] expressed a number of research questions related to the link prediction problem in heterogeneous networks and proposed a model to solve them. Meta-paths were introduced in [14] to analyse heterogeneous networks, extract similarity, predict links and rank their nodes.

In this paper we study about the related work done, in section II, the proposed approach, section III and at final we provide a conclusion in section IV.

II. LITERATURE REVIEW

This paper abstracts complex systems as multi-relational networks, and employs latent space network model to extract low-dimensional factors of sub-networks and adopts likelihood ratio test to examine correlation between factors. Then, regression between target sub-networks and correlated auxiliary sub-networks could be established for cold start link prediction[1].

The general overview of the system is:

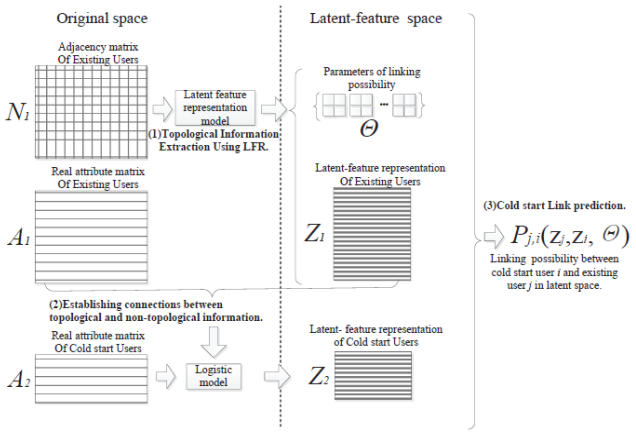


Figure 1. Cold-start Link Prediction: Establishing Connections between Non-topological and Topological Information [1].

Future Aspect: study how to conduct incremental learning on latent-feature representation model so that the model could be adapted to a dynamic circumstance when cold-start nodes consecutively

join a network. Besides, it also has many real applications based on the cold-start link prediction model. For example, our method can contribute to speed up the initial growth for a new online App based on social networking services (SNS).

In [2] construct a hidden space mapping a network into Euclidean space based solely. on the connection structures of a network. Compared with real geographical locations of nodes, our reconstructed locations are in conformity with those real ones. The distances between nodes in our hidden space could serve as a novel similarity metric in link prediction. In addition, we hybrid our hidden space method with other state-of-the-art similarity methods which substantially outperforms the existing methods on the prediction accuracy. Hence, our hidden space reconstruction model provides a fresh perspective to understand the network structure, which in particular casts a new light on link prediction.

In [3] illustrate the use of the rough set theory in predicting links over the Facebook social network based on homophilic features. Other supervised learning algorithms are also employed in our experiments and compared with the rough set classifier, such as naive Bayes, J48 decision tree, support vector machine, logistic regression, and multilayer perceptron neural network. Moreover, we studied the influence of the "common groups" and "common page likes" homophilic features on predicting friendship between users of Facebook, and also studied the effect of using the Jaccard coefficient in measuring the similarity between users homophilic attributes compared with using the overlap coefficient. We conducted our experiments on two different datasets obtained from the Facebook online social network, where users in each dataset live within the same geographical region.

Flow of the system are:

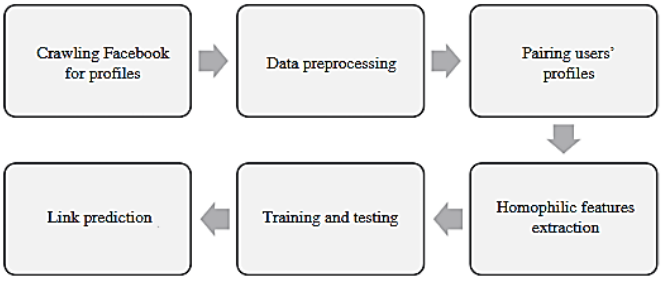


Figure 2. Homophily Based Link Prediction Process[3]

Future Aspect: System will incorporate some topological features in link prediction, such as the Jaccard coefficient for common neighbors and the Adamic-Adar measure for link prediction using the different classifiers stated in this work and the rough sets classifier. We will compare the influence of each category of features (homophilic and topological) separately in link prediction, and then we will investigate what will be the results of combining both categories together when predicting links in online social network.

In [4] for instance, the problem of recommending friends with similar location preference in real world cannot be solved by the method based on the social network topology or non-topological information (such as user profile). A new approach that recommends friends with similar location preference for LBSN's users is proposed, in which both the online friendship information and the offline user behavior are considered. The theories and methods including Markov chain, cosine similarity based on location clustering and threshold evaluation are used in the proposed approach. Finally, rationality and effectiveness of the algorithm is verified by using a real dataset which is from a LBSN (Gowalla).

Future Aspect: the algorithm will be extended for other application domains, such as micro-blog and biological fields' and we will improve the algorithm from the following two aspects: Firstly, in order to get a better user experience, we will try to analyze the periodicity of the user's behavior to better define the user's location preference similarity; secondly, in order to obtain a better

scalability, we will improve the algorithm to make it run on the distributed platform spark.

In [5] author will apply supervised learning algorithms to the link prediction prediction problem using a large set of topological features. Given a network at two different points in time, we train a learning algorithm to identify pairs of edges which appear in the newer network but not in the older network. Then, for a pair of nodes, we use the classification probability of the learning algorithm as our link prediction heuristic. Furthermore, we show that our network-specific heuristics outperform generic heuristics such as the Adamic/Adar coefficient and the Jaccard coefficient.

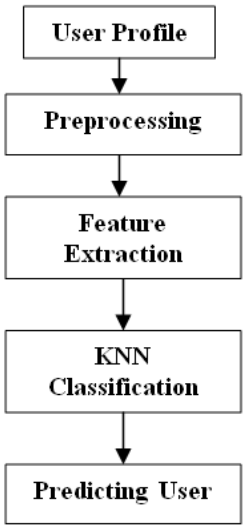
In [6] proposed a multi-attribute ranking method based on the Technique for Order Preference by Similarity to Ideal Object (TOPSIS) to evaluate the local centrality of common neighbor nodes comprehensively. In order to make the local centrality indicator based on TOPSIS achieve better results, we also proposed a new weight calculation method for the attributes normalization matrix.

In [7] study the link prediction problem in multiplex networks. As an example, we consider a multiplex network of Twitter (as a microblogging service) and Foursquare (as a location-based social network). We consider social networks of the same users in these two platforms and develop a meta-path-based algorithm for predicting the links. The connectivity information of the two layers is used to predict the links in Foursquare network. Three classical classifiers (naive Bayes, support vector machines (SVM) and K-nearest neighbour) are used for the classification task. Although the networks are not highly correlated in the layers, our experiments show that including the cross-layer information significantly improves the prediction performance.

III. PROPOSED APPROACH

In this system Latent feature representation technique is used for extracting the features of the facebook users and for link prediction KNN algorithm is used. This system predicts the facebook friend on the basis of description or biography given in the profile.

The flow of the system we have proposed is as follows:



The main objective of the system are:

- An efficient method for network correlation analysis is proposed.
- A novel strategy for cold-start link prediction in multi-relational networks is proposed.
- Classification technique is proposed for predicting the link between the two user.
- Solve the problem of cold-start in link prediction network for online social networking.

This system is used for online social networking, for predicting the friend on the basis of their biography profile.

IV. CONCLUSION AND FUTURE SCOPE

In this paper we discuss the survey on the different link prediction techniques, this survey shows the limitation of the existing system. From the survey it is concluded that there is a need of the system which solve the problem of cold-link prediction problem and improve the performance of the link prediction techniques. This feature is vastly used in social networking sites for recommending the users.

V. REFERENCES

[1]. Shun-yaoWu, QiZhang, MeiWu, "Cold-start link prediction in multi-relational networks", in science direct, Volume 381, Issue 39, 17 October 2017.

[2]. Hao Liao1, Mingyang Zhou1, Zong-wen Wei1, Rui Mao1,Alexandre Vidmer1, Yi- Cheng Zhang, "Hidden space reconstruction inspires link prediction in complex networks", arXiv:1705.02199 cs.SI], March 2017.

[3]. Islam Elkabani and Roa A. Aboo Khachfeh, "Homophily-Based Link Prediction in the Facebook Online Social Network: A Rough Sets Approach", J. Intell. Syst. 2015; 24(4): 491-503.

[4]. Manfang Wu, Zhanquan Wang, HaiLong Hu,"Friend Recommendation Algorithm for Online Social Networks Based on Location Preference", 3rd International Conference on Information Science and Control Engineering, 2016.

[5]. Kyle Julian (kjulian3), Wayne Lu (waynelu), "Application of Machine Learning to Link Prediction", in December 16, 2016.

[6]. Mingqiang Zhou, Rongchen Liu, Xin Zhao, Qingsheng Zhu, "Link Prediction Algorithm Based on Local Centrality of Common Neighbor Nodes Using Multi-Attribute Ranking", in the 12th International Conference on Computer Science & Education (ICCSE 2017).

[7]. Jalili M, Orouskhani Y, Asgari M, Alipourfard N, Perc M. 2017 Link prediction in multiplex online social networks. R. Soc. open sci. 4: 160863.http://dx.doi.org/10.1098/rsos.160863.

[8]. Santo Fortunato. 2010. Community detection in graphs. Physics Reports 486, 3 (2010), 75-174.

[9]. Al Hasan M, Zaki MJ. 2011 A survey of link prediction in social networks. In Social network data analytics pp. 243-275. Berlin, Germany: Springer.

[10]. Liben-Nowell D, Kleinberg J. 2007 The link-prediction problem for social networks. J.Am. Soc.Inf.Sci.Technol. 58, 1019-1031. (doi:10.1002/asi.20591)

[11]. Hasan MA, Chaoji V, Salem S, Zaki M. 2006 Link prediction using supervised learning. In SDM Workshop of Link Analysis, Counterter rorismand Security,Maryland,22April

[12]. Song HH, Cho TW, Dave V, Zhang Y. 2009 Scalable proximity estimation and link prediction in online social networks. In Internet Measurement Conf., Chicago,IL,4-6November.

[13]. Sun Y, Barber R, Gupta M, Aggarwal CC, Han J. 2011 Co-author relationship prediction in heterogeneous bibliographic networks. In Int.Conf. On Advances in Social Networks Analysis and Mining, Taiwan,25-27July , pp. 121-128.

[14]. Sun Y, Han J. 2012 Mining heterogeneous Information networks: principles and methodologies. Williston, VT: Morgan & Claypool.