IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 5, May 2016

An Analytical Study On Link Prediction In Social Networks

Nitin Arora

Amity University Uttar Pradesh, Noida

Abstract: As a result of exponential growth of social networking sites, huge population on earth is now connected with others through social media and form a global e-society. With every instance of time new relationships are initiate and links are getting established and therefore new nodes are added to any online social network. There is natural yearning among the people for the prediction of new relationship in the society, the imitation of which is also obtainable in e-society. Social analytics is an area which is now a day becoming popular among the researchers due to availability of open repositories containing terabytes of data generated by the social networking portals. Link prediction in social network is one of most popular research problem among the knowledge workers and researchers. In last, decade many empirical models for link predication are proposed. This is a survey paper presenting a comprehensive survey of techniques for link prediction based on their approaches, applications and performance in various kind of electronic social networks.

Keywords: Social Networks, Link prediction, Online Social Network Include at least 4 keywords or phrases.

I. INTRODUCTION

Social Network

Social Networks can be defined as the collection of the people, and the ways how they are related to each other. Social Network sites are that can allow to individual:

- create a public profile within the system
- Make a list of users with whom the person can have the connection
- Can view the list of connections which made by the others within the system.

A Social Network is there which consists of a given directed graph G = (V, E) and a set of the attributes for each node in V(for address, name, phone no, native place) and a set of attributes for each edge in E(for instance, type of relationship).

A social network is a composition of persons, which are attached by single or further particular types of interdependency such as friendship, interest, dislikes, home town. In Recent years as we noticed that the online social networks have been growing so faster and exponentially. The term Social network is known as a structure that consists of personal or organizational groups. As the nodes in these networks are connected to each node with another node or by some other dependencies.

It is the best or the popular way of model to interact among various different types community or group. As the community or group can be visualize this interactions as graph, where each vertex correspond to an individual or several collection of groups and an edge correspond to various form of relationship between the equivalent persons or nodes. Although Social network is the most popular way to grow latest opportunity or to do friendships, distribute ideas and doing business online. As Online social media services such as facebook, twitter and flicker, these are just some names that have become part of our daily life of millions of people around the world. Now a day it is difficult to find relationship between people in social networks as each data in the social network needs some methods to analyze.

Link Prediction

Link prediction can be known as the problem of predicting the future links among two individuals. We can consider the attributes/features of the dataset and other observed links. We can be predicting the relations among people in social networks such as to predict the future friendship, guessing the contribution of people in the events. After that we will find some new links which have a more occurrence to built a new links with its other members in near future. It can be known as the problem of link prediction.

Here a network is there known as time t1 and a future time is there which is known as tz, now the problem arises here is to calculate the new relations(link) of friendship that are expected to come into the network within the given time interval as [t, t′].As Liben-Nobell and Kleinberg[] defines that, the problem of link predication is about to what level the social networks can be designed using characteristics which is inherent to network . The below figure demonstrate the simple link prediction as :

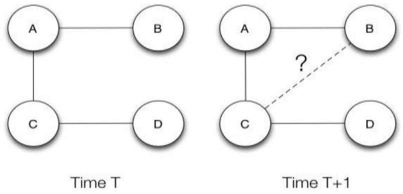


Fig 1: Definition of Links Prediction

DOI 10.17148/IJARCCE.2016.5534

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

Literature Survey:
Liu et.al.[13] presented two algorithms LRW and SRW for Link Prediction Problem based on local random walks on simple networks. The network they considered is defined as follows: Let G(V, E) be a simple network, where V is a set of nodes and E be the set of links. Numerous links and self-relations are not allowed. A random walk can be described as a Markov chain which describes the sequence of nodes visited by a random walker. The algorithms were experimentally compared to previously proposed algorithms. The comparisons were based on two measures AUC and precision.

Algorithm SRW had the best performance on 4 of the 5 datasets based on AUC. Algorithm SRW had the best performance on 2 of the 5 datasets and SRW had the best performance for 1 of the 5 datasets when compared based on Precision. In the cases where the algorithms did not show the best result, the result was still very close to the best. Mohammad et. al.[14] discussed the supervised machine learning algorithm for Link Prediction. They have converted the network into graph. They have used binary classification method. They applied many learning algorithm for link prediction task and did comparative analysis. They also evaluated the attribute rank based on various factors and judge their strength for link prediction.

David et. al.[16] considered five network data from arxiv.org and implemented all the existing technique like chart distance predictor, as prediction based on node-neighbourhood, Predictors based on lane topology and Meta approaches. They have shown the comparative analysis and as a result there is no particular winner among all the techniques. But some techniques performed well. Giovanni Z[2] introduced a game theoretical approach which is based on graph Transduction game. They have tested the proposed work on real world data from Tuenti OSN and compared with standard local measures. The proposed technique performs better than the standard local measures.

Lars B[1] developed a technique based on supervised random walks. They combined the information from network structure with node and edge features. They assigned the score to the edges and introduced training algorithm for learning edge strength estimation function. They have taken facebook and co-authorship network data. The result shows good improvement in predicting the future links.

Mathematical Description of Link Prediction
The problem of link prediction can be defined as: Consider a social network G = (V, E) which contains two sets: i) set of vertices V and ii) set of edges E.

Let we have a dataset which can be organized in the social network $G = (V, E)$ form, where E is defined as the set of boundaries(edges) and V is known as the set of the vertices then the task to guess how likely an unnoticed link $e_{ij} \notin E$ exists among every pair of nodes v_i , v_j in the data network.

Classification of Link Prediction Techniques:

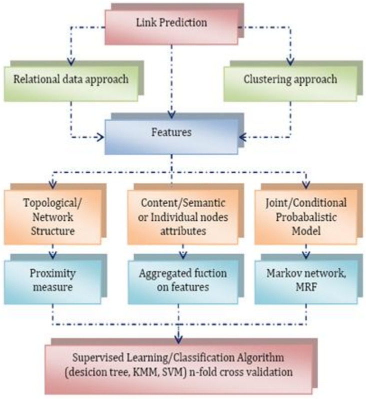
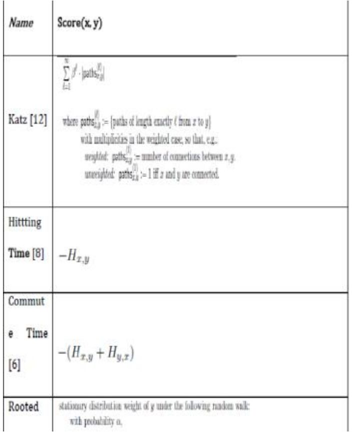

Link Prediction
Relational data approach
Clustering approach
Features
Topological/ Network Structure
Content/Semantic or Individual nodes attributes
Joint/Conditional Probabalistic Model
Proximity measure
Aggregated fuction on features
Markov network, MRF
Supervised Learning/Classification Algorithm (desicion tree, KMM, SVM) n-fold cross validation

Figure 3: Different approaches to link prediction task

There are various techniques for Link Prediction:
a)Predictor based on Graph distance:
The crucial method for calculating node similarity in social networks is by calculating the graph variation among them i.e. we have to discover the pair (x, y) having the length i.e. the direct(short) path by linking them in graph. So the direct(short) path interpreter that selects a random subset of two-distance pairs.

Name	Score(x, y)
Common Neighbours [10]	$\|\Gamma(x) \cap \Gamma(y)\|$
Jaccard Coefficient[9]	$\frac{\|\Gamma(x) \cap \Gamma(y)\|}{\|\Gamma(x) \cup \Gamma(y)\|}$
Adamic/Adar[1]	$\sum_{z \in \Gamma(x) \cap \Gamma(y)} 1/\log\|\Gamma(z)\|$
Preferential Attachment[15]	$\|\Gamma x\|. \|\Gamma y\|$

c) **Path Topology based predictors**



Name	Score(x, y)
Katz [12]	$\sum_{\ell=1}^{\infty} \beta^{\ell} \cdot \|\text{paths}^{\langle \ell \rangle}_{x,y}\|$ where $\text{paths}^{\langle \ell \rangle}_{x,y}$:= {paths of length exactly ℓ from x to y} with multiplicities in the weighted case, so that, e.g.: *weighted:* $\text{paths}^{\langle 1 \rangle}_{x,y}$:= number of connections between x, y. *unweighted:* $\text{paths}^{\langle 1 \rangle}_{x,y}$:= 1 iff x and y are connected.
Hittting Time [8]	$-H_{x,y}$
Commute Time [6]	$-(H_{x,y} + H_{y,x})$
Rooted	stationary distribution weight of y under the following random walk: with probability α,

Higher Level Approaches: There are some meta approaches that can be used for Link Prediction.

a) Low Rank Approximation:
We use the adjacency matrix for representing the graph G. We generally have mapping for each node x to its row r(x) in matrix and defining the score to be inner product of r(x) and r(y). We will choose a small number k and calculate the rank k matrix. We can use singular value decomposition method for this.

DOI 10.17148/IJARCCE.2016.5534

2) **Unseen Bigrams**:

We can define Link Prediction in language modelling as to predict frequencies for unnoticed bigrams i.e. pair of words that co-occur in testing corpus but can not occur in training data. We can increase the estimate of score(x, y) by using the value of score(z, y).

a) Clustering:

We can have the quality of Link Prediction technique by deleting the weak edges in graph with the clustering process. Consider a graph and calculate the score(x, y) for all the edge. Then identify the weak edges and delete from the graph. Recompute the score(x, y) for the graph, then we will have more accurate result for future links.

II. MACHINE LEARNING APPROACHES

Machine learning approaches need to design the program which learn from past data and then to predict the future links of testing data. It is broadly classified into two types: Supervised and unsupervised learning. Supervised learning algorithms need training set where as it is not required in unsupervised learning. There are various techniques under supervised learning: Bayesian classification, Neural Network, Markov based model and pattern discovery. These techniques have been used in past for Link Prediction. Mohammad et. al.[15] Considered two data set BIOBASE and DBLP and extracted the features from the dataset. They evaluated the effectiveness of the features and implemented the classification model(Decision Tree, SVM(Linear Kernel), SVM(RBF Kernel), K-nearest neighbors multilayer perceptron, RBF network, Naïve Bayes, Bagging). They calculated various performances metrics like Accuracy, Precision, Recall, F1-ratio. Ryan N.[5] used the supervised classification scheme. Supervised algorithms have the capability of capturing important relationships between the topology in the network. As they examined the factor by encouraging the use of supervised framework in the network and proposed a high performance framework for Link Prediction task. Nesserine B.[4] have introduced new variation of topological attributes for evaluating the similarity of two nodes in near future. As they have taken two data sets: DBLP Bibliographical database and another is the Bipartite graph 8 years history of transactions. There is clear improvement in the proposed prediction model. Hially R.[3] did the analysis of supervised learning algorithms for weighted networks. They have analyzed supervised Link Prediction for with and without co-authorship network. The result shows the satisfactory results for weighted co-authorship network.

III. DISCUSSION AND CONCLUSION

Among all the approaches machine learning approaches gives the better result. Although the classifier needs to be trained with the data. Machine learning approach has a tendency to require significant computing time for Link Prediction. It is critical to make the balance between training dataset and available resource to train data. It requires effort of making the balance between functionality and effectiveness. Support vector machine are computationally expensive. We presented a comprehensive survey of various approaches used for Link Prediction in Social Network inclusive of traditional and emerging approaches. We have observed that there is a battle between supervised and unsupervised approach. Link Prediction has become very popular area in academia and industry due to the impact of network topology and node attributes on link prediction. We noticed that the above mentioned approaches have been successful in link prediction in different scenarios. It may depend upon the type of network. One approach outperformed than the other for one network. We observed that neuro-fuzzy approach has not been used for link prediction. A neuro-fuzzy approach could be used in combination with the techniques for achieving the better accuracy. We can explore the combination of genetic and evolutionary approaches for better accuracy. This survey may help the researchers to have knowledge about the various techniques for Link Prediction and may give the idea for future work.

REFERENCES

[1]. Lars B., "Supervised Random walks: Predicting and Recommending links in Social Networks", WSDM'11, feb 9-12, 2011, ACM 978-1-14503-0493-1/11/02, Hongkong, China

[2]. Giovanni Z., "Games of friends: A game theoretical approach for Link Prediction in online Social Network", Association for the advancement of artificial Intelligence, 2013

[3]. HiallyR.,"Supervised Learning for Link Prediction for weighted networks"

[4]. NesserineB.,"Supervised Machine learning applied to Link Prediction in Bipartite Social Networks", 978-0-7695-4138-9/10, 2010 IEEE Computer Society

[5]. Ryan. N. L.,"New perspectives and methods in Link Prediction" KDD'10, july 25-28, 2010 Washington DC, USA, ACM 978-1-4503-0055-1/10/07

[6]. Jaccard P., Bulletin de la SocieteVaudoise des Science Naturelles 37, 547(1901)

[7]. M. E. J. Newman, "Clustering and preferential attachment in growing networks," Physical review letters E, vol. 64, 2001.

[8]. G. Jeh and J. Widom, "SimRank: A Measure of Structural Context Similarity," in Proc. the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2000, pp. 538-543.

[9]. L. A. Adamic and E. Adar, "Predicting missing links via local information," Social Networks, vol. 25, no. 3, pp. 211-230, July 2003

[10]. L. Katz, "A new status index derived from sociamatric analysis," Psychometrika, vol. 18, no. 1, pp. 39 – 43, March 1953.

[11]. P. Jaccard, Bulletin De La SocieteVaudoise Des Science Naturelles, Nabu Press, vol. 37, no. 547, 1901.

[12]. M. E. J. Newman, "Clustering & preferential attachment in growing networks," Physical review letters E, vol. 64, 2001.

[13]. W. Liu and L. Lü, "Link prediction based on local random walk," Europhysics Letters, no. 5, 2010.

[14]. Mohammad Al Hasan, and Mohammed J. Zaki, "A Survery of Link Prediction in Social Network", in Social Network Data Analytics, pp 243-275, 2011, Springer

[15]. Mohammad Al Hasan, VineetChaoji, Saeed Salem, Mohammed Zaki , "Link Prediction using Supervised Learning", SIAM Workshop on Link Analysis, Counterterrorism and Security with SIAM Data Mining Conference, Bethesda, MD,2006

[16]. D. L. Nowell and J. Kleinberg, "The link-prediction problem for social network," Journal of the American Society for information science and Technology, vol. 58, no. 7, pp. 1019-1031, 2007.

DOI 10.17148/IJARCCE.2016.5534