IJCSE International Journal of Computer Sciences and Engineering Open Access
Review Paper Volume-3, Issue-10 E-ISSN: 2347-2693

Supervised Random Walks for Predicting Links in Social Networks: A Study

A.Vihashini[1*]2and G.T.Prabavathi

[1*,2]Dept. of Computer Science, Gobi Arts & Science College, Gobichettipalayam

www.ijcseonline.org

Received: Oct/04/2015 Revised: Oct/14/2015 Accepted: Oct/22/2015 Published: Oct/31/2015

***Abstract*—**Predicting is future relationships from a given snapshot of a network or to infer the interactions among existing members that are likely to occur in the near future is called as link prediction. One of the interesting areas of research in social network is prediction of links. There are various techniques for inferring missing links or additional links that are not directly visible but may occur in the future. Random walk is a popular approach which uses node and edge features to solve the problem of link prediction. Supervised random walks combine the network structure with the characteristics of nodes and edges and acts as a powerful tool for predicting the missing and future links. In this paper a study has been made on various algorithms that uses supervised random walk approach for predicting links in social networks.

Keywords*—Social networks, Link prediction, Supervised Random walk*

I. INTRODUCTION

Social networks are graph structures whose nodes or vertices represent people or other entities embedded in a social context and whose edges represent interaction or these entities collaboration between these entities [5]. Social network analysis is an interesting field of study and are dynamic in nature, i.e., they grow or shrink over time through the addition or removal of new edges, indicating the appearance or disappearance of interactions in the underlying social structure. A basic problem in the social networks evolution is link prediction. Link prediction is a sub field of social network analysis concerned with the problem of predicting the (future) existence of links among nodes in a social network [7]. Link prediction in social network is used to predict the possible interaction among the nodes in the near future. It focuses on the interactions between objects rather than objects themselves and has various interesting applications in social networks. Random walk is a stochastic process that is formed by successive summation of independent and identically distributed random variables. The supervised random walk is learn edge strengths with in a graph and these edge strengths are then used for predicting new links.

Link prediction has been applied to social networking websites such as Twitter, Flicker and Face Book to infer the interesting relationships among the nodes in the network. The applications of link prediction differ according to the network to which it is applied. For e.g. predicting interactions between proteins in bioinformatics networks; to create the recommendation system in e-commerce sites; to predict existing yet unknown links (hidden criminal groups) in terrorist networks; prediction of links that may appear in the future of evolving networks in co-authorship networks[9].

This paper reviews various supervised random walk algorithms that predict links in social networks. Chapter 2 explains the use of link prediction problem. Chapter 3 explains the features of supervised random walks. In Chapter 4 a study has been made on various authors' algorithms based on supervised random walks and in Chapter 5 conclusion is given.

II. LINK PREDICTION

Given a snapshot of a social network at time *t*, the problem of link prediction is to infer which interacting members are likely to occur in the future or which among existing interactions are missing during the interval from time t to a given future time t+1[1]. The links can be predicted using supervised or unsupervised approaches. Unsupervised methods assign a score for each pair of nodes with base on neighborhood nodes (local) or path (global) information. Global methods usually achieve higher accuracy than local methods. Global methods are very time consuming and usually infeasible for large-scale networks. Unsupervised strategy considers the link prediction problem as a classification problem. Considering a social network, the network structure can be represented as feature vector for each pair of nodes. These vectors are used to train different classifiers to determine whether the link exist or not between a pair of nodes [6]. Link prediction algorithms can be classified into three categories: Node neighborhood approaches, path based approaches and meta approaches. Link prediction techniques often focus on global properties (graph conductance, hitting or commute times, katz score etc) or local properties (Adamic-Adar and many variations, node feature vectors), but rarely combine these two properties. The main focus on link prediction is to concentrate on link between objects instead of objects

themselves, thus differentiating from traditional data mining techniques which focuses on objects. Commonly link prediction problem algorithms combine information from the network structure with rich node and edge attribute data [10]. The link prediction problem has been extensively studied by various researchers. The first challenge of this problem is the structure of the network. In sparse networks, nodes contain connections to only a very small fraction of all nodes in the network. The second challenge is to what extent the links of the social network is to be modeled using the features intrinsic to the network itself. There are various algorithms that predict links in the areas such as bibliographic domain, molecular biology, protein-protein interaction network, customizing new friend suggestions for bloggers, criminal investigations, etc., Many researchers [2,5,8,11] use co-authorship network as an example since it provides effective data representation for link prediction and readily available on sites like DBLP.

III. SUPERVISED RANDOM WALKS

A random walk is a mathematical formalization of a path that consists of a succession of random steps. Given a graph and a starting point, a neighbor of it is selected at random and moved to this neighbor; then a neighbor of this point is selected at random, and moved to it etc. A Random walk is a stochastic process that consists of a sequence of discrete steps of fixed length. The (random) sequence of points selected this way is a random walk on the graph [6]. Random walk is more suitable for link prediction problems. Random-walk-based algorithm is based on the structure of the network and the node's attributes. Node and link attributes along with node structure information are used for prediction.

The link prediction problem can be approached in to two different strategies using random walks: Supervised and unsupervised. Supervised random walks are trained on a single snapshot of a graph at time t0 and immediately tests on its predictions for new links from a source node s at time t1. This is an important property of supervised random walk as it requires less training data to make same type and quality of predictions. The primary limitation of a supervised approach, compared to network method is the reduced richness of the data representation. In link prediction problem, the supervised random walk is normally used as a method for predicting links in the network as it demonstrates a good generalization and overall performance. Supervised learning strength is assigned to the edges that are likely to have new links. The strength is not set manually, but learned from the features of each edge and nodes between them. Supervised random walks is not limited to link prediction but can be applied to many other problems like graph recommendation anomaly detection, missing link, and expertise search and ranking. Next chapter presents a review of various algorithms that predict links using supervised random walks.

IV. LITERATURE REVIEW

The problem of link prediction from social networks has an extensive literature. This section briefly lists some of the existing works in the area of link prediction using supervised random walks. To address the challenge in both link prediction and recommendation Backstrom and Leskovec et. al [2] has combined the information from the network structure with node and edge level attributes and uses these attributes to guide a random walk on the graph. The supervised learning task has the goal to learn a function that assigns strengths to edges in the network such that a random walker is more likely to visit the nodes to which new links will be created in the future. This method demonstrates a good generalization of the supervised random walks; it can be used as a method for predicting links in the coauthor networks [2]. In this research work, the author designed a unified framework that considers rich node and edge features as well the structure of networks.

In the frame work Supervised Random Walk with Pre-Filtering, first a filter is executed to pick out the most probable links based on certain rules that includes common interest and social connections. In the second step the candidates are ranked based on the attributes and social relationships through supervised Random walk to learn the partial relationship and to rank those relationships for the recommendation [8]. The authors tested the framework from the real world (content-oriented asymmetric) social data collected for a specific period and proved that their framework predicted the appropriate links efficiently outperforming the base links including ordinary SRW with significant margin. The two *steps* reflect the progressive adjustment of prediction and improved the efficiency.

An effective recommendation algorithm should identify factors that influence link creation. Zhijun Yin et. al. [11] proposed an approach that estimates link relevance using random walk algorithm on a social graph that attributes social information to compute link recommendation in social networks. In this work, augmented graph for a social graphs based on a person attributes file is constructed. The edge weights are set for the augmented graph using global waiting in which more weights are attached to more promising attributes. The local importance of attribute specific person based on neighborhood is also defined. Both local and global attributes information are leveraged in to the framework by influencing edge weights. The author used two parameters λ *and* α for link recommendation. In addition to link recommendation, the framework also ranks that personalized attributes. The authors conducted the experiments on DBLP and IMDP datasets. The author suggested that the framework can be improved to identify semantic correlations for link recommendation.

In a coauthor ship social network G=<V,E>, where each edge e=<U,V> is defined as coauthoring a research article. M. A. Hasane et. al. [4] partitioned the range of publication years into two non-overlapping sub-ranges, where the first sub-range is selected as training years and the second as testing years. Then a classification dataset with those pairs was constructed representing positive and negative examples, depending on whether the author pairs published at least one paper in the testing years or not. Two authors are treated as close to each other, if their research works shares a larger set of identical keywords which was treated as proximity feature for predicting the links. The authors tested the link prediction problem with various classification algorithms such as SVM, DT, KNN, Naïve bays using metrics such as Precision-recall, F-values, squared error etc. and suggested SVM outperforms other algorithms. The authors proved that link prediction problem can be handled efficiently by modeling it as classification problem.

Ervin, Tasnadi,Gabor Berend et. al. [3] proposed a supervised machine learning algorithm that uses the structure of the social network to predict non-existing ed**g**es in it and also made use of feature graphs that were constructed based on implicit information provided in the dataset. The authors treated link prediction as a classification task, where task denotes whether the relation 'friendship' holds between a pair of users. Two random sample pair of user sets for which friendship relationship exists and not exists was selected as samples for positive and negative classes. The feature space consists of features derived from different feature groups that serve as different views of the classification instances. The feature space was derived from the auxiliary groups. The author improved the classification results by incorporating the implicit information that does not rely directly on the social network in the unified framework. This paper defined different ways to obtain the similarity scores for pairs of users based on the stationary distribution for rooted random walks on different graphs.

CONCLUSION

The link prediction is a problem of inferring missing links from an observed network to infer additional links that are not directly visible but most are likely to exit, is an interesting area of research. Link prediction problem can be solved through various approaches. Random walk is a prominently used approach in various papers that attempt to predict link in social networks. Random walk algorithms are simple frameworks for unifying the information from ensembles of paths between two nodes. Lot of research has been made in link prediction using supervised and unsupervised approaches. The supervised random walks combine the network structure with the character tics of nodes and edges of the network to predict the links. Random walk based supervised algorithms use node and edge features to learn the edge strengths such that the random walk on a weighted network is more likely to visit "positive" than "negative" nodes. The aim of this paper is to perform a study on algorithms that predicts missing and feature links using supervised random walks.

REFERENCES

[1] Abir De, Niloy Ganguly and Soumen Chakrabarti, Discriminative Link Prediction using Local Links, Node Features and Community Structure, **2013**, pp.1009-1018.

[2] L.Backstrom and J. Leskovec, Supervised random walks: Predicting and recommending links in social networks, WSDM '11 Proceedings of the fourth ACM international conference, **2011**, pp.635-644.

[3] Ervin Tasnadi and Gabor Berend-Supervised prediction of social network links using implicit sources of information, Proceedings of the 24th international conference,**2015**, pp.1117-1122.

[4] M. A. Hasan,V. Chaoji, S. Salem, and M. Zaki-Link prediction using supervised learning, SDM 06 workshop on Link Analysis, Counterterrorism and Security, **2006**.

[5] D. Liben-Nowell and J. Kleinberg, The link prediction problem for social networks, Proceedings of CIKM, **2003**, pp.556–559.

[6] László Lovász- Random Walks on Graphs: A Survey, Paul Erd˝os is Eighty(Volume2)Keszthely (Hungary), **1993**,pp.1–46.

[7] R.N.Lichtenwalter,J.T.Lussier, and N.V. Chawla-New perspectives and methods in link prediction, KDD '10 Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, **2010**,pp.243–252.

[8] Ting Jin, Tong Xu, Enhong Chen, Qi Liu, Haiping Ma, Jingsong Lv, Guoping Hu, Random Walk with Pre-filtering for Social Link Prediction,Proceedings of 2013 9th International conference on computational Intelligence and security,**2013**, pp.139 – 143.

[9] ValdisKrebs,Mapping networks of terrorist cells, Connections,Winter **2002**, pp.24(3):43-52.

[10] C.Wang, V.Satuluri, and S.Parthasarathy,Local probabilistic models for link prediction, Proceedings of the 2007 7th IEEE ICDM,IEEE Computer Society, **2007**, pp. 322-331.

[11] Zhijun yin,Manish Gupta, Tim Weninger and Jiawei Han, A Unified Framework for Link Recommendation Using Random Walks, Advances in Social Networks Analysis and Mining (ASONAM), **2010** International Conference, 2010,pp.152 – 159.