ORIENTAL JOURNAL OF
COMPUTER SCIENCE & TECHNOLOGY
An International Open Free Access, Peer Reviewed Research Journal
Published By: **Oriental Scientific Publishing Co., India.**
www.computerscijournal.org

ISSN: 0974-6471
June 2013,
Vol. 6, No. (2):
Pgs.93-98

Composite and Mutual Link Prediction using SVM in Social Networks

M. RAJENDRAN and K.THIRUKUMAR

Department of Computer Science and Engineering,
Dr.Mahalingam College of Engineering and Technology, (India)
Corresponding author Email : rajendranmscse@gmail.com , thirukumar@drmcet.ac.in

(Received: May 24, 2013; Accepted: June 04, 2013)

ABSTRACT

Link prediction is a key technique in many applications in social networks; where potential links between entities need to be predicted. Typical link prediction techniques deal with either uniform entities, i.e., company to company, applicant to applicant links, or non-mutual relationships, e.g., company to applicant links. However, there is a challenging problem of link prediction among the composite entities and mutual links; such as accurate prediction of matches on company dataset, jobs or workers on employment websites, where the links are mutually determined by both entities that composite entity belong to disjoint groups. The causes of interactions in these domains makes composite and mutual link prediction significantly different from the typical version of the problem. This work addresses these issues by proposing the Support Vector Machine model. By implementing the proposed algorithm it is expected that the accuracy will get increased in the link prediction problem.

Key words: Link prediction, Potential links, Composite, Mutual links, Support Vector Machine.

INTRODUCTION

A social networking service is an online service, platform, or site that target to facilitating the building of social networks or social relations between people who, for example, share importance's, activities, backgrounds, or real-life communications. A social network service consists of a representation of each user, his/her social links, and a variety of new services. This is used to model the interaction among the communities on the social networks. Where the graphs are used to represent the interactions between those communities, in which nodes are representing to people in some communities and links are representing the association between those people.

Understanding the association between two specific nodes by predicting the likelihood of a future but not currently existing association between them is a fundamental problem known as link prediction.

Interaction on the social network involves both positive and negative relationships, e.g., since attempts to establish a relationship may fail due to decline from the expected target. This generates links that signify rejection of invitations, disapproval of applications, or expression of disagreement with others' opinions. Such social networks are mutual since the sign of a link indicating whether it is positive or negative depends on the attitudes or belief of both entities forming the link. Moreover, mutual positive and negative relationships have

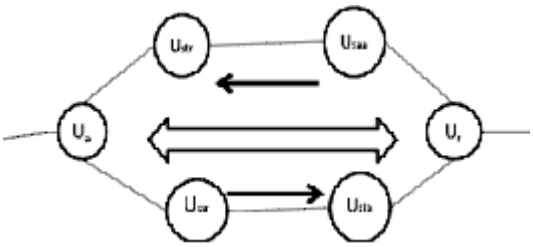

Fig.1: Collaborative Information for Composite and Mutual Link Prediction

InteractionL
Predicted Preference
Predicted match

been even less investigated. Recently, social network analysis has had a variety of applications, such as online dating sites, education admission portals as well as jobs, employment, career and hiring sites, where people in the networks have different roles and links between them can only between people in different roles. Such networks are composite, creating challenges for link prediction since existing approaches focus only on uniform networks where nodes in the networks have the same role and any of them may link to any other.

In this work, we propose a framework of composite and mutual link prediction on the network. Problem of the link prediction addresses the sign of links among the composite entities in the network. We address this problem and it is resolved by machine learning and to construct the structural feature for the machine learning. First we are going to creating the structure of Tetrad, which consists of graph with four nodes in which establish three paths among these nodes on the network based on set of collaborative filtering. Let us consider the first node is company that is the source node and fourth node is target node that is applicant. In between the second node is similar node of applicant and third one is similar node of company. The sign of the link (positive or negative) established between these tetrad paths structures and predicting links that should be satisfied whole interaction among the three path structure to recommending the target node of applicant to the source node of the company. Interest and reliability of these nodes (Company and Applicant) on the graph represent the collaborative information among the communities on the network. Temporality of feature is constructed based on the nodes (people) behavior on network structure. Each node on the network has taken unique time value for making the interactions among the nodes. Recency and Activeness of the nodes behaviors are measured on the network. Finally the properties of these features define the reliability of nodes on the network.

Related work

Recent developments in online social networks such as Facebook and Twitter have raised scalability challenges for link prediction. Large scale link prediction was addressed by Acar et al. in[1], where higher-order tensor models based on matrix factorization were used for link prediction in large social networks.

Hasan, Chaoji et al.,(2006)have proposed the link prediction to find the interaction between the nodes in the dataset. The data in different analysis applications such as social networks, web analysis, and collaborative filtering consists of relationships among the communities, which can be considered as links, between people. For instance, two people may be linked to each other if they exchange their taste and their lifestyles.

The aim is to recommend items that match the taste (likes or dislikes) of users in order to assist the active user, addressed by Cain, Bain et al., (2010) traditional collaborative filtering approach. the user who will receive recommendations, to select items from an overwhelming set of choices. Such systems have many uses in purchasing sites, subscription based services and other online applications, where provision of personalized suggestions is required.

Qiu, Yen et al.,(2011) defining temporal metric statistics by combining traditional statistical measures with measures commonly employed in financial analysis and traditional social network analysis. To use time series to describe node behavior, calculate temporal features from the time series to characterize behavior evolution, and use the temporal features to improve link prediction.

In order to facilitate accurate predictions and explore the different factors that drive link creation we explore the use of three feature sets: social, topical and visibility. Rowe, Stankovic et al., (2012) have proposed the Logistic Regression, high entropy model, Random model. Predicting follower edges within a directed social network by graphs and thereby signi?cantly outperforming models.

s

In this work, we consider the composite and mutual link prediction problem. We propose a framework for addressing this link prediction problem for machine learning to understanding the concept of structural feature for learning. Nodes behavior on the network makes the construction of feature for learning based on collaborative filtering. Machine Learning algorithms to measure the accuracy and precision of the link prediction problem for varies applications of the social networks. These predicted measures varies based on the method that we are going to use such as k-N Nearest, decision tree, logistic regression which has constructed on the existing system and Menon, Elkan(2011) have proposed the support vector machine to calculate the better accuracy of link prediction problem.

Existing System

Link prediction is defined as the inference of new interactions among the members of a given social network. Given a directed graph G = (V,E) with a sign (positive or negative) on each Edge, we let s(u, v) denote the sign of the edge (u, v) from node u to node v. That is, s(u, v) = 1 when the sign of (u, v)) is positive, -1 when negative. For different formulations of our task, we suppose that for a particular edge (u, v), the sign s(u, v) is hidden and that we trying to infer it.

Monadic Features, In the social networks, the liveliness of an entity have impact on the behavior of the entity[4]. First defining the monadic features based on the degree of the outgoing edges, as well as degree of incoming edges. An outgoing edge of a node v is an edge that directs from v to another node. The degree of outgoing edges of a node (do (v)) is the number of outgoing edges from that node v. The degree of positive outgoing edges of a node (d+o (v)) is the number of outgoing edges from that node v with positive sign. Similarly we have to apply this to negative outgoing edge d–o (v) and also for degree of incoming edge with positive and negative sign such as d+i (v), d–i (v). The four monadic features d+o (v), d–o (v), d+i (v) and d–i (v) will be used in our method to represent the general attitude of an entity in a network.

Dyadic Features, We also define dyadic features based on collaborative information. We make use of collaborative information for link prediction and extract dyadic features as in collaborative filtering[4]. For example, in company to applicant's recommendation system, a link only exists between a composite pair, i.e. a company type and applicant type. Therefore, we consider a three step path involving both nodes within a future link, which is defined as a tetrad.

A tetrad t(u, sv, su, v) or t(u, v) is a three step path among four different nodes in a graph, where the source node is u (sender) and target node is v (recipient). A tetrad t(u, v) captures a two step relationship across two types, which is the minimum indirect path between a pair of nodes (u, v).

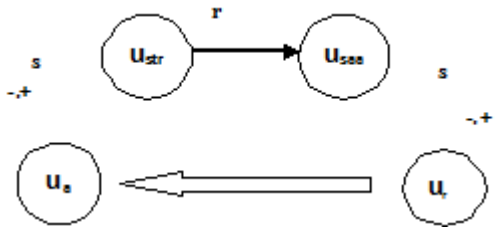

Fig. 2: Tetrad Structure

The typical mutual collaborative filtering for people to people recommendation with preferences of positive edge sign[4, 5] . In the Fig 2, ua is the source node, ur is the target node, ustr is similar node for ur and usaa is similar nodes node for ua . Where S and s mean a link from a precursor to a successor, R and r mean a link is to a precursor from a successor, and their signs, where + means a positive link and ? negative link. To give an example, n(r+s?r+) means the total number of the type of tetrad t(u, sv, su, v) that have a positive link from u to sv, a negative link to sv from su and a positive link from su to v.

Similarly, add one more new type of collaborative information to the sender based inverted collaborative filtering to capture the preference of the recipient as shown in Fig 2. Which allow both positive and negative signs for interaction. Features based on inverted collaborative information are summarized in Table1 connoted by RRS, RSS, SRR and SSR.

The first set of features to capture the recipient's preference and another set of features to capture the sender's preference. We again allow both positive and negative signs in any interaction within the configuration. Features based on preference transmission are connoted by SSS and RRR.

Table 1. Dyadic Features Based on Mutual Collaborative Information.

RSR	SRS	RSS	SRR	SSR	RRS	SSS	RRR
r+s+r+	s+r+s+	r+s+s+	s+r+r+	s+s+r+	r+r+s+	s+s+s+	r+r+r+
r+s+r-	s+r+s-	r+s+s-	s+r+r-	s+s+r-	r+r+s-	s+s+s-	r+r+r-
r+s-r+	s+r-s+	r+s-s+	s+r-r+	s+s-r+	r+r-s+	s+s-s+	r+r-r+
r+s-r-	s+r-s-	r+s-s-	s+r-r-	s+s-r-	r+r-s-	s+s-s-	r+r-r-
r-s+r+	s-r+s+	r-s+s+	s-r+r+	s-s+r+	r-r+s+	s-s+s+	r-r+r+
r-s+r-	s-r+s-	r-s+s-	s-r+r-	s-s+r-	r-r+s-	s-s+s-	r-r+r-
r-s-r+	s-r-s+	r-s-s+	s-r-r+	s-s-r+	r-r-s+	s-s-s+	r-r-r+
r-s-r-	s-r-s-	r-s-s-	s-r-r-	s-s-r-	r-r-s-	s-s-s-	r-r-r-

To Learning and Testing the predict links[4], first calculate the feature values and then calculate a measure of combined feature strength as the weighted combination of feature values, as follows:

$$s = \sum w_i x_i + w_0 \quad \ldots(1)$$

where s is the combined feature strength, xi the value of the ith feature and wi the weight value for xi. To learn the weights and convert this combined feature strength into an edge sign prediction, we use logistic regression, which will output a value in the range of (0, 1) representing the probability of a positive edge sign:

$$p = \frac{1}{1+e^{-s}} \quad \ldots(2)$$

where p is the predicted probability of an positive edge sign. The instances are then classified into positive or negative according to the thresholding of the probability value with respect to a threshold.

Proposed System

Qiu, Yen et al.,(2011) have proposed the temporal feature[5]. Defining temporal metric statistics by combining traditional statistical measures with measures commonly employed in financial analysis and traditional social network analysis. These metrics are estimated over time for a sequence of sociograms. It has shown that some of the temporal extensions of traditional metrics increase the accuracy of link prediction problem.

Temporal features, to use time series to describe node behavior, calculate temporal features from the time series to characterize behavior evolution, and use the temporal features to improve link prediction[5]. Simple Statistics: This type of feature includes simple first-order temporal features such as recency and activeness.

Recency measures the length of time elapsed since a node made its last connection. Activeness measures the number of connections made by a node in the latest time step. Activeness indicates that a node is very active in the last time step and is likely to be active in the future that shown on Fig.3.

Recency and *Activeness* are calculated by making window with 20 time stamp values in fig 3. In this timing window, we are going to calculating the number of occurrences in the entire time stamp for all the nodes. Suppose a particular node with

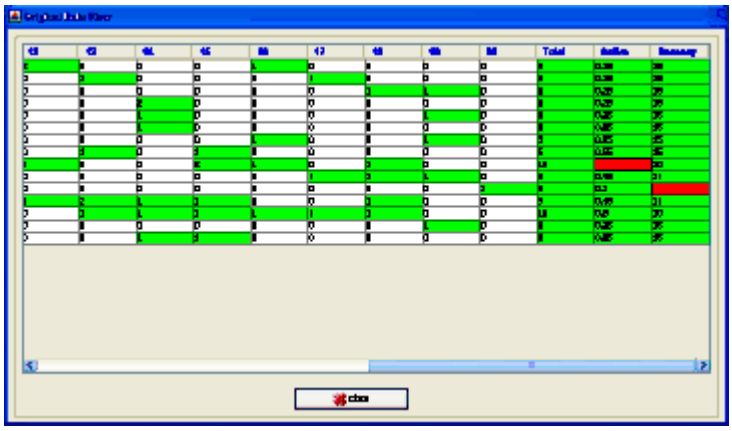

Fig. 3: Activenees and Recency of nodes from the dataset

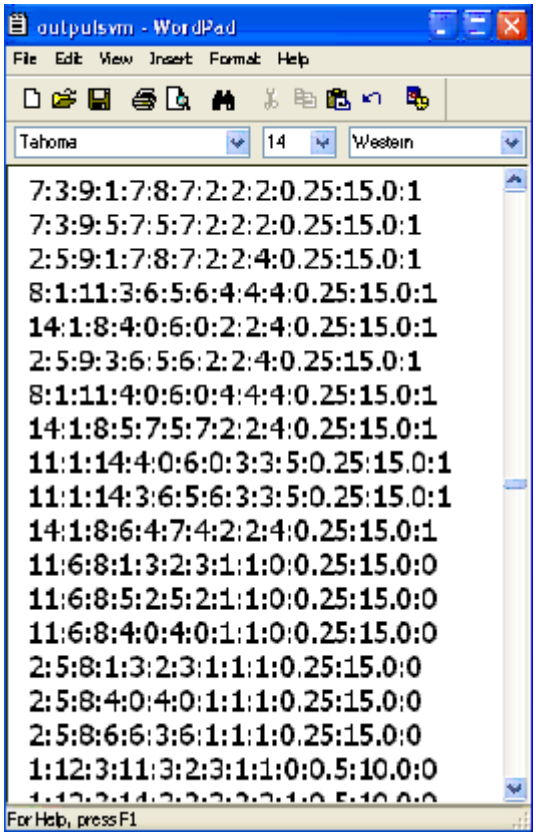

Fig.4 : Construction of features to our algorithms.

the same time stamp values is occur means that will be added in the corresponding time stamp value on the timing window. Then we are going to calculate the total number of occurrences in timing window for each of the node on the data set.

The calculated total occurrences with differentiated with 20 time stamp value for all the node (Recency of nodes) on the data set. The minimum numbers occurrences that mean node with maximum differentiated value is considered as large time inactive node on the network.

Like which. Activeness of the node on the network is calculated by total occurrences of time stamp value is divided by timing window value 20 for the entire node on the dataset. A maximum connection that means node with maximum average value is considered for more active node on the network.

The temporal features to characterize the evolution of node behaviour, and our experimental results suggest that including these temporal features significantly improve link prediction performance.

Support Vector Machine, Menon and Elkan (2011), Support vector machine (SVM) is one of the machine learning algorithms. SVMs are a set of related supervised learning methods used for classification and regression techniques. Given a set of training examples, each considered as belonging to one of two categories, an SVM training algorithm constructs a model that predicts whether a positive link into one category or the other. The algorithm learns a classification model from set of previously labeled (pre-classified) data, and then applies the acquired knowledge to classify the links into two classes: positive links and negative links[6].

Predicted positive links are represented as 1 and the negative links are as 0 on the features construction model from fig 4. In this maximum number of tuples will be considered for the training data that have been given as input into the Support Vector Machine clssifier and other data samples wll be considered for testing samples. Matrix of data points with each row corresponding to a support vector in the normalized data space.

This matrix is a subset of the Training input data matrix, after normalization has been applied according to the 'AutoScale' argument. The sign of the weight is positive for support vectors belonging to the first group, and negative for the second group in Fig 4.

Train a support vector machine, then call the trained machine to classify (predict) new data. In addition, to capture the satisfactory predictive accuracy, we can use different types SVM kernel functions, and we must tune the parameters of the kernel functions. Try different parameters for training, and check via cross validation to predict the best parameters. After obtaining a reasonable initial parameter, we might want to refine our parameters to obtain better accuracy.

RESULTS AND DISCUSSION

The company dataset is used for the required output. The datasets were collected from a commercial social network site containing interactions between users. For the collected data set, evaluation of the logistic regression method gives the accuracy in the testing set. So it is expected to increase the accuracy than the existing method by using support vector machine.

REFERENCES

1. Cai .X, Bain .M, Krzywicki .A, Wobcke .W, Kim Y.S, Compton.P and Mahidadia. A "Learning Collaborative Filtering and Its Application to People to People Recommendation in Social Network", Proceedings of IEEE- International Conference on Data Mining, pp 743-748 (2010).
2. Hasan M.A, Chaoji .V,Salem .S and Zaki .M "Link Prediction using Supervised Learning" , Processing SDM 06 workshop on Link Analysis, Counterterrorism and Security, pp 394-415 (2006).
3. Leskovec .J, Huttenlocher .D and Kleinberg .J "Predicting Positive and Negative Links in Online Social Networks", Proceedings of the 19th International Conference on World Wide Web, pp 641-650 (2010).
4. Cai .X, Bain .M, Krzywicki .A, Wobcke.W and Kim Y.S "Reciprocal and Heterogeneous Link Prediction in Social networks", Proceedings of the 16th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining, Vol.II, pp 193-204 (2012).
5. Qiu.B, He.Q and Yen.J "Evolution of Node Behaviour in Link Prediction", Proceedings of the Twenty-Fifth conference on Artificial Intelligence (2011).
6. Rowe .M, Stankovic .M and Alani.H , "Who will follow whom? Exploiting semantics for link prediction in attention-information networks", Proceedings of the 11th International Semantic Web Conference, pp 476-491 (2012).
7. Menon .A.K and Elkan .C "Link Prediction via Matrix Factorization", Proceeding of the European Conference on Machine Learning and Knowledge Discovery in Databases, Vol. II, pp 437-452 (2011).
8. Ouyang .T.Y "Leveraging Temporal Features for Link Prediction in Communication Networks", Massachusetts Institute of Technology Cambridge (2007).
9. Wang .C, Satuluri .V, Parthasarathy .S "Local Probabilistic Models for Link Prediction", Proceedings of the 2007 Seventh IEEE International Conference on Data Mining, pp 322-331 (2007).
10. Hopcroft .J, Lou .T, Tang .J "Who Will Follow You Back? Mutual Relationship Prediction", Proceedings of the 20th ACM International Conference on Information and Knowledge Management, pp 1137-1146 (2011).