Physica A 460 (2016) 361–373

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Link prediction based on temporal similarity metrics using continuous action set learning automata

Behnaz Moradabadi *, Mohammad Reza Meybodi

Department of Computer Engineering, Amirkabir University of Technology, Tehran, Iran

ARTICLE INFO

Article history:
Received 30 March 2015
Received in revised form 1 February 2016
Available online 13 April 2016

Keywords:
Social network
Link prediction
Similarity metric
Temporal data
Learning automata

ABSTRACT

Link prediction is a social network research area that tries to predict future links using network structure. The main approaches in this area are based on predicting future links using network structure at a specific period, without considering the links behavior through different periods. For example, a common traditional approach in link prediction calculates a chosen similarity metric for each non-connected link and outputs the links with higher similarity scores as the prediction result. In this paper, we propose a new link prediction method based on temporal similarity metrics and Continuous Action set Learning Automata (CALA). The proposed method takes advantage of using different similarity metrics as well as different time periods. In the proposed algorithm, we try to model the link prediction problem as a noisy optimization problem and use a team of CALAs to solve the noisy optimization problem. CALA is a reinforcement based optimization tool which tries to learn the optimal behavior from the environment feedbacks. To determine the importance of different periods and similarity metrics on the prediction result, we define a coefficient for each of different periods and similarity metrics and use a CALA for each coefficient. Each CALA tries to learn the true value of the corresponding coefficient. Final link prediction is obtained from a combination of different similarity metrics in different times based on the obtained coefficients. The link prediction results reported here show satisfactory of the proposed method for some social network data sets.

1. Introduction

Nowadays people and organizations could communicate more efficiently using social networks. Social network can be visualized as graphs, where the nodes correspond to persons and the edges represent some form of communications between the corresponding persons [1,2]. Since this kind of network is generally complex and highly dynamic, it is really important to understand its behavior along time [1,3]. Social Network Analysis (SNA) is a research area that tries to study the dynamics of the network structure [3]. In this paper, our goal is to use dynamics of the network to predict what communications are most likely to be formed in the future. This problem is a well-known problem in SNA area that is called link prediction problem [1].

To deal with link prediction problem, there are several approaches [4]: the most common approach is based on applying topological similarity metrics to non existence links at time t to determine if a link will appear at a time t' ($t' > t$). Such methods generate scores for each link and use the scores to perform prediction task either by an unsupervised or a supervised technique. There are many similarity metrics including local similarity metrics: Common Neighbors [5], Salton

* Corresponding author.
E-mail addresses: moradabadi@aut.edu (B. Moradabadi), mmeybodi@aut.edu (M.R. Meybodi).

http://dx.doi.org/10.1016/j.physa.2016.03.102
0378-4371/

Index [6], Jaccard Index [7], Hub Depressed Index [8], Hub Promoted Index [9], Leicht–Holme–Newman Index (LHN1) [9], Preferential Attachment Index [10], Adamic-Adar Index [11] and Resource Allocation Index [12], global similarity metrics: Katz Index [13], Leicht–Holme–Newman Index (LHN2) [9], Matrix Forest Index (MFI) [14] and Quasi local metrics that do not require global topological information but use more information than local indices: Local Path Index [15], Local Random Walk [16], Superposed Random Walk [16], Average Commute Time [17], Cos+ [18], random walk with restart [19], SimRank [20], Resource Allocation index and Local Path index [15].

The main problem of traditional methods for link prediction is that they cannot explore the network evolution, because they only consider the current network structure without considering the links occurrence time. On the other hand the different similarity metrics have shown different performance on different social networks. Our work tries to overcome these limitations by using CALA to determine the coefficients of different times and similarity metrics. The intuition of the proposed approach is that the final prediction depends on the past similarity scores but this dependency is not uniform between different times and different similarity metrics and it can be differently influenced by difference coefficients. In other words we do link prediction by exploring how different topological metrics in different periods influence the final similarity score.

Continuous Action set Learning Automata (CALA) is an adaptive decision making unit with real-valued actions that tries to learn the optimal action from a set of allowable actions by interacting with a random environment [21]. In each step, it selects an action from its action-set. The action selection in CALA is based on a probability distribution over the action set. The selected action is applied to the environment and then a reinforcement signal is produced by the environment. CALA updates the probability distribution of its actions according to both reinforcement signal and a learning algorithm and again chooses an action. These steps are repeated until CALA converges to some action.

In the proposed method, in order to accomplish link prediction in time $t + 1$, the link prediction problem is modeled as a noisy optimization problem by considering one coefficient for each of similarity metrics and time periods. The goal of optimization problem is to determine the coefficients of the different similarity metrics as well as the coefficients of different past times such that the overall error of the prediction task is minimized. This noisy problem is solved by a team of CALA: one CALA for each coefficient. Each CALA tries to learn the true value of the corresponding coefficient. Final link prediction is done using the scores that is calculated from the past similarity scores with considering times and similarity metrics coefficients. In order to evaluate the result of the proposed approach, we conducted some experiments and compared our results with the ones achieved by other link prediction approaches. In general, the experiments showed that our approach performs better than other strategies.

The rest of the paper is organized as follows. Section 2 introduces the time series link prediction problem and its literature review. The learning automata and continuous action set learning automata are described in Section 3. Section 4 introduces the proposed link prediction method based on CALA. Section 5 presents the experimental study for some social network data sets. Section 6 summarizes the main conclusion and discusses the future directions of our research.

2. Time series link prediction methods

In this section we first introduce the time series link prediction problem and then review the related works in this area.

2.1. Problem formulation

As it is defined in Ref. [22], the time-series link prediction problem is formally introduced as follows: Let V be the list of nodes, $V = \{1, 2, \ldots, N\}$. A graph series is a list of graphs $\{G_1, G_2, \ldots, G_T\}$ corresponding to a list of adjacency matrices $(M_1, M_2, \ldots, M_T)$. Each M_t is an $N \times N$ matrix with each element $M_{t(i,j)}$ corresponds to edge $E_{t(i,j)}$ in G_t. The value of $M_{t(i,j)}$ is from the set $\{0, 1\}$ and it is the indicator of existence or not existence of the edge (i, j) during the period t. Then in the time series link prediction, we try to predict the occurrence or not occurrence of the links in time $T + 1$ using previous times $M_1, M_2, \ldots, M_T$.

2.2. Related works

In the following subsection we review the recent time series link prediction methods:

Ref. [23] has proposed a method for time-aware link prediction. They proposed an extension of the local probabilistic model that is introduced in Ref. [24] by using temporal information. Their proposed method used the link existence and the time of existence to generate a probabilistic model for link prediction. An empirical evaluation of technique was performed over two collaboration networks and they showed that link time occurrence can be considered as a main feature in the prediction result.

The authors of Ref. [25] have formulated the link prediction problem as a periodic temporal link prediction and studied that if the data have underlying periodic. In the proposed method they have introduced two matrix and tensor-based methods for predicting future links and summarized the data of multi time periods into a single matrix using a weight-based method. Then they have used a CANDECOMP/PARAFAC tensor decomposition to illustrate the usefulness of using natural three-dimensional structure of temporal link data and showed the superiority of their method on some bibliometric social networks.