*Forest Chemicals Review*
*www.forestchemicalsreview.com*
*ISSN: 1520-0191*
*July-August 2022 Page No. 1648-1654*
*Article History: Received: 02 April 2022, Revised: 15 April 2022, Accepted: 24 April 2022, Publication: 04 May 2022*

# Link Prediction Method for Social Network

**Yuemei Ren[1,2], Lei Li[1,2*],Qingquan Bian[2]**
[1]Henan Province Engineering Technology Research Center of IIOT, Nanyang, Henan, China
[2]School of Electronic Information Engineering, Henan Polytechnic Institue, Nanyang, Henan, China
*Corresponding Author.

***Abstract:***

Aiming at the problem of low prediction performance due to the lack of negative link feature fusion and mining effective information in symbolic social networks, a new negative link prediction method based on feature fusion is proposed in this paper. Based on the classical structural balance theory and social status theory, this method constructs four characteristics related to negative symbols. Including node characteristics, structure characteristics, similarity characteristics and scoring characteristics, the negative link prediction is realized by using logistic regression algorithm. Its effectiveness is verified on two typical symbolic network data sets, epinions and Slashdot. The experimental results show that compared with the benchmark method, the accuracy of the extracted method is improved by about 4.5% and 10.4% respectively on the two data sets. The F score increased by about 27.3% and 31.5% respectively. The purpose of improving the prediction effect of negative link is achieved..

***Keywords**: Social Network, Link Prediction, Deep Learning, Logistic Regression Algorithm..*

## I. INTRODUCTION

Opportunistic networks evolved from Delay Tolerant Networks (DTN) [1-2]. Different from the traditional MANET, in the opportunistic network, nodes can communicate without a complete link [3]. The source node and destination node can not only deliver data through direct encounter, but also complete the indirect delivery of messages through multi hop routing. Link prediction in the network is to predict whether there are undiscovered links and future links in the network through known node attributes and network node information [4-6]. By predicting the links between network nodes, we can mine the unknown relationships in the network, and then obtain more valuable information. Link prediction is an important task in network analysis. In opportunistic networks, link prediction can provide good direction and theoretical support for solving the difficulties of routing strategy [7]. However, in different networks, the methods of link prediction will also be different. Therefore, we should design a good link prediction method according to the experience of relevant fields and the actual situation of the network [8]. How to design a good link prediction algorithm according to the characteristics of opportunistic network is the focus of this paper.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2022 Page No. 1648-1654
Article History: Received: 02 April 2022, Revised: 15 April 2022, Accepted: 24 April 2022, Publication: 04 May 2022

## II. SIMILARITY INDEX OF OPPORTUNISTIC NETWORK LINK PREDICTION

2.1 Construction of similarity index ASLP

At present, the existing similarity indicators (such as CN, AA, RA, etc.) have good results in networks with relatively large density and more common neighbors. However, the opportunistic network is very sparse, and there may be few common neighbors between the two nodes at a certain time. The similarity index is difficult to describe the structural information of the network, which leads to a certain error in link prediction in opportunistic network. In the research of Zheng et al., it is considered that for sparse networks, the computational complexity will not increase when only the common neighbors and their third-order neighbors are considered. Therefore, this paper selects the local path index as the basic index, because after considering the high-order path, there will be more structure information available in the network and more link information. Therefore, this paper combines this index with the average separation time, and the constructed aslp index is shown in Formula 1 [9-14].

$$S_{a,b}^{ASLP} = \sum_{z\in\Gamma(a)\cap\Gamma(b)} \frac{AS_{az}+AS_{zb}}{2} + \alpha \sum_{m,n\in l_{a\to b}} \frac{AS_{am}+AS_{mn}+AS_{nb}}{3} \tag{1}$$

Where z is the common neighbor of node a and node b, $AS_{az}$ and $AS_{zb}$ are the average separation time of the link, m and n are the third-order neighbors between the two nodes, and $AS_{am}$, $AS_{mn}$ and $AS_{nb}$ are the average separation time of each link in the third-order neighbors. Aslp index improves the original similarity index according to the characteristics of opportunity network. On the one hand, this index can well describe the similarity between network links. On the other hand, it can also expand the local structure information of the network, which establishes a good foundation for constructing deep learning samples later.

2.2 Link prediction and evaluation index

After proposing the index, we need to prove whether the index can accurately describe the link situation through experiments, and verify the effectiveness of the index by comparing with other similarity indexes. AUC (area under ROC curve) and precision are two common indicators to evaluate the quality of link prediction algorithm. AUC is the most widely used evaluation index at present. In the experiment, the independent comparison is not repeated N times. The number of times to record 1 point is n 'times, and the number of times to record 0.5 point is n' times. The calculation formula of AUC is shown in 2.

$$AUC = \frac{n'' + 0.5n'}{n} \tag{2}$$

According to the AUC principle, if the link score is generated randomly, the AUC score is about 0.5. If the AUC value of the algorithm is more than 0.5, it indicates that the algorithm is effective. And the higher the AUC value, the better the index effect. AUC measures the quality of the index as a whole, while

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2022 Page No. 1648-1654
Article History: Received: 02 April 2022, Revised: 15 April 2022, Accepted: 24 April 2022, Publication: 04 May 2022

precision only considers whether the top L edge is predicted accurately. Precision index is the ratio of the number of successful links predicted to the total number of links. The principle is: calculate the similarity index value of each link, then sort it in descending order, and finally check the correct proportion predicted in the first L-bit. The formula is shown in Figure 3.

$$Precision = \frac{m}{L} \tag{3}$$

Where m is the number of successful links predicted in the previous L. The higher the precision value, the higher the prediction accuracy.

The data sets used in this paper are infocom05 and MIT (the details of the data sets are described in Chapter 5), and the comparison of prediction index performance is carried out in the two data sets. In this experiment, CN, AA, RA, Katz and LP indexes are compared with aslp indexes to analyze the performance of these indexes in opportunistic network link prediction. The proposed indicators are compared by AUC and precision. The performance of aslp index in precision is inferior to the best AA index and superior to other similarity indexes. However, the gap between aslp and AA index is not very large, and the difference from other traditional indexes is that aslp is more stable than other indexes with the change of time window. In order to display the results more intuitively, average the calculated AUC and precision. As shown in Table 1.

**TABLE I. AUC and precision mean**

| | **MIT** | | **Infocom05** | |
|---|---|---|---|---|
| | AUC | Precision | AUC | Precision |
| **CN** | 0.7822 | 0.268 | 0.8369 | 0.2819 |
| **AA** | 0.7913 | 0.3131 | 0.839 | 0.3665 |
| **LP** | 0.8035 | 0.2725 | 0.8707 | 0.3254 |
| **Katz** | 0.8039 | 0.241 | 0.8708 | 0.2301 |
| **ASLP** | 0.8287 | 0.2835 | 0.9001 | 0.3485 |

To sum up, the ASLP index proposed in this paper is more effective and stable than the traditional similarity index in the link prediction of opportunistic networks. ASLP index can accurately describe the sparsity and intermittent connectivity of the network. Therefore, this paper uses ASLP as the basic training sample. Through the training of deep learning model, the potential law of link dynamic change is extracted, so as to infer whether the link exists or not at the next moment.

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2022 Page No. 1648-1654
Article History: Received: 02 April 2022, Revised: 15 April 2022, Accepted: 24 April 2022, Publication: 04 May 2022

## III. CONSTRUCTION OF OPPORTUNITY LINK PREDICTION MODEL IN NETWORK

3.1 Cyclic neural network

Although the aslp index proposed in this paper has a good performance in the opportunity network, the problem is that the index only reflects the state in each time window and does not explore the internal relationship between each time window. In other words, it does not study the law of link evolution at any time. Therefore, this paper uses the deep learning model to mine the potential law of opportunistic network links changing with time. In the traditional fully connected neural network, the connection of neurons only exists between layers, and there is no connection among neurons in layers. The performance of neural network with this structure in dealing with timing problem is not satisfactory. Recurrent neural network (RNN) is a deep learning model to deal with time series problems. The model has achieved great success and wide application in natural language processing and various prediction problems. Therefore, this paper uses LSTM in RNN to explore the internal law of opportunity network.

$H^{(t)}$ represents the state of the hidden layer of the model at time t, which is jointly determined by time t and time t-1, $o^{(t)}$ represents the output of the hidden layer at time t, $L^{(t)}$ represents the loss function of the gap between the real value and the predicted value at time t, and U, V and W are the shared parameters in RNN. With the above introduction, the final predicted output value at time t is shown in formulas 4, 5 and 6.

$$h^{(t)} = \sigma\left(Ux^{(t)} + wh^{(t-1)} + b\right) \tag{4}$$

The hidden layer state at time t is determined by the hidden layer state at time t-1 and the input data at time t, where σ is the activation function of RNN and b is the offset parameter of linear relationship.

$$o^{(t)} = Vh^{(t)} + c \tag{5}$$

The formula is the expression of the predicted output value at time t.

$$\hat{y}^{(t)} = \sigma\left(o^{(t)}\right) \tag{6}$$

If it is a classification problem, the □ of this formula is the activation function, which is generally the softmax function. Then the gap between $y^{(t)}$ and $\hat{y}^{(t)}$ can be quantified through the loss function. It can be seen from the formula that the state of the hidden layer in the cyclic neural network is related not only to the input at the current time, but also to the state of the hidden layer at the previous time.

3.2 Short and long term memory network

Although RNN can use historical information and use the back-propagation algorithm over time, the

Forest Chemicals Review
www.forestchemicalsreview.com
ISSN: 1520-0191
July-August 2022 Page No. 1648-1654
Article History: Received: 02 April 2022, Revised: 15 April 2022, Accepted: 24 April 2022, Publication: 04 May 2022

model can not control the transmission of information in practical application. When the time series is too long, there will be the problem of gradient dispersion or gradient explosion. In the process of back propagation of traditional recurrent neural network, when the time series is too long, with the progress of back propagation, the parameters of the weight matrix of the hidden layer close to the output layer update quickly. With the passage of time, its parameters will soon reach the optimal state, while the weight matrix of the hidden layer close to the input layer will update slowly, and the value of the final weight matrix will be smaller and smaller, which is gradient dispersion. On the contrary, if the weight matrix of the hidden layer close to the input layer is very large, the weight of the hidden layer behind it will become larger and larger, resulting in the failure of convergence of the model, which is gradient explosion.

In theory, recurrent neural network can deal with the problem of time series, but when dealing with the situation of long time series, it can not save too long time information. LSTM can be regarded as a cyclic neural network with specific structure, which avoids the problems of gradient disappearance and gradient explosion in standard RNN, and LSTM adds three control units: input gate, output gate and forgetting gate. As the time series samples enter the model, the units in LSTM will judge the information, the information that meets the rules will be recorded, and the information that does not meet the rules will be forgotten. Therefore, it has great advantages in solving the problem of time series. LSTM is composed of multiple units. Each unit has three gates, namely input gate, forget gate and output gate.

Forgetting gate: decide how much information in the previous state should be discarded. Represents sigmoid function, which will output the number between [0,1]. 0 is the cell state before completely discarding, and 1 is the cell state before completely retaining. The calculation formula is as follows:

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{7}$$

Input gate: determines the proportion of input information retained in the cell state Ct. The calculation formula is as follows:

$$\begin{aligned} i_t &= \sigma\left(W_i \cdot [h_{t-1}, x_t] + b_i\right) \\ \tilde{C}_t &= \tanh\left(W_c \cdot [h_{t-1}, x_t] + b_C\right) \\ C_t &= f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \end{aligned} \tag{7}$$

Output gate: the output is based on the updated cell state. Determines the input for the next cell. The calculation formula is as follows:

$$\begin{aligned} o_t &= \sigma\left(W_o \cdot [h_{t-1}, x_t] + b_o\right) \\ h_t &= o_t \cdot \tanh\left(C_t\right) \end{aligned} \tag{7}$$

It can be seen from the above introduction that LSTM model accumulates the information with time series in linear form. And each gate of LSTM model has its own weight and bias. With the progress of model training, the parameters of the model will be adjusted, so as to solve the problem of long-term dependence and overcome the shortcomings of traditional RNN. Therefore, LSTM can effectively deal with the link prediction problem of opportunistic networks.

## V. CONCLUSION

The time variability of opportunistic networks makes link prediction difficult. In this paper, the historical information of nodes on time and space is fully considered, and the similarity index aslp is proposed by combining the frequency, duration and distribution of connections. On this basis, the LSTM deep learning model is applied to the link prediction of opportunistic networks. Aslp index combines time and space dependence, which can better reveal the internal law of link evolution in opportunity network, so as to effectively predict the future link change trend of opportunity network.

## ACKNOWLEDGEMENTS

This research was supported by Science and Technology Key Project of Henan Province (Grant No. 202102311007, No. 222102210128), Key Scientific Research Project of Universities in Henan Province (Grant No. 22A520026), Funding Program for Young Backbone Teachers of Higher Vocational Schools in Henan Province (No. 2020GZGG026), Science and Technology Key Project of Nanyang (No. KJGG020).

## REFERENCES

[1] Zheng Jianguo, Zhu junxuan, Cao ruzhong Research on prediction of social network information dissemination link based on situation. Information theory and practice, 2018, 41 (6): 6
[2] He Jun, Liu Yezheng Research on link prediction of online social networks based on multi-dimensional social relations. Modern intelligence, 2017, 37 (7): 7
[3] Meng Xuying, Zhang Qijia, Zhang Hanwen, et al Personalized privacy protection method for social network link prediction. Computer research and development, 2019, 56 (6): 8
[4] Zheng Wei, pan Qian, Deng Yufan Link prediction method based on common neighbor network centrality in mobile social networks. Computer research (2016, 9, 4): 33
[5] Gu Qiuyang, Ju Chunhua, Wu Gongxing Social network link prediction method based on subgraph evolution and improved ant colony optimization algorithm. Journal of communications, 2020, 41 (12): 15
[6] Li XueGuo, Feng Gang Research on data mining method for social network privacy protection. Science and Technology Bulletin, 2013, 29 (1): 4
[7] Xiong Liyan, Chen Jianghua, Huang Weichun, et al Analysis and Research on a resource service push method for social networks. Science, technology and engineering, 2014 (26): 5

*Forest Chemicals Review*
*www.forestchemicalsreview.com*
*ISSN: 1520-0191*
*July-August 2022 Page No. 1648-1654*
*Article History: Received: 02 April 2022, Revised: 15 April 2022, Accepted: 24 April 2022, Publication: 04 May 2022*

[8] Wang Yao, Kou Yue, Shen Derong, et al Cross social network link prediction based on meta path selection and matrix decomposition. Computer science and exploration, 2019, 13 (9): 12
[9] Jiang Ruoran, Zhang Lingling Recommendation algorithm based on link prediction and node degree in social attribute network. Management review, 2019 (2): 11
[10] Yang Jiahong, Hu Tao, Ren Sha, et al A friend recommendation algorithm based on meta path link prediction mechanism is introduced. Small microcomputer system, 2017, 038 (004): 726-731
[11] Ding Yue, Huang Ling, Wang Changdong Link prediction method based on generative countermeasure network. Computer science and exploration, 2019
[12] Zhu Yuhang, Liu Shuxin, Ji Lixin, et al A sequential link prediction algorithm integrating local topology influence. Journal of electronics and information, 2021, 43:1-13
[13] Li Chunying, Tang Yong, Tang Zhikang, et al Community discovery model for large-scale academic social networks. Computer application, 2015 (09): 157-160 + 165
[14] Shu Jian, Zhang xuepei, Liu Linlan, et al Link prediction method between multiple nodes based on deep convolution neural network. Acta electronica Sinica, 2018, 46 (12): 8