TEST Engineering & Management

July - August 2020
ISSN: 0193-4120 Page No. 4182 - 4187

Construction of the Prediction Model of Traffic Flow by Using Computer Deep Learning Algorithm and Its Application in Logistics Management

Yanqi Zhang[1], Yuqing Fan[1*], Zhigang Song[1], Zhichao Wang[1], Yunjiao Xu[2]
[1]Zhengzhou University of Aeronautics, School of Management Engineering, Zhengzhou 450046, People's Republic of China
[2]Tianjin Chengjian University, School of Economics and Management, Tianjin 300384, People's Republic of China
***Corresponding Author: （Yuqing Fan yqfan@zua.edu.cn）**

Article Info Volume 83
Page Number: 4182 - 4187
Publication Issue:
July - August 2020

Article History
Article Received: *25 April 2020*
Revised: *29 May 2020*
Accepted: *20 June 2020*
Publication: *10 August 2020*

Abstract

Through the deep learning algorithm of computers, the prediction model of traffic flow in the road network is constructed to predict the traffic flow in the network more accurately and efficiently, and solve the "last kilometer" problem in logistics distribution. Firstly, the prediction model of traffic flow is studied by using graph convolution network (GCN) in deep learning algorithm. In the process of prediction, the improved track graph convolutional network (TGCN) model is introduced to better process data in deep layers and predict the complex traffic flow. The efficiency of data processing and prediction accuracy of the TGCN model are analysed by using the traffic flow data and prediction parameters in the actual road network. The mean absolute percentage error (MAPE) of the TGCN model is 4.8%, indicating that the accuracy of the TGCN model is high. Compared with other models of deep learning methods, the TGCN model is proved to have great advantages in the accuracy of data prediction. In the actual logistics transportation and distribution, the prediction model of traffic flow by using deep learning algorithm has better prediction performance.

Keywords: *Deep learning; GCN; prediction model of traffic flow; logistics management; distribution efficiency*

1. INTRODUCTION

With the advent of the big data era of transport, cars are increasing continuously and traffic data are growing explosively, which brings huge pressure to the urban road traffic system. If the current limited transportation space in the city only relies on the construction of transportation infrastructure, it is difficult to alleviate the problems of traffic congestion and frequent traffic accidents. To address this situation, the rapid development and deployment of intelligent transportation system (ITS) play a great role in optimizing the scheduling control process of traffic and improving the traffic efficiency [1]. Through the combination of automatic control technology, data communication technology and sensor technology, ITS can quickly collect, analyze and predict the traffic flow to provide information for people's safe travel, and elevate the control level of traffic management department [2].

The premise of theprecise prediction by ITS is the accurate acquisition of traffic flow information and data. In traditional prediction methods, the global positioning system, radar and social media are mainly used to predict traffic flow, resulting inlow accuracy and efficiency, but the development of sensor technology greatly improves the speed of information collection [3].The common sensors used in traffic system include side looking radar and ultrasonic detector, which can efficiently obtain real-time traffic data and predict road traffic status. However, to carry out long-term prediction of traffic flow, besides relying on ITS and sensor technology, a prediction model of traffic flow[4,5] needs to be established. In recent years, the technical breakthrough of deep learning method has been manifested in the fields of computer vision, speech recognition and natural language processing, which also provides the possibility for traffic flow prediction [6].GCN based on deep learning is an end-to-end learning method, which can effectively identifycompleted images in complex traffic scenes. This abstract method of extracting the characteristics of task objects can accurately detect and predict traffic flow [7,8].

With the rapid development of China's domestic logistics industry, how to realize accurate and efficient distribution with low cost is the current focus of the industry. To avoid the

July - August 2020
ISSN: 0193-4120 Page No. 4182 - 4187

problems of long time and poor safety during distribution, efficient and economic scheduling arrangement for vehicles should be adopted, which can effectively reduce the cost of transportation and human resources of enterprises, save transportation resources, and improve the efficiency of enterprises. The computer deep learning method is used to establish the prediction model of traffic flow, and an algorithm model which can predict the traffic flow with high accuracy is proposed to realize the vehicle positioning, communication, speed measurement and scheduling in logistics distribution, as well as the road flow prediction, vehicle scheduling intelligent, etc. Moreover, the prediction model also has profound practical significance for the research of path optimization in logistics distribution.

2. METHOD

2.1 Principle of traffic flow prediction

Traffic parameters are used in traffic flow prediction, mainly including traffic flow velocity, traffic flow density and the rate of flow. The observation of traffic flow velocity mainly includes average velocity and instantaneous velocity, and the average velocity can be divided into time average velocityand space average velocity [9]. If n is regarded as the number of vehicles on the road section within a period of time, t_i as the driving time of the i thvehicle, and v_i as its instantaneous velocity, the calculation methods of its average time velocity $\overline{v_1}$, average space velocity $\overline{v_2}$, and instantaneous velocity v are as follows:

$$\overline{v_1} = \frac{1}{n}\sum_{i=1}^{n} v_i \quad (1)$$

$$\overline{v_2} = \frac{x}{\frac{1}{n}\sum_{i=1}^{n} t_i} \quad (2)$$

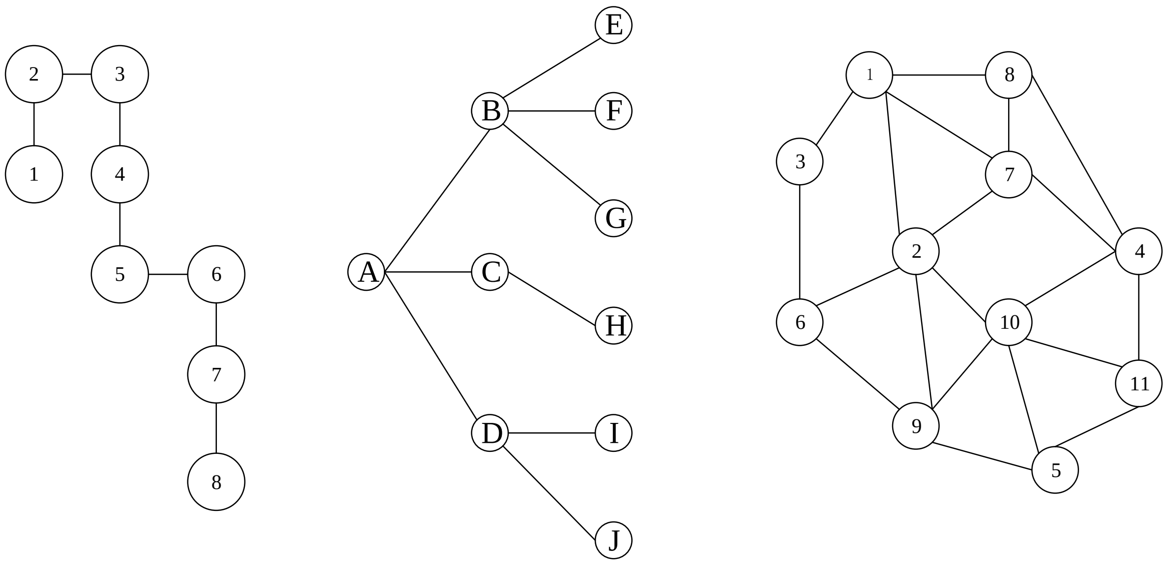
$$v = \frac{\mathrm{d}x}{\mathrm{d}t} = \lim_{t_1 \to 0, t_2 \to 0} \frac{x_2 - x_1}{t_2 - t_1} \quad (3)$$

Traffic flow density reflects the number of vehicles on a certain length of road. Because of its instantaneity, it can reflect the current lane occupancy, which is also known as time occupancy [10,11]. It refers to the ratio of the time spent by vehicles passing through the monitor to the total monitoring time within the observation time *t*. k represents lane occupancy, L is the total monitoring time, t_i is the driving time of the i th vehicle passing a certain distance, and P is the number of vehicles on a certain section of road, and the calculation method of traffic flow density is as follows:

$$k = \frac{1}{L}\sum_{i=1}^{P} t_i \quad (4)$$

2.2 GCN model based on deep learning

Graph is a kind of data structure. It is composed of vertex and edge, which is usually expressed as D = (V, E). G is the identification of the graph, V represents vertex set, E represents edge set, vertex set V is finite and non-empty, edge set E can be empty, and the relationship between any two vertices can be represented by edge [12,13]. Graph has the following common structures:



(a)Linear structure (b)Tree structure (c)Graphic structure

Figure 1. Common structures of the graph

Among the deep learning methods, there are three typical models, namely,convolutional neural network(CNN), deep belief nets (DBN) and stack auto-encoder network model [14,15]. GCN, the optimization method of CNN, is used to establish the prediction model of traffic flow. The concept of GCN was first proposed in 2017. GCN can be divided into spectrum-based convolution (spectral convolution) and space-based convolution (spatial convolution) according to convolution methods. It can implement convolution operation on topology structure of arbitrary graphs and deeply dig target features. Especially in the aspects of node classification and

Published by: The Mattingley Publishing Co., Inc.

July - August 2020
ISSN: 0193-4120 Page No. 4182 - 4187

edge prediction, the prediction performance of GCN on dataset is more accurate than other deep learning methods [16-18].

Both GCN and CNN can use multi-layer structure to extract characteristics of the graph. Suppose that there are N nodes in a set of graph data, and each node has different characteristics. The characteristics of all nodesform a $N\times D$ matrix X, whileall nodes form a $N\times N$ matrix A, which is called adjacency matrix. H is the characteristic of each layer,and the input of GCN in each layerincludes adjacency matrix A and node characteristic H. For the input layer, $H=A$, while for the output layer, $H=Z$. $H^{(l)}\in R^{N\times d^{(l)}}$ represents the node level of the graph at the l thlayer, $D^{(l)}$ is the dimension expressed by the node at the l th layer, σ is the nonlinear activation function, and W is the parameter matrix. Then, the mapping function equation of GCN is shown in (6), and the propagation equation between layers is shown in (7):

$$f\left(H^{(l)},A\right)=\sigma\left(AH^{(l)}W^{(l)}\right) \quad (6)$$

$$H^{(l+1)}=\sigma\left(\hat{D}^{-\frac{1}{2}}\hat{A}\hat{D}^{-\frac{1}{2}}H^{(l)}w^{(l)}\right) \quad (7)$$

2.3 Prediction model of logistics and traffic flow based on GCN

To establish a prediction model of traffic flow and accurately predict traffic flow,it is necessary to get familiar with the network structure of road traffic, and then deal with the traffic flow data by using CGN. Sensors are used to record the corresponding velocity of a specific time period on both sides of the road. Numbers are given to each sensor and the time interval F of data acquisition is set, and then the data are organized into a dataset table. If N sensors are regarded as a sensor network, the relationship between adjacent sensors can be represented by an undirected graph $G=\left(V,E,W\right)$. $|V|=n$, and V represents the set of all nodes, namely, sensor nodes; E represents the set of all edges, namely, the connectivity between sensors; $W\in\mathbb{R}^{n\times n}$ represents a $n\times n$ adjacency matrix, namely, the adjacency matrix of G. In the process of calculating the value of W according to connectivity, when the topological relationship of vertices cannot be estimated, W needs to be constructed according to the distance between sensors.Therefore, the obtained traffic flow data V is defined on G and composed of F data framesin the structure of graph, as shown in Figure 2.

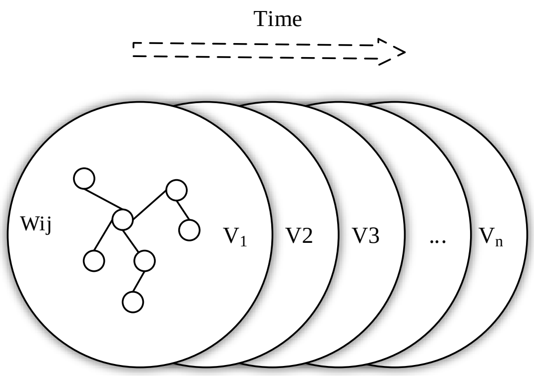


Figure 2. Traffic flow prediction in the structure of graph

Therefore, when predicting the traffic flow at some place, the traffic flow data measured in F time segments can be expressed as follows:

$$V=\left[v_1,v_2,......,v_F\right]^T \quad (8)$$

According to the historical traffic flow data, the prediction of the future traffic flow in this area can be expressed as follows:

$$\hat{v}_{F+T}=argmaxp\left(v_{F+T}\,|\,v_1,v_2,...,v_F,G\right) \quad (9)$$

However, considering the actuality in different regions, the traditional GCN cannot be used to collect traffic flow data. The reasons are as follows. First, different regions have different landforms, terrain undulations and traffic flow directions, so the road traffic networks are complex, which may be in different forms such as ring, radial pattern, and square. To predict traffic flow accurately, multiple factors must be considered. Second, it is easy to collect traffic flow data by using the original GCN method, that is, when the layer is shallow, GCN can obtain the task nodes quickly. However, when the huge traffic flow information needs to be predicted, it is difficult for GCN to distinguish the eigenvectors of the nodes, which will take too much time due to the layer depth. Therefore, to ensure the accuracy and efficiency of data prediction and make it more feasible to obtain the deep structure by increasing the number of layers in GCN, the concept of TGCN is introduced by improving the node layer and memory. In TGCN, the state of the nodes in the upper layer is required to serve in the aggregation and update of the next layer'snodes. The structure of the TGCN model is shown in Figure 3.

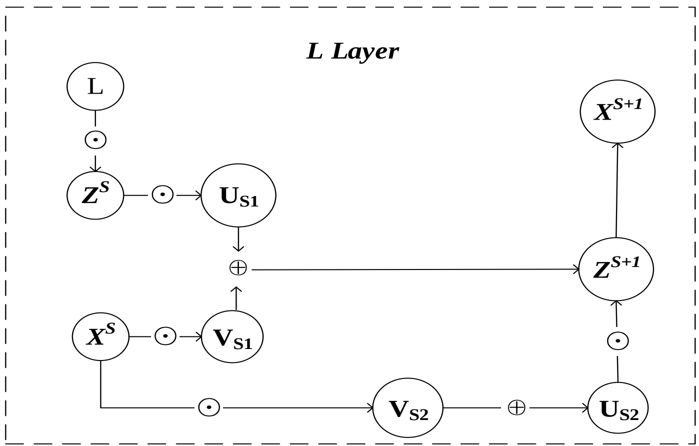


Figure 3. Structure of the TGCN model

Published by: The Mattingley Publishing Co., Inc.

Use "⊕" to represent matrix addition, "⊙" to represent matrix multiplication, and circle to represent operation. To express more conveniently, nodes are divided into two states: X and Z, where X represents normal state and Z represents recessive state. p represents that the two states have the same vector. If n vertices in the graph have these two states respectively, then,for the TGCN in the S th layer, the normal matrix of the node is $X^S \in \mathbb{R}^{n\times p}$, the recessive state matrix is $Z^S \in \mathbb{R}^{n\times p}$, and the adjacency information correlation matrix of the graph is $N \in \mathbb{R}^{n\times n}$. The output new normal and recessive state information are $X^{S+1} \in \mathbb{R}^{n\times p}$ and $Z^{S+1} \in \mathbb{R}^{n\times p}$. f is set as the activation function. $U_{S1} \in \mathbb{R}^{p\times p}$, $U_{S2} \in \mathbb{R}^{p\times p}$, $V_{S1} \in \mathbb{R}^{p\times p}$ and $V_{S2} \in \mathbb{R}^{p\times p}$ are all training parameters. Now, the recessive state of the node is updated according to equation (10):

$$Z^{S+1} = f\left(NZ^S U_{S1} + X^S V_{S1}\right) \quad (10)$$

There is a difference between equation (10)and the GCN calculation equationinmessage passing mechanism, that is, the state vector$p$ of the node is added to the node recessive state. At this time, the update of the node normal state mainly depends on the upper node's state and recessive state, which is as follows:

$$X^{S+1} = f\left(Z^S U_{S2} + X^S V_{S2}\right) \quad (11)$$

The basic TGCN model is established. In practical application, each sensor node can better capture the characteristics of the information flow of road traffic, and the state of the next node is automatically updated after the data pass through the upper convolution layer. The multi-layer TGCN structure is shown in Figure 4, which improves the speed and accuracy of deep prediction and makes up for the shortcomings of the existing GCN in deep-seated calculation.

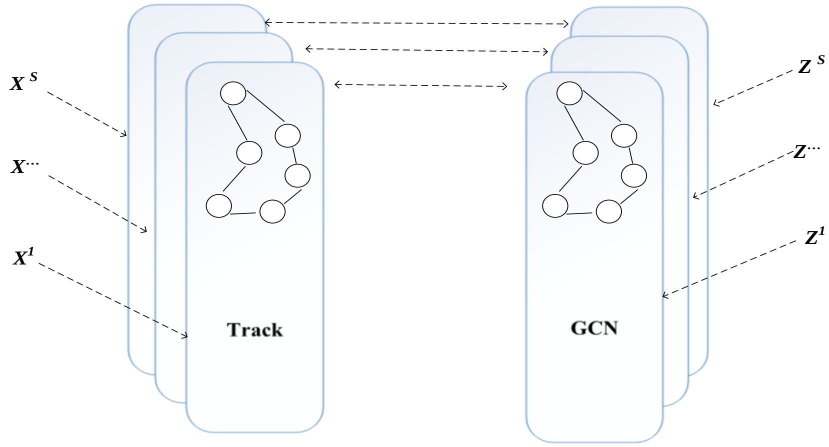


Figure 4. Multi-layer TGCN model

2.4 Experimental settings

(1) Data sources

Since the sensor monitoring data of most of the domestic traffic network is not public, and a large dataset is needed to verify the deep learning method, the traffic network data of California and Seattle in the United States are adopted to test the TGCN model, and the specific datasets are as follows:

①Pems-bay is a road network dataset in California, which is collected by the PeMS system of the California department of transportation, including the flow datasets of 24 sections in the California road network. A total of 61 days of data are collected from May 1, 2018 to June 30, 2018, with the interval of 5 minutes.

② Seattle dataset is collected by 323 sensors in the Seattle road network system. The monitoring data of traffic flow from January 10, 2017 to April 10, 2017 are collected, with the interval of 5 minutes.

③ Guangdong dataset is collected by Foshan Highway Bureau. The monitoring data of traffic flow of all national roads in Foshan, Guangdong province from September 1, 2018 to December 3, 2018 are collected every 5 minutes.

(2) Comparison of evaluation indexes

Common evaluation indexes in traffic flow prediction include mean absolute error (MAE), MAPE and root mean square error (RMSE). These three indexes are used to calculate the error value of prediction, which is then compared with the real value to evaluate the accuracy of TGCN model. If $\hat{n}_i$ is the prediction value of traffic flow in a certain period of time in the future,$n_i$is the real observation value from the road, and m is the number of samples, three kinds of indexes are as follows:

$$MAE\left(x,\hat{x}\right) = \frac{1}{m}\sum_{i=m}^{m}\left|x_i - \hat{x}_i\right| \quad (12)$$

$$MAPE\left(x_i,\hat{x}_i\right) = \frac{1}{m}\sum_{i=m}^{m}\frac{\left|x_i - \hat{x}_i\right|}{\hat{f}_i} \quad (13)$$

$$RMSE\left(x_i,\hat{x}_i\right) = \sqrt{\frac{1}{m}\sum_{i=m}^{m}\left|n_i - \hat{n}_i\right|^2} \quad (14)$$

3. RESULTS AND DISCUSSION

3.1 Comparison of the results of traffic flow prediction by using deep learning algorithm

TGCN model is compared with other machine learning methods and deep learning methods such as autoregressive integrated moving average model (ARIMA), spatial temporal graph convolutional networks (STGCN), GCN and long short term memory (LSTM). The comparison results of predictive performance of each algorithm on PEMs-Bay dataset and Seattle dataset are shown in Table 1.

Published by: The Mattingley Publishing Co., Inc.

TEST Engineering & Management

July - August 2020
ISSN: 0193-4120 Page No. 4182 - 4187

Table 1. Comparison of predictive performance of each algorithm

	PeMS-Bay			Seattle			Guangdong		
Methods	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE
ARIMA	3.89	9.2	8.35	4.69	13.62	11.25	4.46	11.28	10.59
STGCN	3.21	7.04	5.68	4.52	7.81	6.24	3.99	7.61	6.08
GCRNN	3.63	7.62	8.26	3.74	9.13	7.54	3.67	8.85	7.37
LSTM	3.52	9.2	6.28	4.28	9.64	7.89	3.73	9.65	6.88
GCN	2.26	5.46	3.97	4.21	7.81	5.67	3.62	5.56	5.24
TGCN	**2.09**	**4.98**	**3.54**	**3.64**	**6.02**	**5.15**	**3.47**	**5.06**	**4.79**

In Table 1, through the quantitative analysis of the three evaluation indexes, the TGCN model is proved to have better prediction performance compared with the other five models. Among them, the ARIMA Kalman filter method is a kind of regression model, which is widely applied in the prediction of time series. Compared with the deep learning algorithms LATM, DCRNN and GCN, the prediction accuracy of the time series method in early times is lower, and its effectiveness is limited. Because of the uncertainty of traffic flow direction and the complexity of real-time road conditions, time series method cannot deal with these nonlinear variation characteristics.Due to the improvement of deep learning algorithm in long-term prediction, the MAPE values of the prediction results of LSTM, GCN and TGCN models are all less than 7.5%, which is more accurate than the time series method used in early times, proving that GCN based on deep learning has advantages in the modeling of nonlinear traffic flow prediction.In the three datasets, the prediction errors of Seattle dataset are larger than that of PEMs Bay dataset, which suggests that the actual situation of Seattle dataset and Guangdong dataset is more complex in traffic flow prediction. The TGCN model can obtain the structure information of the target better by taking the advantage of the improved deep calculation. Compared with other baseline methods, the TGCN model has outstanding strengths in accuracy.

3.2 Analysis of prediction results of TGCN model in logistics management

To observe the prediction performance of the TGCN model more intuitively, the traffic flow data of PEMs-Bay on July 30, 2018 are adopted for testing again, and MAPE and RMSE are used for quantitative calculation. The MAPE value is 4.8%, suggesting that the mean error is about 5%, and the accuracy of prediction reaches 95%, which is high for the model. The prediction results of the TGCN model are shown in Figure 5. The value of the traffic flow predicted by the model is close to the actual value. Although there are a small number of predictive values with large error, the overall prediction results can meet the requirements of traffic flow prediction in the future.

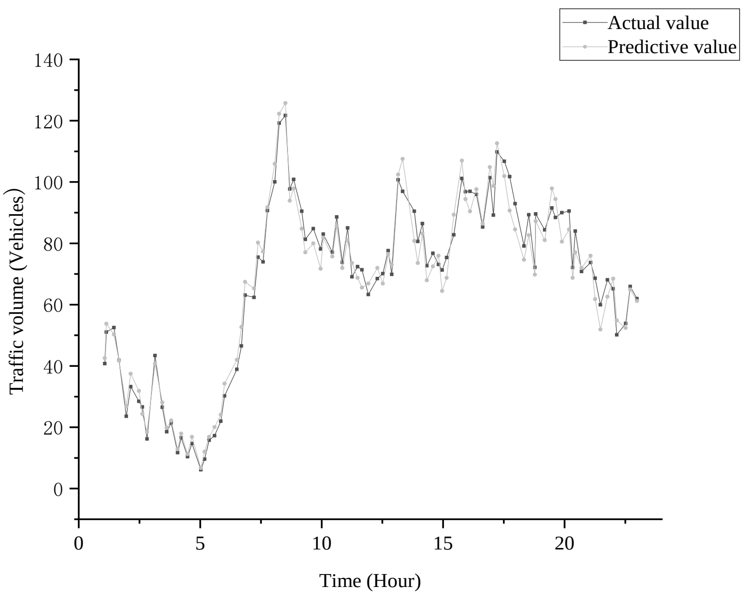


Figure 5. Prediction results of the traffic flow by TGCN

4. CONCLUSION

The whole process of logistics management includes procurement, warehousing, packaging, loading and unloading, processing, transportation and distribution. The prediction model of traffic flow is mainly applied in the terminal link of the logistics process, namely, transportation and distribution. In transportation and distribution, selecting distribution route is easily affected by the weather, accidents, road conditions and other factors. Unreasonable distribution scheme may lead to low efficiency and poor safety. At present, traffic laws and regulations limit the velocity of transport vehicles. In this case, besides changing the mode of transportation,more accurateprediction models of traffic flow are needed to improve the efficiency of goods transportation.

Computer deep learning algorithm is combined with emphasis to evaluate the road status in the process of transportation and distribution. Duringthe establishment of the prediction model of traffic flow, the advantages and development of GCN are described in detail. On the basis of GCN, the optimized TGCN model which can capture the features of objects in deep layers is introduced. Then, extensive actual data are used to test the model, and the advantages and disadvantages of each model are compared and analyzed. Finally, the optimized TGCN model is proved topre-plan the most reasonable transportation routethrough data acquisition, analysis and prediction, avoid traffic jams and roads with special speed limits in advance,ensure that the goods are delivered on time, and achieve the goal of subdividing target characteristics and high-precise prediction. The results demonstrate that this method is effective in logistics distribution.

REFERENCES

1. ChaoL Z,Jia R,lin H J. The simulation model on delay time of road accessibility based on intelligent traffic control system.Cluster Computing, 2019, 22(2), pp. 5919–5929.
2. Satyanarayana M S, Murali Mohan B M, Raghavendra S N. Intelligent Traffic System to Reduce Waiting Time at Traffic Signals for Vehicle Owners. International Journal of Educational Technology in Higher Education, 2018, 3, pp. 281-287.
3. Yang C X, Qing C R. A Review on Traffic Prediction Methods for Intelligent Transportation System in Smart Cities. CISP-BMEI, 2020, 1, pp.152.
4. LeeY. Traffic flow models with looking ahead-behind dynamics, Mathematics, 2019, 3, pp.1903.
5. Pu H L.Progress of intelligent transportation system key technologies.Science and Technology Review, 2019, 6, pp.495.
6. Bianchini, Pierre, Pelletier, et al. Deep Learning in Science. Computer Science, 2020, 10(41), pp. 29.
7. KahngH, Kim SB. Semi-supervised Learning with End-to-end Graph Convolution. Journal of the Korean Society of Management Science and Technology, 2018, 4, pp. 1643-1646.
8. Ting S R, Ding L M, Zhao L. Research on Logistics Distribution Path Optimization Based on PSO and IOT. International Journal of Wavelets, 2019, 17(06), pp. 1950.
9. Yan L X. Traffic flow prediction based on graph neural network. Logistics Sci-Tech, 2019, 3, pp. 58.
10. Lei H,Shao H Y. Short-term traffic flow prediction of road network based on deep learning. Iet Intelligent Transport Systems, 2019, 01(33), pp. 1049.
11. Xiao L. Short-term traffic flow forecast model based on integrated LSTM. Journal of South China University of Technology, 2019, 6, pp. 20.
12. Peng C Y, Rohrbach, Marcus, Cheng Z Y, Yannis, et al. Graph-Based Global Reasoning Networks. IEEE Access, 2019, 10(52), pp. 1109.
13. Zhe Q, Dong G X. Graph Convolution and Self Attention Based Non-maximum Suppression. International Symposium on Neural Networks, 2019, 6, pp. 77-86.
14. Paramita Ray, Amlan Chakrabarti. A Mixed approach of Deep Learning method and Rule-Based method to improve Aspect Level Sentiment Analysis. Applied Computing and Informatics, 2019, 2(2), pp. 1016.
15. Abualshour, Abdulellah. Applications of Graph Convolutional Networks and DeepGNC's in Point Cloud Part Segmentation and Upsampling.Febs Journal,2019, 6, pp. 1033-1035.
16. Jing Y, Hang L Y, Chang Q Z, et al. Modeling Text with Graph Convolutional Network for Cross-Modal Information Retrieval.2018, 9, pp.985-992.
17. Yang L, Fang G Y, Hua G, et al. Probabilistic Graph Convolutional Network via Topology-Constrained Latent Space Model. IEEE Transactions on Cybernetics, 2020, 3(33), pp. 1-14.
18. Jiao X W, Jie M, Nan W N, et al. Deep Learning Model for Point Cloud Classification Based on GraphConvolutional Network Laser and Optoelectronics Progress, 2019, 21(18), 543-545.