ISSN 2616-5767 Vol.4, Issue 4: 30-40, DOI: 10.25236/AJETS.2021.040407

Traffic forecasting and road management based on graph neural networks and big data

Changkun Wang*, Zhichao Qiu, Yu Xiao, Xuxu Dong, Yuqian Wang

School of Jilin University, Changchun 138000, China
**Corresponding author: 1712537966@qq.com*

Abstract: *With the continuous development of China's society and economy, urban traffic problems have become increasingly serious. Analyzing the data gnerated by the city will help us better solve the urban traffic problem. This paper defines a key road in the city, and installs data acquisition equipment on these roads. By understanding the traffic conditions of key roads, we can understand the traffic conditions of other roads in the city. In this way, only a small number of roads need to be monitored in the city, which can reduce the workload of urban data analysis, greatly improve the efficiency of data analysis and reduce the cost of urban construction. This paper combines the common traffic flow prediction scenarios in traffic problems to verify the importance of these key roads. The GPS data of taxis in Changchun City are mainly used for statistical analysis of urban road traffic flow. The main work includes the following aspects: [1] Firstly, the parallel processing method is used to match the taxi GPS data, which greatly improves the matching efficiency. The POI interest point data, road network data and ground induction coil data of Changchun City are crawled. The road flow data is obtained according to GPS data statistics. Grid the city. Road matching is carried out on the ground induction coil data. It is considered that the roads with ground induction coils are the key roads selected by human experience, and these roads are defined as the initialized key roads. Secondly, using the road network map data, the embedded coding features of the road are obtained through the graph embedding series models in the graph neural network method. The embedded features and road attributes are used to cluster the roads, and the key roads based on graph neural network are selected according to the clustering results. Then, the road flow prediction model is mainly divided into full data model and sparse data model. The full data mode uses all road history data, and the sparse data mode uses key road history data. The feature engineering of sparse data model is constructed, which mainly includes five parts: Road inherent attribute characteristics, road network diagram relationship characteristics, POI interest point characteristics, graph embedding characteristics and key road flow characteristics. A variety of machine learning and deep learning methods are tried to construct the road flow prediction model. Finally, the total cost evaluation standard of key road selection results is defined, which is composed of time cost, data cost and precision cost. [2]The time cost is the time taken for model training, the data cost is the proportion of the number of key roads to the total number of roads, and the accuracy cost is the accuracy of the road flow prediction model. The total cost experimental results of full data mode and sparse data mode are compared to verify the feasibility of sparse data mode in display scene. Comparing the key roads of initialization mode selection with the key roads of graph neural network mode selection, it is verified that the key roads of graph neural network mode selection have lower total cost, and can reduce or optimize the existing key roads in the existing cities, so as to achieve the purpose of reducing the cost of urban construction. This paper establishes a traffic prediction model to calculate the flow of other urban roads according to the sparse data of key roads. By reducing the number of monitoring equipment to reduce the cost of urban construction, we can also calculate the flow information of roads without monitoring equipment in the past. Since only road network map data and road attribute data are required for the selection of key roads, this method can also be used for auxiliary design in the planning of new urban roads, so as to reduce the workload and avoid the error of manual operation.*

Keywords: *Figure neural network, big data, sparse data, traffic network, flow prediction, deep learning*

.

ISSN 2616-5767 Vol.4, Issue 4: 30-40, DOI: 10.25236/AJETS.2021.040407

1. Introduction

1.1 Problem background and research significance

1.1.1 Problem background

With the rapid development of China's social economy, China's economy has gradually transformed into a development period dominated by information industry. Urban life has become more intelligent and convenient, and various problems have also arisen [1]. For example, traffic congestion, industrial waste gas pollution, energy distribution and other issues. In the past, due to the bottleneck of technical methods, we were unable to formulate effective strategies to solve these problems, but in recent years, with the development of big data science, we have the ability to face these challenges. Hundreds of millions of data are generated in cities every day. By using the professional technology of computer science to process and analyze these data, find the essential causes of these urban problems, and then formulate plans to solve them. In short, this is the main task of urban computing [2].

The research field of urban computing includes many disciplines, including urban planning and intelligent transportation, among which intelligent transportation is closest to people's daily life [3]. Scientifically and effectively predict the traffic flow of urban roads, which is conducive to people's daily travel and work. Considering the urban construction cost and other reasons, the traffic data of not all roads in the city can be obtained in real time. Therefore, it is necessary to establish a road flow prediction system to infer the flow data of all roads from the flow data of some roads. We define these "partial roads" as the key roads in the city, and how to choose these key roads has become the primary problem to be solved.

1.1.2 Research significance

Thanks to China's modernization and information construction, there are data acquisition equipment on some roads in cities at this stage. Through these data acquisition devices, real-time traffic data can be obtained [4]. These roads with data acquisition equipment can be considered as key roads in the city.

At this stage, most cities mainly use the strategy of installing data acquisition equipment on urban trunk roads. This paper holds that although these trunk roads are important, they may not fully reflect the traffic situation in the city. Moreover, this installation strategy will also produce some problems. Firstly, there are a large number of branch roads, auxiliary roads and other roads in the city without data acquisition equipment, which makes these roads unable to be monitored. Secondly, most trunk roads are connected, and the road flow data connected upstream and downstream often have high similarity, resulting in a waste of resources of data acquisition equipment. Finally, with the increase of the number of urban roads, how to plan and install data acquisition equipment for new roads is a challenge [5].

Therefore, this paper proposes a method for selecting urban key roads based on graph neural network. And establish a set of road flow prediction system to infer the flow data of other roads in the city from the flow data of key roads, and indirectly verify the rationality of the selection of key roads through the accuracy of the road flow prediction system.

Through the road flow prediction system, only the flow data of key roads can be used to predict the flow information of other roads in the city. It can predict the flow information of non main roads such as branch roads and auxiliary roads, which is convenient for us to have a more comprehensive understanding of urban traffic conditions and is conducive to the dispatching and command of urban traffic command center.

The key roads selected by graph neural network method do not need to limit the road attributes such as road length and road grade, which can make the types of key roads in the city more diversified. Moreover, the selected roads are often more scattered in space, so as to avoid the problem of high similarity of collected data and improve resource utilization. Moreover, when using the graph neural network method to select key roads, only the road network structure information is required, which is also applicable to the planning of new urban roads that have not been put into use, so as to avoid the situation that planning can only be carried out through human experience in the past, that is, reducing the workload of labor and reducing the risk of wrong judgment.

ISSN 2616-5767 Vol.4, Issue 4: 30-40, DOI: 10.25236/AJETS.2021.040407

2. Data Processing

2.1 Taxi GPS data matching

Data Description:

The data in this paper comes from the GPS data of 2000 surplus taxis of a taxi company in Changchun from May 1, 2017 to July 29, 2017. The time interval is about 30s. The vehicle identification ID (a) and five fields (b) of time, precision, dimension, passenger status and driving direction are recorded.

2.1.1 GPS data track cleaning

In this paper, the mean filtering algorithm is used to clean GPS data, and the idea of sliding window is used to clean the latitude and longitude field of GPS data_ I and log_ The predicted value is the average of the first n-1 windows $(\tilde{LAT})$ And $(\tilde{log})$。 We first encode the longitude and latitude with 9-bit geohash code, so that the error of the distance data can be controlled at about 5m, which can keep the original information as much as possible, finally, we decode the 9-bit geohash to get the modified latitude and longitude data.

2.1.2 Track matching of GPS data

In this paper, GPS data points are combined into GPS trajectory data by using two modes of no-load state and passenger state. Due to the fact that there are many taxis in no-load state, we should formulate a targeted no-load state merging strategy. After analysis, it is reasonable for us to use 120s as the truncation time of trajectory data, because the signal lamp time in Changchun is generally within 120s, which will not cause the truncation of data due to waiting for the signal lamp, and the GPS position point will not change within 120s, We judge it as a resting or closed state, cut off two tracks, and remove the redundant data in the same position.

2.1.3 GPS data track filling

When we match, we should fill the missing sections in the track results. When the matching strategy is to match GPS track data, we should not only consider which road the individual GPS data points should match, but also whether the roads after the adjacent GPS data points match are connected.

First, we calculate the path that may pass between two data points by Dijkstra shortest path algorithm; Secondly, we consider the road length, road grade and other road attributes. Thirdly, we consider the direction field of GPS data points. Combined with three factors, the GPS track data are matched by using ArcGIS map management and related space calculation module.

2.2 POI data crawling

POI -- point of interest. In geographic data, a POI point can be a shopping mall, a station, etc. Through POI analysis, we can better understand the characteristics and traffic mode of a certain area. This paper crawls Changchun POI data through Baidu map's developer interface and Baidu map API interface.

2.3 Urban grid region division

After analysis, it is found that the influence ability of key roads is limited, and the influence is not strong for the roads that are far away. Therefore, in order to simplify the complexity of the research, the urban road network is divided into grids, which are interconnected within the grid and do not affect each other between the grids. The following figure shows the vector map of Changchun road network:

Figure 1: Vector map of Changchun Road Network

ISSN 2616-5767 Vol.4, Issue 4: 30-40, DOI: 10.25236/AJETS.2021.040407

2.4 Initialize key roads

There is a shock circuit in the ground sense coil equipment. When a vehicle passes, the current will change, and a shock signal is returned to record the flow data. The ground sensing coils in Changchun City are divided into two types: one is the ground sensing coil of common road, with 659 monitoring points in total, and the other is the ground sensing coil of viaduct expressway, with 42 monitoring points in total.

After matching, the ground sense line of Changchun City is as shown in the figure below. The green dot is the common road ground sense coil, and the Red Square is the elevated road ground sense coil. According to statistics, 443 roads have the ground sense coil monitoring points, which are mainly distributed in the fourth ring area of Changchun.

Figure 2: Geographical coil distribution in Changchun City

2.5 Python data processing tools

Multiprocessing module, network x module, Matplotlib module.

3. Key road selection based on graph neural network

3.1 Graph Embedding

In this paper, Node2VEC based on random walk and Graph Sage based on neural network are adopted to learn the Embedding vector of the road.

3.1.1 Node2Vec Model

The NODE2VEC algorithm is improved by DeepWalk algorithm. In order to show the homophily and structural equivalence of the network, NODE2VEC modified the weight of different nodes. Both DFS and BFS control the jump probability between nodes, and the probability is shown in Formula 1

$$\pi_{ab} = \alpha_{pq}(t,b) \cdot w_{ab} \tag{1}$$

3.1.2 Graph SAGE Model

Node2vec can only operate on a fixed network to get the Embedding of a node. So we adopt a Graph Sage model which is suitable for large-scale networks.

Graph Sage represents the Embedding of each node with the Embedding of its neighbor nodes. It trains an aggregation function to quickly retrieve the Embedding of a new node.

3.2 Road Clustering Model Based on K-Means

3.2.1 K-means clustering model

K-means algorithm is also called K-means algorithm [38] the mathematical expression of its cost function is shown in Formula 2:

$$J = \sum_{i=1}^{k} \sum_{j \in c_k} \left(x^{(j)} - u_i\right)^2 \tag{2}$$

ISSN 2616-5767 Vol.4, Issue 4: 30-40, DOI: 10.25236/AJETS.2021.040407

3.2.2 Road clustering method

In this paper, Embedding features and road inherent attributes are used as clustering algorithm features.

First, the road is input, each section is abstracted as a word, and then the vehicle trajectory is used as a sentence to obtain the Embedding of the road and the road grade as the inherent attribute features to select the key road.

3.3 Analysis of road clustering experiment results

Grid 46 is selected for analysis. Its main coverage for the Changchun City Guilin Road and comrade street business circle. The indicators are shown in Figure 3. below.

FID	label	class1	class2	class3	class4	class5	class6	class7
1724	2	12.81565	13.93967	6.5513	17.2583	14.65653	15.21974	14.6325
1725	15	13.23841	14.52009	10.21574	17.74477	14.54122	15.67381	15.463
1726	15	13.89279	15.48316	11.68607	17.91301	14.7961	15.94783	15.29957
1727	15	13.28723	13.99491	11.10801	16.99261	13.97054	15.01363	14.72225
1728	15	13.72325	14.93697	9.94944	17.71241	14.91353	15.83832	15.19356
1729	8	13.95977	14.94197	11.38535	19.20855	15.58582	16.63926	17.65431
1730	8	13.1989	13.80865	10.80094	17.59934	14.22975	14.98895	15.56315
1731	8	13.75561	15.08965	9.0396	18.80262	15.49759	16.53064	16.82566
1732	19	15.31109	15.17442	12.67297	16.83798	13.62155	14.51984	13.49012
1733	19	15.87903	16.01823	13.03406	17.69133	14.5637	15.43378	14.6484
1734	19	16.08042	16.66592	12.85464	18.26286	15.53962	15.96081	15.43362
1736	15	13.00205	13.48121	11.59512	16.46097	13.45267	14.75883	14.15943

Figure 3: Schematic diagram of clustering indexes

3.3.1 Road flow distribution

Figures 4, 5 and 6 below show the results of traffic distribution clustering in categories 8, 14 and 18 on July 25, 2017, respectively.

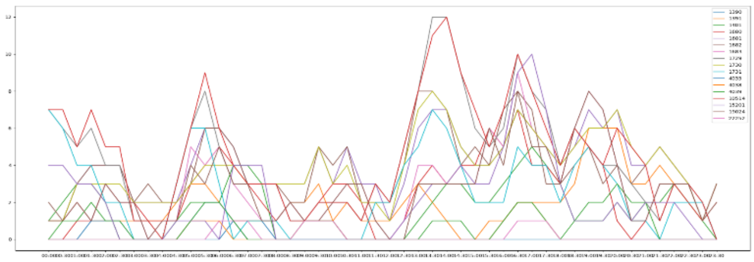


Figure 4: Category 8

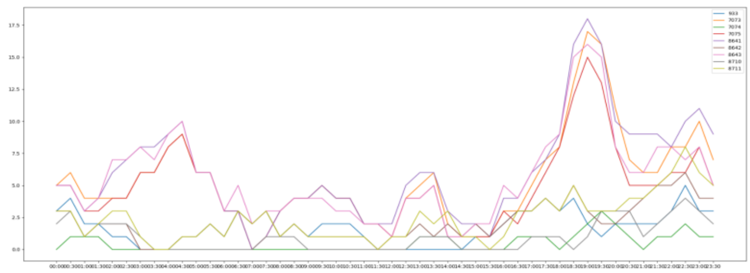


Figure 5: Category 14

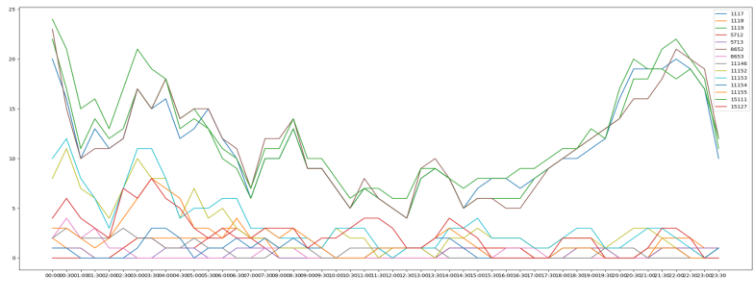


Figure 6: Category 18

ISSN 2616-5767 Vol.4, Issue 4: 30-40, DOI: 10.25236/AJETS.2021.040407

3.3.2 Road spatial distribution

Figure 7 below shows the results of clustering. Different colors are used to distinguish different categories.

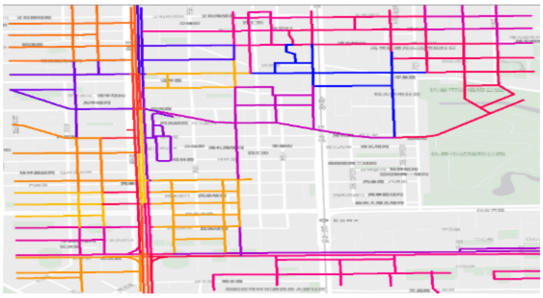


Figure 7: Spatial distribution of roads of various categories

3.3.3 Different methods are selected for critical road distribution

Grid 46 is selected again. The red is the key road and the green is the ordinary road

Fig.9 shows the distribution selected by the method of graph neural network.

Figure 8: Grid 46, through the initialization method

Figure 9: Grid 46, through the graph neural network

3.4 The summary of this chapter

This chapter mainly introduces the definition of the critical road, Graph Embedding method, and how to select the critical road. K-means algorithm is used to carry out cluster analysis on the road, and the key road is obtained. The differences of the key paths chosen by the two methods are also analyzed through a sample.

4. Road flow forecast model

4.1 Full data traffic prediction model

In this paper, the real road traffic in recent time slices is used as the feature of the full data traffic prediction model.ARIMA model and LSTM model are used to forecast road flow with full data.

Autoregressive moving average (ARIMA) model: ARIMA (p, d, q) autoregressive moving average model, where p is the number of autoregressive terms, d is the difference order [39], and q is the number of moving average terms.Is a common time series model.Its mathematical description is shown in Formula 3:

$$\Delta^d y_t = \theta_0 + \sum_{i=1}^{p} \emptyset_i \Delta^d y_{t-1} + \sum_{j=1}^{q} \theta_j \epsilon_{t-1} \tag{3}$$

In this paper, the traffic data of each road in June are used to train the model and predict the traffic data of each road in July. Each model only learns its own characteristics.

Long and Short Term Memory Network (LSTM) model: the structure of LSTM network is shown in Figure 4.1 below, which can solve the dependency problem of long distance [40].

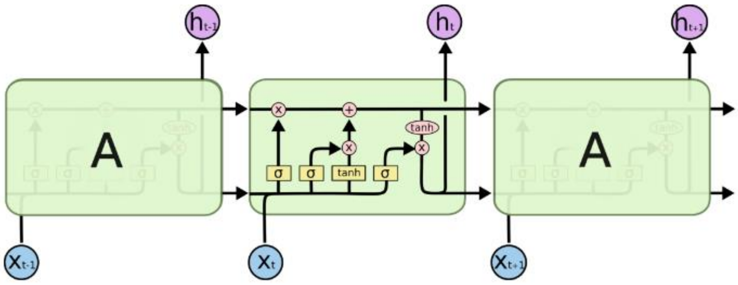


Figure 10: LSTMNetwork structure diagram

The LSTM oblivion gate determines what historical information is discarded.LSTM entry gates 1 and 2 determine which new information is to be retained and updated.Finally, we need to filter and other operations and LSTM output gate to determine what the final output is.It can keep distant historical information, but it can also update the latest data constantly.

4.2 Feature ngineering

Under the system of sparse data traffic prediction model, most of the roads are non-critical roads, and there is no historical traffic information.This paper mainly adopts the following five features to learn the flow law of each road.

Characteristics of road inherent attributes: Based on the road network data provided by ArcGIS, we selected relevant fields that mainly included 9 key fields to represent road properties.

Characteristics of road network graph relation: this feature mainly considers the relation between non-critical roads and critical roads in the network topology of each grid area.Four features are extracted, namely, connected road grade feature, connected road quantity feature, connected road distance feature and connected road hop feature.

Characteristics of POI points of interest: The distribution of POI points of interest can be used as a reference to reflect the social attributes of the surrounding blocks.According to the statistics, the POI data of Changchun contains 17 large categories and 122 small categories.In this paper, the number and density of POI interest points in 17 categories within the diameter range of 100m, 300m and 500m around each road were respectively counted.

Graph Embedding feature: Attempts to represent the implied relationship between roads with a set of low-latitude vectors.In this paper, the Graph Embedding model is used to learn the relationship between traffic transfer between roads, and each road is represented by a set of low-dimensional hidden vectors.

Key road flow characteristics: This paper mainly extracts the statistical characteristics of the historical flow of key roads.It is believed that the number of these features is proportional to the number of key roads.

4.3 Sparse data traffic prediction model

Use a compromise of 30 minutes as a reasonable time slice length.Various machine learning and deep learning methods are used to build the sparse data traffic prediction model.

Polynomial regression (Poly) model: Polynomial regression model includes non-linear relationships by adding high powers of features.Polynomial regression belongs to generalized linear regression

[42].The mathematical expression is shown in Equation 4 below.

$$\Delta^d y_t = \theta_0 + \sum_{i=1}^{p} \emptyset_i \Delta^d y_{t-1} + \sum_{j=1}^{q} \theta_j \epsilon_{t-1} \tag{4}$$

Random Forest Model (RF): The most classic implementation of bagging method in ensemble learning model is Random Forest model [43]. Bagging's random sampling is a sampling method with put back. Each time partial data is used to build the base model, resulting in about 36.8% of the data not participating in the model training, which is called out-of-bag data.This part of data is used as validation data to test the generalization ability of the model.

Gradient lifting decision tree (Light GBM) model (LGB): Light GBM mainly makes the following four improvements to improve its prediction accuracy and model training efficiency, 1.The data structure built by replacing pre-sorted with Histogram algorithm, 2.3. EFB is used to preprocess sparse data. 4.The tree splitting process was decided by the leaf-wise strategy.

Neural Network Model (NN): Firstly, the five-part feature is embedded and its output is fixed to the unified dimension.The Embedding layer is initialized with random weights.Subsequently, three layers of network were constructed in the Dense layer, and the ReLU function was used as the activation function.In order to prevent over-fitting results of the model, the parameter of Dropout layer added after the first two Dense layers is 0.2.Early stop mechanism: After 5 epochs are trained, the MSE index does not decrease, indicating that the model has reached the optimal fitting effect.The network structure of the neural network model in this paper is shown in Figure 4.2 below.

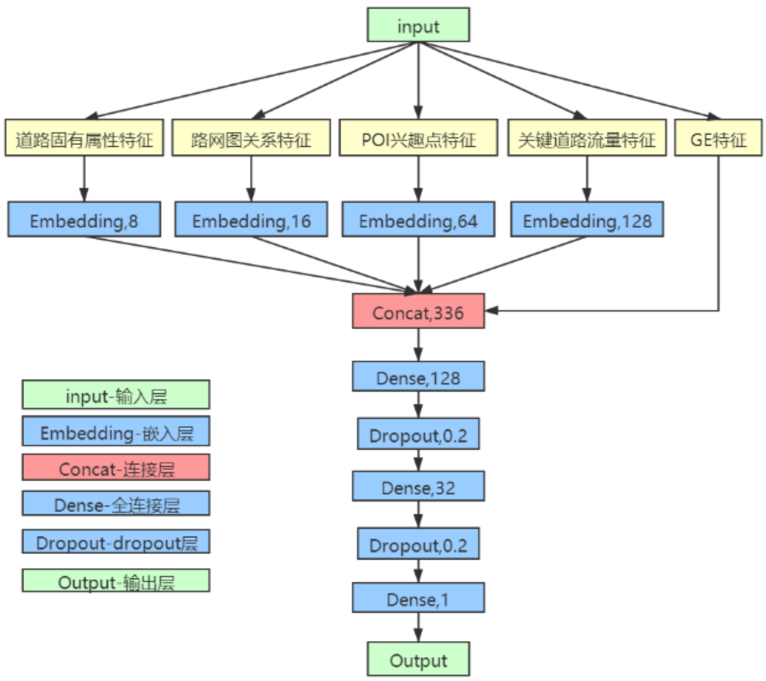


Figure 11: Neural network junction

4.4 Experimental evaluation criteria and Result Analysis of flow prediction model

Evaluation criteria: Since road flow prediction is a standard regression problem, this paper adopts MAE, RMSE (and MAPE as the evaluation indexes of the model, and adds MAPE for joint evaluation.

In this paper, the data of Changchun taxi from May to July 2017 were used.The data of June were used as the training data, the data of July 1, BBB 0 and 15 were used as the verification data, and the data of July 16 to 29 were used as the test data.

Analysis of the experimental results of the full data traffic prediction model: Table 4.1 shows that in most cases, the ARIMA model has better fitting effect than the LSTM model.

The error RMSE can be guaranteed to be between 0.5 and 1.8 when making full data road flow prediction.The total time of ARIMA model training is about 1319 minutes and 25 seconds, and the average cost of a single road is 57.8 seconds.The total training time of the LSTM model was 81 minutes and 32 seconds, and the average cost of a single road was 3.57 seconds.The accuracy of LSTM is close to that of ARIMA, and LSTM is more suitable for application scenarios requiring real-time performance in efficiency.

Analysis of experimental results of sparse data flow prediction model: The prediction errors of different regions in each model are shown in Table 4.2, Table 4.3, and Table 4.4.The proportion of critical roads (known data) in these grids is between 2 and 7 per cent. LGB and RF models can achieve the optimal performance in most grid areas. LGB-V5 takes the average prediction of the five primary models (obtained by data subset training) as the final prediction result.

The LGB (0.29 seconds) and RF (15.3 seconds) with the best performance of the predicted results have a great difference in time efficiency.The RF model is not suitable for the scene requiring real time in this data set.The accuracy of LGB-V5 is only improved by 1%, and the time efficiency is about 8 times that of LGB.Neural network model is slightly inferior to LGB in prediction accuracy, and can be used as a sub-model of model fusion to improve the robustness of the overall prediction results.Polynomial regression model is very efficient in construction, but its accuracy fails to meet the requirements.

Fig. 4.3 is the feature importance graph of the Light GBM model on the grid 65 data set.Among them, Graph Embedding features accounted for 49.3%, key road flow features accounted for 26.5%, PoI points of interest features accounted for 11.9%, road network relationship features accounted for 6.9%, and road inherent property features accounted for 5.4%.

Embedding features and critical road traffic characteristics were used to construct the model.The total contribution rate of feature importance is about 75%, which shows that it is possible to learn the hidden connections between road networks through Embedding features and then interact with the model with key road flow characteristics to achieve a better prediction effect.

5. Experiment results analysis

Our group decided to prove the rationality of key road selection according to the fitness of the traffic flow prediction model. When the prediction results of the sparse data model and the full data model are similar, it proves that the sparse data model can complete the experiment with less data, and it can be considered that it is appropriate to select key roads.

We divided the evaluation criteria for quantifying the accuracy of these two models into three aspects, including data cost, time cost and precision cost. Data cost represents the proportion of roads being selected of all roads. Time cost represents the time-consuming of model training. The accuracy cost measures the error of the model, and RMSE is used as the reference. Since the accuracy cost is the more important one, all these three costs are set with corresponding weights, with time cost accounting for 20%, data cost accounting for 20%, accuracy cost accounting for 60%. The total cost formula is as follows:

$$COST = Time * 20\% + Data * 20\% + Accuracy * 60\%$$

By comparing the cost value, we can compare the key roads, and calculate the cost of sparse data model and full data model. Light GBM model was used to analyze and compare the results.

As the units of these kinds of costs are different, it is necessary to standardize the three kinds of cost data.For the cost of data form I in the number of cost items J, the specific formula is as follows:

$$C_{\mathrm{ij}} = \frac{O_{ij}}{Max(O_{ij})_{i=1,2\ldots m}}$$

The following table is the prediction error table of different critical road proportion in sparse data model. It can be seen from the table that the greater the proportion of key roads have, the more accurate the model prediction is. Due to the lack of data, we can only try to approach the effect of the full data model.

RMSE of flow prediction in different key roads proportion

ISSN 2616-5767 Vol.4, Issue 4: 30-40, DOI: 10.25236/AJETS.2021.040407

Serial number / Proportion	13	43	46	51	65	77
3%	0.5824	2.0675	1.4936	1.6245	2.1845	2.3775
5%	0.5287	2.0121	1.4675	1.5982	2.1425	2.3764
7%	0.5376	2.0456	1.4093	1.4983	2.1334	2.3342
10%	0.5004	1.9703	1.3824	1.5673	2.1354	1.9984
Full Data	0.4367	1.4534	1.1824	1.2394	1.5374	1.7643

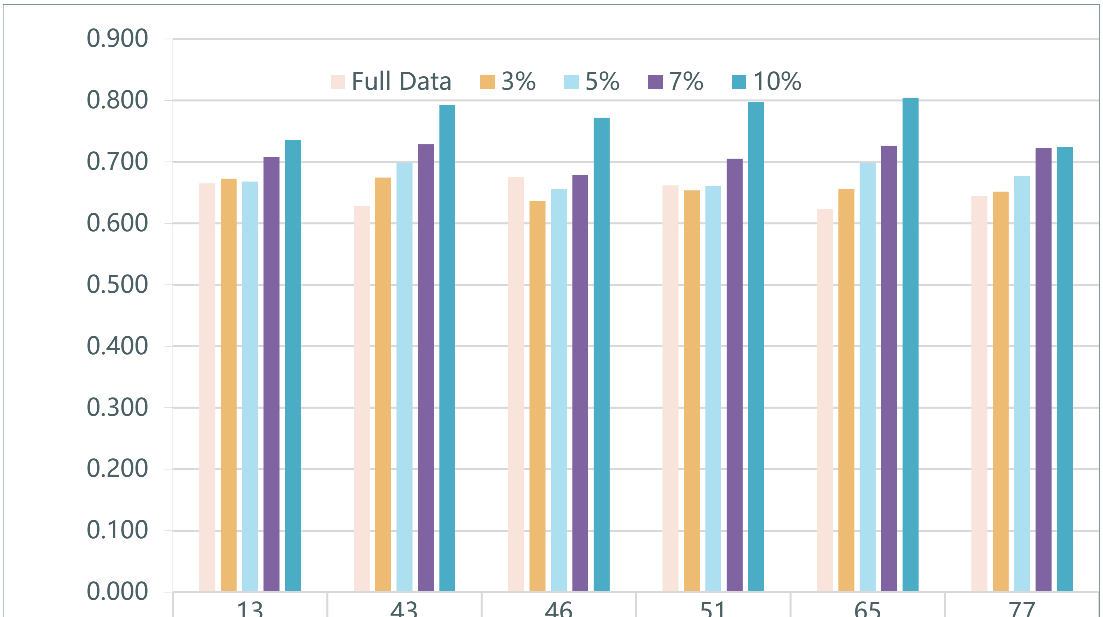
Figure12: Time consuming table of flow prediction in different key roads proportion (seconds)

The following table shows the training duration of different key road proportions in sparse data model. The data dimension of sparse data pattern depends on the number of key roads and the total number of roads in the grid. The full data model uses the recent time slice as a fixed model feature, so its data dimension only depends on the total number of roads in the grid. When the number of key roads increases, the data dimension increases and the training time of the model is relatively prolonged.

Serial number / Proportion	13	43	46	51	65	77
3%	25	53	215	15	161	20
5%	44	79	374	21	316	29
7%	51	88	657	47	394	49
10%	75	149	1378	64	606	84
Full Data	5	5	11	2	7	1

Figure13: Time consuming table of flow prediction in different key roads proportion (seconds)

Figure13 By calculating the formula, we can obtain the total cost. The total cost of different grid areas in different data models is shown in the figure below.



	13	43	46	51	65	77
Full Data	0.665	0.6285	0.6751	0.6617	0.6229	0.6447
3%	0.6727	0.6744	0.6366	0.6536	0.6562	0.6516
5%	0.6679	0.6986	0.6557	0.6604	0.6986	0.6767
7%	0.7083	0.7286	0.6788	0.7053	0.7263	0.7227
10%	0.7355	0.7927	0.7718	0.7972	0.8045	0.7242

Figure14: Cost comparison chart of full data mode and sparse data mode

Through the chart, we can see that although the cost of sparse data model has a certain accuracy gap with the cost of full data model, the sparse data model can save much cost , and the less the proportion of key roads, the more accurate the result is. After comprehensive consideration, the sparse data model

ISSN 2616-5767 Vol.4, Issue 4: 30-40, DOI: 10.25236/AJETS.2021.040407

needs less data acquisition equipment to complete the experiment, which meets the reality and feasibility of urban construction.

References

[1] YuZheng, XiaofangZhou. Computing with Spatial Trajectories [M]. Springer New York, 2011.
[2] An XL, Zhang L, et al. Research on urban public traffic network with multi-weights based on single bus transfer junction [J]. PHYSICA A-STATISTICAL MECHANICS AND ITS APPLICATIONS, 2015, 436: 748-755.
[3] Tao R, Xi YG, et al. Simulation Analysis on Urban Traffic Congestion Propagation Based on Complex Network. JUL 10-12, 2016[C]. IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI):2016.
[4] Osat S, Faqeeh A, et al. Optimal percolation on multiplex networks [J]. NATURE COMMUNICATIONS, 2017, 8: 1540.
[5] Zanin M, Papo D, et al. Combining complex networks and data mining: Why and how [J]. PHYSICS REPORTS-REVIEW SECTION OF PHYSICS LETTERS, 2016, 635: 1-44.
[6] Boccaletti S, Bianconi G, et al. The structure and dynamics of multilayer networks [J]. PHYSICS REPORTS-REVIEW SECTION OF PHYSICS LETTERS, 2014, 544: 1-122.
[7] Pan SJ, Yang QA. A Survey on Transfer Learning [J]. IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, 2010, 22: 1345-1359.
[8] Turnbull Laura, Hutt Marc-Thorsten, et al. Connectivity and complex systems: learning from a multi-disciplinary perspective [J]. Applied network science, 2018, 3: 11.
[9] Colak S, Lima A, et al. Understanding congested travel in urban areas [J]. NATURE COMMUNICATIONS, 2016, 7: 10793.
[10] Yang YJ, Xu YB, et al. Efficient traffic congestion estimation using multiple spatio-temporal properties [J]. NEUROCOMPUTING, 2017, 267: 344-353.
[11] Huang LP, Yang YJ, et al. Mining the Relationship between Spatial Mobility Patterns and POIs [J]. WIRELESS COMMUNICATIONS & MOBILE COMPUTING, 2018, 10: 4392524.
[12] Huang LP, Yang YJ, et al. Comparing Community Detection Algorithms in Transport Networks via Points of Interest [J]. IEEE ACCESS, 2018, 6: 29729-29738.
[13] Du ZW, Yang YJ, et al. Modeling and inferring mobile phone users' negative emotion spreading in social networks [J]. FUTURE GENERATION COMPUTER SYSTEMS-THE INTERNATIONAL JOURNAL OF ESCIENCE.2018, 78: 933-942.
[14] Ma C, Yang YJ, et al. Mobility-based sinknode-aided routing in disaster network under the background of big data [J]. CLUSTER COMPUTING-THE JOURNAL OF NETWORKS SOFTWARE TOOLS AND APPLICATIONS, 2019, 22: 11583-11590.
[15] Yang YJ, Xu YB, et al. Improving Existing Collaborative Filtering Recommendations via Serendipity-Based Algorithm [J]. IEEE TRANSACTIONS ON MULTIMEDIA, 2018, 20: 1888-1900.
[16] Liu ZD, Li ZJ, et al. Mining Road Network Correlation for Traffic Estimation via Compressive Sensing[J].IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, 2016, 17: 1880-1893.
[17] Li YH, Zheng Y, et al. Location Selection for Ambulance Stations: A Data-Driven Approach. NOV 03-06, 2015 [C]. ACM SIGSPATIAL GIS: 2015.
[18] Lv YS, Duan YJ, et al. Traffic Flow Prediction with Big Data: A Deep Learning Approach [J]. IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, 2016, 16: 865-873.
[19] Yi XW, Zhang JB, et al. Deep Distributed Fusion Network for Air Quality Prediction. AUG 19-23, 2018[C]. KDD'18: PROCEEDINGS OF THE 24TH ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY & DATA MINING: 2018.
[20] Shanmukhappa T, Ho IWH, et al. Multi-layer Public Transport Network Analysis. MAY 27-30, 2018[C]. IEEE INTERNATIONAL SYMPOSIUM ON CIRCUITS AND SYSTEMS (ISCAS):2018.
[21] Pink O, Hummel B. A statistical approach to map matching using road network geometry, topology and vehicular motion constraints [C]// International IEEE Conference on Intelligent Transportation Systems. IEEE, 2008:862-867.
[22] Sharma K P, Pooniaa R C, Sunda S. Map matching algorithm: curve simplification for Frechet distance computing and precise navigation on road network using RTKLIB [J]. Cluster Computing, 2018(2): 1-9.
[23] Aleta A, Meloni S, Moreno Y. A Multilayer perspective for the analysis of urban transportation systems [J]. Scientific Reports, 2017, 7:44359.
[24] Liu X, Gong L, Gong Y, et al. Revealing travel patterns and city structure with taxi trip data [J]. Journal of Transport Geography, 2015, 43: 78-90.