

Traffic jam prediction based on analysis of residents spatial activities

Zhijin Lv

College of Software, Jilin University
Key Laboratory of Symbolic Computation and Knowledge
Engineering of Ministry of Education
Changchun 130012
lvzj2020@foxmail.com

Wei Tang

College of Software, Jilin University
Key Laboratory of Symbolic Computation and Knowledge
Engineering of Ministry of Education
Changchun 130012

Hao Fu

College of Software, Jilin University
Key Laboratory of Symbolic Computation and Knowledge
Engineering of Ministry of Education
Changchun 130012

Xiaoxu Chen

College of Software, Jilin University
Key Laboratory of Symbolic Computation and Knowledge
Engineering of Ministry of Education
Changchun 130012

Abstract—The prediction of urban traffic congestion has always been one of the important contents in the research of intelligent transportation systems. The difficulty in predicting urban traffic congestion is that urban traffic operation is essentially a collection of spatial activity planning for residents under certain conditions. The huge group of residents themselves has great complexity and uncertainty. Traditional neural networks mostly focus on road characteristic data and road condition data, and lack of in-depth exploration of the fundamental factors of traffic congestion for residents' travel. So we propose a traffic congestion prediction model based on the analysis of residents' spatial activities. Starting from the residents' activities, the simulation of urban traffic operation conditions can more realistically reflect the traffic congestion situation of the city at specific times and roads, and quickly generate model results. The experimental results show that the model analyzed by residents' spatial activities runs fast and has high prediction accuracy. The results are in line with the actual situation and have strong practical value.

Keywords-component: *Traffic congestion, Raster data, Resident travel, Big data applications, Multinomial Logistic Regression Model*

I. Introduction

With the continuous progress of China's urbanization, the urban population and vehicles have surged in recent years. The traffic problems in high-density cities like Beijing have attracted more and more attention. The increasing traffic congestion and the subsequent decline in residents' travel efficiency and frequent traffic accidents have reflected the city's saturation of traffic space is getting higher and higher. Therefore, intelligent transportation has become a focus of attention and research in the field of transportation. With the gradual coverage of big data on the Internet, more and more data can be used to support research on traffic prediction. In this area, there are many prediction models and methods, such

as Kalman filtering, time series, historical average, linear regression, non-parametric models, etc^[1-7].

Because neural networks have the characteristics of learning complex non-linear systems, many models in the field of traffic prediction are currently based on neural network prediction techniques. Smith and others first applied back-propagation neural networks to short-term traffic flow prediction in 1994. The neural network models currently applied in the field of traffic prediction include multilayer feedback neural networks, radial basis RBF neural networks, backpropagation algorithms, recurrent neural networks, time-delay neural network models, and so on.

However, neural networks still have disadvantages that cannot be ignored, such as high accuracy of data requirements, large data input volume, slow training speed, and easy to fall into local optimum. Lack of adaptability to vehicle congestion in local areas and peak hours. There are still shortcomings in solving practical problems, so that neural networks have not been applied on a large scale in traffic prediction, and their role in solving traffic congestion and congestion is very limited.

Urban traffic flow arises from the daily travel needs of the majority of residents and is related to the choice of each resident's travel mode and the specific destination. Therefore, from the perspective of residents, studying their travel planning under the specific spatial conditions of the city, and studying road traffic issues from the micro to the macro, provides us with another plan to predict traffic congestion^[8]. With the maturity and application of modern information technologies such as the Internet and big data, we can quickly obtain a large amount of data on the spatial activities of residents and data on real-time traffic congestion. Therefore, this research method not only ensures the reliability of the research but also reduces the difficulty of data collection and

improves the model calculation speed. It is very suitable for studying traffic congestion at specific times and roads.

Based on the urban planning data and residents distribution data of the Second Ring Road in Beijing, this paper uses the big data such as POI points, APP evaluation, and check-in to study the characteristics of residents' spatial activities, and proposes a traffic congestion prediction model based on the analysis of residents' spatial activities^[9]. To more easily quantitatively study the activities of residents in cities, we will rasterize all the data through a GIS system, and each grid data independently calculates its interaction with the surrounding grid.

II. Analysis of residential space activities

A. The urban residents' mobility model

To more realistically simulate the activities of the residents in the city, we divided the residents into two categories: "the aimless crowd" and "the purposeful crowd". The former focuses on simulating the free flow of people in population-intensive areas, while the latter focuses more on simulating people with clear destinations and arrival methods. The first part of the urban resident flow model interacts by computing each raster data with neighboring data.

Through the GIS system, we can rasterize all geographic information data. So we can more intuitively see the data changes of various regions and roads in the Second Ring Road of Beijing. And this makes data interaction between specific areas more convenient. In the model, we treat each raster data as a cell and identify the raster data covered with the target road as a road cell. In each iteration, the cell interacts with neighboring cells.

The characteristic of the cell is reflected in its ability to attract surrounding people. It is mainly affected by the number of POI points, population, and regional popularity in the area.

For each cell:

$$u = \frac{p}{c} \times (1 + 0.1 \times t) \quad (1)$$

Among them, u represents the cell's ability to attract the urban population, p represents the cell's POI points, c represents the number of people contained in the cell, and t represents the regional popularity of the cell.

Although the flow of people in cities is very random and disorderly. The population between each cell is macroscopically expressed as a certain number of inflows and outflows per unit time. And there is no doubt that this is mainly affected by the attraction between cells.

For the interaction between two cells:

$$c_{a-b} = c_a \times k \times (1 + u_b - u_a) \quad (2)$$

Among them, c_{a-b} represents the population flowing from a to b in this iteration, c_a represent the current number of people in a , and k represents the population's free spread ratio.

In consideration of accuracy and computational performance, in this study, Moorish neighborhoods were

selected to calculate the interactions between cells, and the cell size was set to 100m×100m. Each iteration will update the population and regional popularity of the cell.

B. Travel model of urban residents

Based on the urban resident flow model, we simulated the purposeful trips that residents in each cell may produce. Since the car dominates the traffic flow in the city center, we will focus on analyzing private car travel for residents. To more realistically simulate the occurrence of resident travel events, we established a resident Agent simulation model. Agent represents a kind of intelligent subject, it is an active object, with autonomy, reactivity, and initiative. That is, each resident Agent decides the travel destination and travel mode independently and is not affected by other factors.

Resident Agents, as the basic unit for autonomously selecting transportation modes, usually choose the most effective travel scheme among many transportation modes. Therefore, we use the Logit model to simulate the travel modes of residents. Under certain conditions and locations, the factors that affect the choice of residents' travel modes mainly include gender, age, income, occupation, car ownership, and bus change number of stops. SPSS's principal component analysis method was used to select the characteristic variables, and finally, the influencing factors were determined as follows: car ownership, travel distance, and the number of transfer stations. So we can get:

$$p = \frac{e^{0.345X_1 + 0.224X_2 + 0.321X_3 - 2.452}}{1 + e^{0.345X_1 + 0.224X_2 + 0.321X_3 - 2.452}} \quad (3)$$

Among them, p represents the probability of residents choosing the car, and X_1 , X_2 , and X_3 respectively represent the car ownership, travel distance, and the number of transfer stations of individual residents.

So based on the urban residents' mobility model, we can get the population travel data at a specific moment and in a specific area. The collection of travel data generated by all cells in a unit time can be reflected in the traffic conditions in the second ring of Beijing.

Combining the residents' flow and residents' travel model, we can simulate all the residents' spatial activities in Beijing Second Ring Road. The next step is to get the traffic conditions of the road cells.

III. Short-term congestion prediction based on analysis of residents' spatial activities

A. Experiment

This article uses the real-time traffic congestion data of Beijing's specific roads published on the AMAP Big Data platform. The morning and evening peak data of 6:00 to 9:00 and 17:00 to 20:00 are selected in one day and obtained every five minutes. Traffic congestion is a contradiction between transportation supply and demand caused by the increase in motor vehicles, insufficient transportation supply, etc. The specific manifestations are the phenomenon of vehicle congestion, slow driving, and delays compared to smooth traffic conditions. A higher traffic congestion delay index indicates a larger proportion of travel

delay as a proportion of travel time, and also indicates that traffic congestion is more severe.

The purpose of this article is to use the residents' spatial activity analysis model to predict traffic congestion at specific periods and roads within the Beijing Second Ring Road. The road monitoring data, traffic congestion index evaluation, road network and other information accumulated every day by AMAP Big Data Platform provide solid data support for the prediction of traffic congestion index.

In each iterative calculation, various types of data of the urban residents' mobility model will be used as the input of the travel model of urban residents. And the resident travel data generated by the travel model of urban residents will be reflected in the urban residents' mobility model as the result of this iterative calculation.

Combining the urban residents' mobility model with the urban residents' travel model, in each iterative operation, each cell will generate several residents' travel intentions. For each travel intention, the shortest road is selected, and the corresponding road cells are added corresponding traffic flow. As an independent individual in the overall model, each road cell only realizes the prediction of a specific road segment, and its functions are limited; all the cells are gathered together to form a huge complex network that can predict the entire road network. In the experiment, we compared the prediction result of the test sample with its actual congestion situation. If the actual congestion situation is consistent with the predicted situation, the road segment prediction can be considered accurate, otherwise, it is considered inaccurate. The cells of each road are judged independently, and the statistical accuracy is finally synthesized.

B. Comparison of residents' spatial activity analysis with other algorithms

We compare the prediction algorithm based on the analysis of residential space activity with the classic BP neural network and logistic regression algorithm under the same data situation. The main comparison objects are prediction accuracy and model calculation time. The experimental results are shown in the table.

TABLE I. COMPARISON OF DIFFERENT MACHINE LEARNING ALGORITHMS

Algorithm	Training time/s	Accuracy/%
ARSA	3.26	79.82
LR	0.96	74.15
Bp-network	236.71	63.54

As can be seen from the table, due to the use of the logistic regression algorithm in the basic algorithm of the model, the running time will increase slightly. But comparing with BP neural network, not only greatly reduces the model operation time, but also improves the prediction accuracy to a certain extent. Therefore, the prediction algorithm based on the analysis of the residents' spatial activities, while ensuring the accuracy, greatly simplifies the model calculation process, avoids a large amount of data calculation, and has a fast calculation speed.

Logistic regression algorithms are commonly used, linear models. Logistic regression algorithms are generally considered to be neural networks without hidden layers. Due to the fast calculation speed, these linear models are widely used in large data scenarios such as statistical analysis, calculating advertisements, search rankings and so on. In this data set, its training speed is so fast that it only takes a few seconds, but its prediction accuracy is very different from this model. At the same training time, its accuracy is significantly lower than the model proposed in this paper. From the comparative analysis of the model and the linear model, it can be seen that the algorithm based on the spatial activity of the residents not only has the training speed of the linear model but also can guarantee the approximation ability of the neural network.

It is worth mentioning that we also tried to run the Adaboost-SVM on the same experimental data set, but the experiment still failed to output complete and valid results after 2080 minutes of running. This shows that our model is not only faster than Adaboost-SVM but also more suitable for applications on large data sets.

IV. Conclusion

We elaborated and analyzed the algorithm of resident space activity analysis. Inspired by resident travel research, we proposed a traffic congestion prediction model based on resident space activity analysis. This model uses a GIS system to rasterize the research data. Each raster data is considered as an independent cell, and it is only responsible for the residents' movement and residents' simulation operations in its area. The interaction data between each cell will be reflected in the road cell involved, and the result of the operation can be integrated to predict the complex urban road network. Since it is not just a simple linear regression of data, our model can obtain higher prediction accuracy than the logistic regression algorithm. At the same time, each cell is independent of each other, suitable for parallel operations, and can meet large-scale data operations.

The experimental results show that the prediction method based on the analysis of the residents' spatial activities is highly accurate in predicting specific periods and roads, and the prediction results are in line with actual conditions, have high reliability, and have certain practical value. Compared with other algorithms using only default values, the proposed algorithm has better results. This article does not show the comparison results after calculation by other algorithms. In the future, we are going to further compare with other algorithms to prove whether they still have better results and make the experimental results more convincing.

Reference

- [1] Zhanquan Sun, Geoffrey Fox. Traffic Flow Forecasting Based on Combination of Multidimensional Scaling and SVM[J]. International Journal of Intelligent Transportation Systems Research,2014,12(1).
- [2] Monoj K. Pradhan, Sonajharia Minz, Vimal K. Shrivastava. A Kernel-Based Extreme Learning Machine Framework for Classification of Hyperspectral Images Using Active Learning[J]. Journal of the Indian Society of Remote Sensing,2019,47(10).
- [3] Chuan Luo, Chi Huang, Jinde Cao, Jianquan Lu, Wei Huang, Jianhua Guo, Yun Wei. Short-Term Traffic Flow Prediction Based on Least

- Square Support Vector Machine with Hybrid Optimization Algorithm[J].
Neural Processing Letters,2019,50(5).
- [4] Jieni XUE, Zhongke SHI. Short-Time Traffic Flow Prediction Based on Chaos Time Series Theory[J]. Journal of Transportation Systems Engineering and Information Technology,2008,8(5).
- [5] Feng Jin, Baicheng Zhao. Short-Term Traffic Flow Prediction Based on Road Network Topology[J].Journal of Beijing Institute of Technology,2019,28(03):383-388.
- [6] Huang Guang-Bin, Zhou Hongming, Ding Xiaojian, Zhang Rui. Extreme learning machine for regression and multiclass classification.[J]. IEEE transactions on systems, man, and cybernetics. Part B, Cybernetics: a publication of the IEEE Systems, Man, and Cybernetics Society,2012,42(2).
- [7] Azadeh Emami, Majid Sarvi, Saeed Asadi Bagloee. Using Kalman filter algorithm for short-term traffic flow prediction in a connected vehicle environment[J]. Journal of Modern Transportation,2019,27(3).
- [8] Yuqi Liu, Yongping Zhang, Scarlett Ting Jin, Ye Liu. Spatial pattern of leisure activities among residents in Beijing, China: Exploring the impacts of urban environment[J]. Sustainable Cities and Society,2020,52.
- [9] DAI Dandan, ZHOU Chunshan, YE Changdong. Spatial-temporal Characteristics and Factors Influencing Commuting Activities of Middle-class Residents in Guangzhou City, China[J].Chinese Geographical Science,2016,26(03):410-428.