Recurrent city traffic congestion propagation analysis

Xing Ji¹, Jing Sheng Wang¹*, Tong Mei Zhou¹, Peng Jun Chen¹

¹ School of Traffic & Management, People's Public Security University of China, Beijing 102623, China
Email:243263495@qq.com

Abstract. In order to analyze the traffic congestion of urban road network, the data preprocessing method is improved, and the three-dimensional spatio-temporal state data is transformed into one-dimensional data, and the spatio-temporal constraint is added on the basis of FP-Growth algorithm to achieve more efficient mining congestion propagation law. The purpose of this paper is to study the traffic congestion propagation law in the morning and evening peaks of the working day and all day and rest days through two experiments, and obtain the traffic congestion propagation law in working days and rest days. The research results show that using the urban road network traffic state data, the analysis method of this paper can more effectively mine the propagation law of recurrent traffic congestion.

1. Introduction

Traffic congestion has been plaguing people's daily travel. Over the years, countries have adopted a variety of methods, such as road construction, fee management, traffic control, network balance, vehicle purchase restrictions, etc., although the traffic congestion has been alleviated to a certain extent, but congestion is still serious. At present, the impact of recurrent traffic congestion is gradually increasing, which seriously affects road traffic efficiency; research on recurrent traffic congestion is one of the focuses of road traffic management.

In view of the law of recurrent traffic congestion, the previous research is mainly based on mathematical statistics and simulation. In terms of mathematical statistics, Morales [1] started from the spatial extent of congestion diffusion, and analyzed the sudden congestion queue length and total delay caused by the arrival rate and departure rate curve estimation, provided that the road arrival rate and departure rate were unchanged. Nam et al. [2] used a risk model to estimate the duration of congestion after a traffic incident, and analyzed the volatility and reliability of the congestion duration. In terms of simulation, Chen et al. [3] integrated the concept of thermodynamic dissipation and aggregation into cellular automata simulation to analyze the law of urban traffic congestion dissipation and aggregation. Zhang [4] constructed the urban road network model based on the complex theory, adjusted the cellular transmission model according to the traffic characteristics, simulated and analyzed the propagation process of traffic flow in the road network under different traffic demands, and then obtained the temporal and spatial characteristics and propagation law of traffic network congestion. Wei [5] used speed as the basic quantitative index of traffic state, divided the traffic state, and improved the SDM model. Based on the measured data, he put forward the analysis method of Traffic Congestion Propagation characteristics. Zhang et al. [6] improved SIS models were used to analyze the influence of propagation time on Congestion Propagation under different traffic conditions.
However, under the increasingly complex background of urban road traffic system, it is difficult to achieve ideal results only through the location detection and processing measures of traffic incidents. For the analysis of single section, the influence between adjacent roads is not fully considered. In view of the research on regional road network, the relationship between roads in road network is not considered. Based on the method of traffic simulation, many ideal hypotheses are added, and it is difficult to accurately describe complex urban road traffic systems. At the same time, in the process of urban informatization, a large amount of traffic state data has been accumulated, and the law of traffic congestion is hidden behind the data. How to use the massive traffic data to more deeply explore the propagation mechanism of traffic congestion is even more important.

This paper will make use of the traffic state data of urban road network and apply the improved algorithm to analyze the mechanism and process of traffic congestion propagation.

2. Improved data preprocessing method

To analyze and study the traffic congestion propagation law, it is necessary to obtain the traffic data of the urban road network, including ID attribute, link name attribute, time period attribute and traffic status attribute, and process the original data within 5 minutes to obtain a complete day’s data. As shown in Table 1.

| ID | ROAD | TIME | STATUE |
|----|------|------|--------|
| 1  | 1    | $t_1$ | S1     |
| 2  | 1    | $t_2$ | S2     |
| 3  | 1    | $t_3$ | S1     |
| 4  | 2    | $t_1$ | S2     |
| 5  | 2    | $t_2$ | S3     |
| 6  | 2    | $t_3$ | S1     |
| 7  | 3    | $t_1$ | S2     |
| 8  | 3    | $t_2$ | S3     |
| 9  | 3    | $t_3$ | S1     |
| ...| ...  | ...  | ...    |

The dataset in Table 1 is a typical traffic dataset that contains time, space, and values. Analysis of such three-dimensional attribute data cannot be solved using traditional methods because traditional association analysis algorithms can only handle transactions with one-dimensional attribute item. Therefore, this paper will improve the data preprocessing method, transform the three-dimensional data into one-dimensional attributes, simplify the difficulty of data processing, and improve the efficiency of data mining.

First, the road segment name is coded, and the corresponding road is represented by a fixed number. Then, according to the requirements of the rule mining, the time is processed. If it is necessary to mine the law of multiple days, the time stamp is used as the time identifier, and the data of one day. As a transaction, a multi-day data set is transformed into a data set with multiple transactions. Finally, the state data is discretized.

This article treats the data for one day into the form of Table 2.

| TID | Items |
|-----|-------|
| 1   | $t_1\_s1,t_1\_s2,t_1\_s1,t_2\_s2,t_2\_s3,t_2\_s1,t_3\_s2,t_3\_s3,t_3\_s1$ |
Data for a certain period of time corresponds to a transaction, and analyzing multiple days of data will generate multiple transactions. The t in Table 2 represents a time period, such as 8:00-8:10, by discretizing the time, merging the space and state. In this way, the basic items in the transaction are converted from three-dimensional to one-dimensional attributes, and the time period can be divided into multiple ranges, and different ranges can mine different meanings.

3. Improved FP-Growth algorithm

The traditional classic FP-Growth algorithm only mines association rules for transaction data sets, and cannot mine spatio-temporal data, and cannot mine association rules that meet traffic conditions. In this paper, the FP-Growth algorithm is improved, and spatio-temporal constraint rules are added to the algorithm to better explore the spatio-temporal relationship between traffic states in the road network.

The FP-Growth algorithm is a memory-based association rule mining algorithm proposed in 2000. It is based on the Apriori algorithm, but its data structure uses the FP tree to store frequent item sets. Compared with the Apriori algorithm, the algorithm does not need to generate a large number of candidates, which can significantly reduce the search overhead. Some studies have shown that the FP-Growth algorithm is effective and scalable for both long mode and short mode, and its processing speed ratio The Apriori algorithm is an order of magnitude faster. The FP-Growth algorithm can be divided into two steps: (1) constructing an FP tree based on data; (2) mining frequent item sets from the FP tree. The improvement of the algorithm in this paper focuses on step (2), that is, mining frequent item sets from the FP tree, adding constraints to this step, and filtering out frequent items that meet the traffic characteristics and instant space feature conditions.

(1) Building an FP tree based on data. The data set is scanned for the first time, and the frequent item sets are obtained. At the same time, the root node of the tree is created and marked with "null"; the database is scanned for the second time, a branch is created for each transaction, and a header list is created at the same time;

(2) Mining frequent item sets from the FP tree. Beginning with a frequent pattern of length 1, construct its conditional pattern base, then construct an FP tree and recursively mine on that tree. The mode growth is achieved by the suffix mode and the frequent mode connection generated by the conditional FP tree.

The resulting FP-Growth mining steps are:
1. the transaction data set are scanned firstly, get frequent 1 item set and its support count;
2. FP tree is constructed based on frequent 1 item sets;
3. transaction data sets are scanned secondly, read transactions one by one;
4. according to the result of step 3, a branch is created one by one, and a head linked list is created;
5. starting from the frequent pattern with a length of 1, its conditional pattern base is constructed;
6. the FP tree is constructed according to step 5, and recursively mine on the tree. After obtaining frequent items, filter out frequent item sets that meet the conditions based on the constraint rules;
7. to determine whether the algorithm is finished or not, jump to the next step;
8. according to the frequent item sets obtained in step 6, skip to step 5 to execute;
9. algorithm ends.

The improved FP-Growth algorithm flow is shown in Figure 1.
Figure 1. Improved FP-Growth algorithm flow chart

The constraint rules of the algorithm are:
1 Frequent items can contain up to two or three items;
2 Frequent items must contain congestion status;
3 In an association rule, the preceding and following items cannot be the same place, and the place is adjacent;
4 In an association rule, the time of the previous item is earlier than the time of the latter item.

Through the constraint of the constraint rule in step 6 of the above algorithm flow, the congestion propagation law that satisfies the time-space constraint of the demand can be filtered out.

4. Case study
This paper selects 12 sections of a certain area in Qing Shan District of Baotou City for analysis, as shown in Figure 2.
The urban traffic state data of the region in December 2017 is selected for analysis. The road segment has a state data every 5 minutes, and 12 road segments are divided into different flow directions, totally 115,000 records. The specific sample data is shown in Table 3.

Table 3. Example analysis sample data

| ID  | TIME          | ROAD                                           | STATUE |
|-----|---------------|------------------------------------------------|--------|
| 1   | 2017120111010 | Xing4fu2lu4qing1nian2lu4gang1tie3da4jie1       | 1.3    |
| 2   | 2017120111100 | qing1nian2lu4fu4qiang2lu4xing4fu2lu4           | 0.3    |
| 3   | 2017120111110 | shao3xian1lu4xing4fu2nan2lu4fu4qiang2zhong1lu4| 1.3    |
| 4   | 2017120111110 | xing4fu2lu4fu3lu4wang4yuan2nan2dao4qing1nian2lu4| 2.1    |
| 5   | 2017120111110 | wang4yuan2nan2dao4fu4qiang2lu4jin4bu4dao4     | 5      |

In Table 3, ID is the name of the link, and the ID is converted into a numeric symbol, corresponding to Figure 2; the status data is determined according to the actual vehicle speed and the free vehicle speed in the road, as shown in formula (1).

\[ TPI = K \times TSI = K \times \left(1 - \frac{v_{ij}}{v_{ij}^f}\right) \]  \hspace{1cm} (1)

Among them, TSI is the travel speed index; K is the traffic index value range factor, the value is >0, if configured according to K=10, then TPI∈[0,10], the larger the value, the more congested the road; \( v_{ij} \) is the ‘i’ grade road, the actual speed of the road is numbered as ‘j’; \( v_{ij}^f \) is the ‘i’ grade road, and the free speed of the road number is ‘j’.

The state preprocessing is divided according to the state level, and the state data of the state data \( TPI \geq 2.5 \) is selected for analysis.

In the case study, we choose the early and late peak period and all day, and consider the different flow direction, such as the clockwise direction 5 and 7, but the No. 7 road has no east to West data, No. 12 road clockwise no data, carry out the following experiments.

4.1 Experiment 1: mining the law of Congestion Propagation in the working day
In this paper, the data of 15 working days (December 4-8, 11-15 and 18-22) in December 2017 are used to mine clockwise and anticlockwise respectively, and the data of slow to severe congestion state are screened out. The rules of early and late peak periods and the whole day are mined respectively.

4.1.1 Early peak
The data of the morning peak from 7:30 to 10:00 recorded a total of 1842 recorded data, as shown in Table 4.

Table 4. Sample data for early peak

| ID  | TIME          | ROAD                                           | STATUE |
|-----|---------------|------------------------------------------------|--------|
| 1   | 201712010815  | xing4fu2nan2lu4gang1tie3da4jie1shao3xian1lu4  | 2.6    |
| 2   | 201712010750  | fu4qiang2zhong1lu4shao3xian1lu4gang1tie3da4jie1| 2.5    |
| 3   | 201712010835  | wang4yuan2nan2dao4jin4bu4dao4fu4qiang2lu4     | 3.2    |
| 4   | 201712010830  | xing4fu2nan2lu4gang1tie3da4jie1shao3xian1lu4  | 2.1    |
| 5   | 201712010830  | xing4fu2lu4qing1nian2lu4gang1tie3da4jie1      | 3.2    |

We use the data preprocessing method proposed in this paper to get the transaction table (Table 5).

Table 5. Sample transaction data for early peak

| TID | Items             |
|-----|-------------------|
| 1   | '0825_9_1', '0815_8_2', '0820_7_1', '0800_8_4', '0830_11_1', ... |
The support of the algorithm is set to 0.5 with a confidence level of 0.5. The clockwise transaction data of morning peak is excavated, and the results are shown in Table 6 and 7.

### Table 6. Clockwise association rules for early peak

| Rule | Confidence |
|------|------------|
| ('0915_6_4') --> ('0925_8_1') | 1.0 |
| ('0915_6_4') --> ('0915_8_1') | 1.0 |
| ('0915_8_1') --> ('0915_6_4') | 0.714285714286 |
| ('0835_6_2') --> ('0840_8_1') | 1.0 |
| ('0835_6_2') --> ('0840_7_1') | 1.0 |
| ('0835_8_1') --> ('0835_6_2') | 0.625 |
| ('0835_6_2') --> ('0835_8_1') | 1.0 |
| ('0750_7_1') --> ('0755_8_4') | 0.666666666667 |
| ('0755_7_1') --> ('0755_8_4') | 0.666666666667 |
| ('0755_8_4') --> ('0755_7_1') | 0.857142857143 |
| ('0815_8_1') --> ('0820_6_2') | 0.714285714286 |
| ('0820_6_2') --> ('0830_7_1') | 0.857142857143 |
| ('0820_8_1') --> ('0820_6_2') | 0.777777777778 |
| ('0820_6_2') --> ('0820_8_1') | 1.0 |

### Table 7. Anticlockwise association rules for early peak

| Rule | Confidence |
|------|------------|
| ('0800_8_4') --> ('0810_9_1') | 0.6 |
| ('0845_8_4') --> ('0845_7_1') | 1.0 |
| ('0845_7_1') --> ('0845_8_4') | 0.75 |
| ('0840_7_1') --> ('0845_8_4') | 0.6 |
| ('0800_8_4') --> ('0800_9_1') | 0.7 |
| ('0800_9_1') --> ('0800_8_4') | 1.0 |
| ('0800_7_1') --> ('0800_8_4') | 0.7 |
| ('0800_8_4') --> ('0810_7_1') | 0.7 |
| ('0745_9_1') --> ('0750_8_4') | 0.857142857143 |
| ('0745_9_1') --> ('0755_8_4') | 1.0 |
| ('0750_8_4') --> ('0755_7_1') | 0.8743 |
| ('0750_8_4') --> ('0750_7_1') | 1.0 |
| ('0750_7_1') --> ('0750_8_4') | 0.9 |
| ('0750_7_1') --> ('0755_8_4') | 1.0 |

After screening, the following useful rules are obtained:

Clockwise direction:
Thus it can be seen:

1) at 7:50 on the working day, West Route 7 will transmit the congestion to the south of the 8 Road north of 7:55 to the South with a 66.7% possibility.

2) at 8:35 on the working day, 100% of North Road 6 will spread the traffic to the north of the 8 Road north of 8:40.

3) at 9:15 on the working day, 100% of North Road 6 will spread the traffic to the north of the 8 Road north of 9:25.

4) at 7:45 on the working day, West Route 9 spread eastward to 85.7% of the possibility of traffic congestion to the north of route 8 South of 7:50.

5) at 7:45 on the working day, West Route 9 will send a 100% possibility to traffic to the north of route 8 South of 7:55.

4.1.2 Late peak

Similarly, a total of 2100 data were screened from late peak to 16:30 to 19:00 and slowly to severe congestion. The support degree of the algorithm is set to 0.5 and the confidence level is 0.5. Mining the late peak transaction data is shown in Table 8 and 9.

| Table 8. Clockwise association rules for late peak |
|--------------------------------------------------|
| ('1840_7_1') --> ('1845_8_2') conf: 0.714285714286 |
| ('1825_8_1') --> ('1830_9_2') conf: 0.714285714286 |
| ('1835_8_1') --> ('1845_6_2') conf: 0.625 |

| Table 9. Anticlockwise association rules for late peak |
|------------------------------------------------------|
| ('1840_7_1') --> ('1845_8_4') conf: 0.857142857143 |

After screening, the following useful rules are obtained:

1) at 18:40 on the working day, West Route 7 will transmit the congestion to the south of the 8 Road north of 18:45 to the South with a 71.4% possibility.

2) at 18:25 on the working day, north of route 8 will transmit the congestion of 71.4% to the south to the east of the road 9 to 18:30.

4.1.3 All day

Similarly, the whole day 07:30 to 20:00, and slow to severe congestion state data, a total of 8218 data. The support of the algorithm is set to 0.5 and the confidence level is 0.5. Mining the clockwise transaction data on the whole day, we get the following useful rules:

| ('0750_7_1') --> ('0755_8_4') conf: 0.667 |
| ('0835_6_2') --> ('0840_8_1') conf: 1.0 |
| ('0915_6_4') --> ('0925_8_1') conf: 1.0 |
| ('0745_9_1') --> ('0750_8_4') conf: 1.0 |
| ('0745_9_1') --> ('0755_8_4') conf: 1.0 |
| ('1825_8_1') --> ('1830_9_2') conf: 0.714 |
| ('1840_7_1') --> ('1845_8_2') conf: 0.714 |
| ('1925_8_1') --> ('1935_9_2') conf: 1.0 |
It can be seen that compared with the data mining results of morning peak and evening peak, an item has been added:
At 19:25 on the working day, 8 North Road to the south to 100% of the possibility of congestion will be transmitted to 19:35 Road 9 to the West.

The algorithm's support is set to 0.5 and the confidence level is 0.5. Mining the anticlockwise transaction data of the whole day, we get the following useful rules:

\([0745_9_1'] \rightarrow [0750_8_4']\) conf: 0.857
\([0745_9_1'] \rightarrow [0755_8_4']\) conf: 1.0

The results are in accordance with the rules of the previous mining.

4.2 Experiment two: mining the law of Congestion Propagation in rest days
Using the rest day data of the 8 days of 2017 (2-3, 9-10, 16-17 and 23-24), the data of clockwise and anticlockwise were excavated respectively, and the data of slow to severe congestion state were screened out, and the rules of the day were excavated. Among them, the data of 5950 records from 7:30 to 22:00 hours were recorded.

In the same way as experiment one, the support of the algorithm is set to 0.5, and the confidence level is 0.5. Data mining is done on all day affairs, and the results are shown in tables 10 and 11.

Table 10. Clockwise association rules for all day periods of rest day

| Rule                  | Confidence |
|-----------------------|------------|
| ([1805_6_1']) \rightarrow ([1815_8_2']) | 0.571428571429 |
| ([1810_9_1']) \rightarrow ([1815_8_2']) | 0.571428571429 |

Table 11. Anticlockwise association rules for all day periods of rest day

| Rule                  | Confidence |
|-----------------------|------------|
| ([1815_8_1']) \rightarrow ([1825_9_2']) | 0.571428571429 |
| ([1840_8_1']) \rightarrow ([1850_9_2']) | 0.666666666667 |

The following useful rules are obtained by screening:
Clockwise direction: \([1805_6_1']) \rightarrow ([1815_8_2']) conf: 0.571428571429;
Thus, on the day of rest, at 18:05, north of Route 6 will have a 57.7% chance of spreading north to the south of the road 8.

4.3 Experiment three: comparison of rule mining operation time on rest day
Taking the data of the rest day in the second experiment as the experimental data set, using the FP-growth algorithm without spatio-temporal constraint and adding spatio-temporal constraint algorithm, the following running time is obtained, as shown in Figure 3.

![Figure 3. Comparison of time constraints between adding spatio-temporal constraint algorithm and adding constraint algorithm](image)

As can be seen from the above diagram, the FP-growth algorithm with spatio-temporal constraints significantly improves the efficiency of mining rules. The FP-growth algorithm without spatio-
temporal constraints can dig more rules, but most of the rules have no practical value. After artificial screening, the results of the algorithm with spatio-temporal constraints are consistent. Therefore, the analysis method proposed in this paper can more effectively mine the propagation law of frequent traffic congestion.

5. Conclusion

In view of the analysis of the regular traffic congestion propagation law, a data preprocessing algorithm is proposed, which can transform the three-dimension attribute data into one-dimensional properties, and add the spatio-temporal constraints to the FP-Growth algorithm to achieve more efficient traffic congestion propagation law mining. On this basis, the application algorithm is studied and the experimental study is carried out. The correlation between various sections of the rest day road network is obviously lower than that of the working day, and the correlation of each road in the middle and late peak road network of the working day is obviously lower than that of the early peak. In addition, in addition to the rules of Congestion Propagation in the experiment, the analysis results of the temporal and spatial correlation of traffic flow are also obtained. The results show that the improved algorithm proposed in this paper can effectively analyze the traffic congestion propagation law in road network, and can be used for reference to alleviate traffic congestion in urban roads.

References

[1] Morales, J.M. (1987) Analytical procedures for estimating freeway traffic congestion. ITE J., United States, 57:1(1):55-61.
[2] Nam, D., Mannering, F. (2000) An exploratory hazard-based analysis of highway incident duration. Transportation Research Part A., 34(2):85-102.
[3] Chen, T., Chen, S.F. (2004) Research on Dissipative Structure Characteristics of Urban Road Traffic System. China Civil Engineering Journal., 37(1):74-77.
[4] Zhang, J.F. (2016) Urban traffic jam propagation and control strategies based on complex networks. Doctoral dissertation, Lanzhou Jiaotong University.
[5] Wei, W. (2017) Analysis method of road traffic status characteristics and congestion propagation law based on measured data. Doctoral dissertation, Beijing Jiaotong University.
[6] Zhang, J.F., Ma, C.X., Wu, F., Pu, H., Jia, F.Q. (2015) Improved SIS Model for Traffic Congestion in Complex Urban Traffic Networks. Journal of Transportation Research., 1(6):20-25.