Research Article

Impact Analysis of Land Use on Traffic Congestion Using Real-Time Traffic and POI

Tianqi Zhang, 1 Lishan Sun, 1 Liya Yao, 2 and Jian Rong 1

1 Beijing Key Laboratory of Traffic Engineering, Beijing University of Technology, 100 Pingleyuan, Chaoyang District, Beijing 100124, China

2 School of Mechanical and Vehicular Engineering, Beijing Institute of Technology, 5 South Zhongguancun Street, Haidian District, Beijing 100081, China

Correspondence should be addressed to Lishan Sun; hill310@sina.com

Received 30 March 2017; Revised 26 June 2017; Accepted 1 October 2017; Published 26 October 2017

Academic Editor: Dongjoo Park

Copyright © 2017 Tianqi Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper proposed a new method to describe, compare, and classify the traffic congestion points in Beijing, China, by using the online map data and further revealed the relationship between traffic congestion and land use. The data of the point of interest (POI) and the real-time traffic was extracted from an electronic map of the area in the fourth ring road of Beijing. The POIs were quantified based on the architectural area of the land use; the congestion points were identified based on real-time traffic. Then, the cluster analysis using the attributes of congestion time was conducted to identify the main traffic congestion areas. The result of a linear regression analysis between the congestion time and the land use showed that the influence of the high proportion of commercial land use on the traffic congestion was significant. Also, we considered five types of land use through performing a linear regression analysis between the congestion time and the ratio of four types of land use. The results showed that the reasonable ratio of land use types could efficiently reduce congestion time. This study makes contributions to the policy-making of urban land use.

1. Introduction

As the urbanization of China is accelerating, the expense of urban land leads to excessive concentration of public functions, causing the growing occurrences of congestions in urban traffic. The land use affects the attracted direction, the ratio of traffic flow, and the travel model, which are the factors related to public traffic demand. Reasonably planning urban land is essential to ensure the efficient operation of urban traffic. Therefore, understanding the correlation between the land use and the traffic congestions can help optimize urban traffic.

The traffic data on congestions involves large-scale and complex space-time information, making mining the traffic data difficult. Besides, the source of the traffic data is not readily available. Previous studies [1–6] have focused on traffic flow through traditional methods (conventional four-step travel demand model) without considering geographic information (the coordinates of longitude and latitude, the categories, and the specific location information). The traditional four-step travel demand model typically operates on the individual survey data that is of high cost, low accuracy, and low efficiency. So it is necessary for us to develop a method of judging congestion point using geographic information systems (GIS), which serves as a quick, precise alternative to the conventional four-step models.

In literature, limited models on traffic congestions have been proposed to investigate the relation between traffic congestions and urban land use. For example, Wing Lowdon established the economic model on how transportation, location, and urban land use affected the travel of consumers from their residences to workplaces [7]. Alonso [8] improved this model by considering the value of the urban land, finding that the value of different urban plots was negatively correlated to the transportation cost to the city center. Izraeli and McCarthy [9] (1985) found that the residential land had an effect on congestion; there was a significant positive correlation between population density and commuting time. Handy [10] analyzed the impact of land use on travel characteristics and discovered that the
frequency of traveling decreased as the density of land use increased and that the distance of traveling increased as the speed of traveling decreased. Gordon et al. [11] (1989) analyzed satellite data of 82 US metropolitan areas in 1980 to extract the information on the densities of different types of the land use (the type of the resident, the industry, and the commerce). When considering the employment rate at that time, it was found that the increment of the industrial density would lower the car commuting time, as well as the residential and commercial densities. Ewing et al. [12] (2003) investigated the impact of land use on commuting time and pedestrian delay using cross-sectional data of 83 metropolitan statistical areas at the years of 1990 and 2000. The results showed that the commuting time during these two years was negatively correlated with the mixed land utilization index and was positively correlated with the street accessibility.

Unlike the urban land attribute data that is complex for analyzing and classifying, the point of interest (POI) data, which is closely related to urban land attributes and urban planning guidance, can be easily quantified and analyzed. Yu and Ai [13] discussed the characteristics of the spatial distribution of urban POI data and proposed a model to estimate network kernel density for providing guidance to land planning. Ma et al. [14] proposed a visual search model for POIs of highway transportation to help reduce the costs of transportation. Liu et al. [15] computed the attractiveness of POIs according to the number of times of taxis stopping nearby the POIs.

No literature has integrated the real-time traffic data to the POI data. Using the urban road network data, the real-time traffic data, and the POI data, this study explored the correlation between traffic congestion and different attributes of urban land use and established the evolution of urban traffic congestion geographic model. The outcomes of this study would contribute to the policy-making in the planning of urban land use.

2. Introduce and Extraction of POI

2.1. Introduce of POI. In a Geographic Information System, the POIs include the houses, the scenic spots, the shops, and the mailbox. The data from the POIs contains the coordinates of the longitude and latitude, the categories, the specific location information, and the User Identification (UID). This study used the electronic map of Beijing POIs since it recorded an enormous amount of information on city locations.

2.2. Classification and Extraction of POI. The default classification of Beijing POIs is of 16 categories as the first class and 96 categories as the second class, most of which are not related to the traveling of residents. This study conducted a detailed survey to obtain a better classification of Beijing POIs well representing resident travel purposes. The result shows that “work,” “school,” “shopping,” “leisure,” and “return home” are the primary traveling purposes of Beijing residents, accounting for about 85% of the total travel, as listed in Table 1. Thus, new categories of POI for classification, including education, work, shopping, residential, and recreation, were created, as shown in Table 2.

This study also extracts about 90,000 POIs within fourth ring road in Beijing using web-crawler software. The POI data, which contained four types of information, name, category, longitude, and latitude, were processed to determine characteristics of land use for revealing the relationship between traffic congestion and different types of land use.

2.3. Quantization of the POI. The quantization of POI focuses on the architectural area, which is the main city index for planning stage and objectively reflects current indexes of the population, jobs, the social economy, and the development of specific regions. Because of the large number of POIs, it is not practical to conduct an investigation using all POIs in the architectural area. Hence, POIs were randomly sampled for investigating. The sampling size n is calculated according to the following equation:

\[
n = \frac{P(1-P)}{(e^2/Z^2 + P(1-P)/N)},
\]

where \(e\) is the percentage of expected error, \(Z\) is the bilateral quantiles corresponding to a particular confidence, \(P\) is the degree of sample variation, and \(N\) is the population number. In this study, we assume that \(e\) is 10% and \(Z\) is 1.64 (the

| Travel purpose          | Percentage | Cumulative percentage |
|-------------------------|------------|-----------------------|
| Commuting to work       | 36.34%     | 36.34%                |
| Returning               | 28.13%     | 64.47%                |
| Commuting to school     | 7.88%      | 72.35%                |
| Shopping                | 6.96%      | 79.31%                |
| Leisure                 | 6.44%      | 85.75%                |
| Picking/taking somebody up | 3.20%     | 88.95%                |
| Business affairs        | 3.18%      | 92.13%                |
| Visiting friends        | 3.05%      | 95.18%                |
| Personal affairs        | 2.51%      | 97.69%                |
| Delivery and pickup     | 1.97%      | 99.66%                |

| Original type          | New type   | Corresponding trip purpose |
|------------------------|------------|---------------------------|
| Education              | Work       | Return trip               |
| Elementary education   | Going to school |                        |
| Secondary education    |             |                           |
| Higher education       |             |                           |
| Firm                   |             |                           |
| Business               |             |                           |
| Retail store           |             |                           |
| Outlet                 |             |                           |
| Real estate community  |             |                           |
| Leisure and recreation |             |                           |

Table 1: Proportions of residents travel purpose in Beijing (2014).

Table 2: Classification of POI based on main travel purpose.
Table 3: Sampling questionnaire of various land uses.

| Classification      | Total number | Sampling number | Ratio  |
|---------------------|--------------|-----------------|--------|
| Preprimary education| 1499         | 64              | 4.3%   |
| Elementary education| 976          | 63              | 6.5%   |
| Secondary education | 1435         | 64              | 4.5%   |
| Higher education    | 416          | 58              | 14.0%  |
| Firm                | 19276        | 67              | 0.3%   |
| Retail store        | 5884         | 66              | 1.1%   |
| Outlet              | 6037         | 66              | 1.1%   |
| Residential         | 8358         | 67              | 0.8%   |
| Leisure and recreation| 8225        | 67              | 0.8%   |
| Sum                 | 52106        | 582             | 1.1%   |

Table 4: The data format of the architectural area of POI.

| Object ID | POI name                              | Type   | Construction area/m² |
|-----------|---------------------------------------|--------|----------------------|
| 3089      | The Art Museum of H.M                 | Leisure| 1000                 |
| 3090      | Mosque                                | Leisure| 500                  |
| 3091      | Sihui Karting Club                    | Leisure| 2000                 |
| 3092      | Tsinghua University                   | Education| 16000              |
| 3093      | Peking University                     | Education| 18000              |
| 3094      | Beijing forth school                  | Education| 3000                |
| 3095      | Tianannmen Square                     | Leisure| 2000                 |
| 3096      | Dongdan Park                          | Leisure| 1500                 |
| 3097      | National Museum                       | Leisure| 2000                 |
| 3098      | Sihui Shopping Plaza                  | Business| 1600                |
| 3099      | New World Wholesale Mall              | Business| 1700                |
| 3100      | China Central Academy of Fine Arts    | Education| 7000                |
| 3101      | China World Trade Center Exhibition Hall| Leisure| 2000               |
| 3102      | Beijing Economic and Trade Company    | Company | 700                  |
| 3103      | Haidian Primary School                | Education| 1500                |

confidence is 90%). Since the degree of sample variation is unknown, let the maximum value of \( P \) be 0.5. After applying the above values to (1), the maximum samples rate is 14.0% and the minimum is 0.3%. Based on the above calculation, 58–67 POIs for each type of land use were randomly selected in the field survey in May 2016. Table 3 lists the specific sampling rates of various types of land use.

The investigation methods included network inquiry, telephone inquiries, site investigation, and site inquiries. Table 4 shows the data format. Table 5 lists the average architectural area of each type of land use calculated based on the survey results. The total architectural area in every congestion region was calculated according to the number of instances of each type of POIs in this region, which provides conditions for analysis of the correlation between congestion points and land use in its coverage.

3. Introduce and Extraction of Real-Time Traffic

3.1. Real-Time Traffic Profile. The road states in the real-time traffic information from web search engines are the congestion, amble, and smooth, which have colors of red, yellow, and green, respectively, reflecting the traffic conditions in the region. The map application (GPS or smartphone) uses the information of the road states to navigate roads and plan the trips for drivers.

Table 5: Average architectural area of all types of POI.

| Original type          | New type | Average construction area/m² |
|-----------------------|----------|-----------------------------|
| Education             |          |                             |
| Preprimary education  | Education| 3715                        |
| Elementary education  | Education| 6782                        |
| Secondary education   | Education| 37936                       |
| Higher education      | Education| 170772                      |
| Company               | Work     | 2399                        |
| Business              |          |                             |
| Retail store          | Shopping | 254                         |
| Outlet                |          | 56756                       |
| Real estate community | Residential| 149686                      |
| Leisure and recreation| Leisure  | 4518                        |
3.2. Real-Time Traffic Data Acquisition and Processing. Map images on the web search engines are composed of different layers: the bottom layer displays basic geographic information; the middle layer shows names, comments, and other information of different venues; the upper layer (it was used in this study) contains traffic information depicted with colors red, yellow, and green. We downloaded all blocks of the upper layer and then spliced them according to their naming rules, generating a map containing the real-time network running state for a particular area. The real-time data were collected every 5 minutes, a total of 60 times a day, at weekday morning and evening peak in the first week of March 2016. Note that Beijing City morning peak period is 7:00 am–9:30 am and evening peak period is 17:00 pm–19:30 pm. The weather in that week was sunny without rain, snow, or other unusual weather conditions. Figure 1 shows the data format of real-time traffic. Table 6 lists the schematic color-codes of congestion levels.

![Figure 1: The schematic of real-time road traffic on working day morning peak.](image_url)

### Table 6: The schematic color codes of congestion levels.

| Congestion levels | Color code | Color sample |
|-------------------|------------|--------------|
| Red (congestion)  | 243, 48, 48| —            |
| Yellow (amble)    | 255, 158, 25| —            |
| Green (smooth)    | 23, 191, 0 | —            |

![Figure 2: The schematic of traffic vector on working day morning peak.](image_url)

Congestion time is an evaluation index of congestion for reflecting congestion intensity. The change forms of congestion properties about one congestion point in the adjacent moment are 1-3, 2-3, and 3-3, respectively, denoted as a five-minute jam. The three forms of the point represent that the state turns from smooth passing to congestion and from slow passing to congestion and is in continuous congestion.

Based on the real-time traffic data, the vector attribute table of congestion properties at different time instances was obtained. The Python program superposed the congestion time into single accumulated time per hours per morning/evening peak.

4. Congestion Distinguishing and Influence Scope

4.1. Congestion Points Discrimination and Its Spatial Distribution. The road traffic status is only affected by a previous step time or a forward direction on the road. From the time dimension, congestion is divided into primary congestion and secondary congestion. The primary congestion point has an increasing congestion intensity, while the segment ahead of it has the same or reduced congestion intensity. On the contrary, the secondary congestion point has a growing congestion intensity, while the segment ahead of it also has an increasing congestion intensity. Following the above definition, the types of the congestion points are related to the time variance and the adjacent spatial space. However, a congestion point in a vector layer is discrete, and the adjacent spatial space is unsearchable. Therefore, it is necessary to record the topological relationship among points by the actual traffic flow direction.

The directions of the road traffic flows were monitored on the electronic map; the discrete points were edited. The road point attribute at the end of the traffic flow was marked as “0,” because the roads were not completely closed.

Based on the above definition, this paper set up criteria for the types of congestion points. Let $C_i(t)$ be the degree of congestion in road point $i$ at the $t$th time and road point $i$ is
A decision tree is constructed as shown in Figure 3. When $C_i(t) \leq C_i(t-1)$, the road point $i$ is not a congestion point; when $C_i(t) > C_i(t-1)$, it is necessary to consider the degree of congestion in road point $j$ to judge the type of congestion. If $C_j(t) \leq C_j(t-1)$, the road point $i$ will be defined as primary congestion point; vice versa, if $C_j(t) > C_j(t-1)$, the road point $i$ will be defined as secondary congestion point.

The Python program modeled the criteria tree for simulating the real-time traffic at morning and evening peaks. A distribution map of congestion points at the morning and evening peaks was also generated, as shown in Figure 4.

According to a sample survey in the city of Beijing, the public considers a congestion time longer than 65 minutes being extremely unacceptable and a congestion time shorter than 30 minutes being acceptable. Hence, we grouped the congestion time into 65, 95, 125, 155, 185, 215, and 245 minutes with 30 minutes as the interval. The congestion points with congestion time longer than 65 minutes were extracted for visualization. Figure 5 presents the visualization of the congestion time at the morning and evening peaks in Beijing. Within the 2nd ring of Beijing, ITC, the commercial districts of Asian, Xidan, Wangfujing, and Zhongguancun had severe congestion. And the regions of Songjiazhuang, Fangzhuang, and Dahongmen, which are residential and recreation districts in Beijing, had slight congestion. It is hypothesized that the congestion intensity is related to the proportion of different land types. Thus, the congestion time was used to cluster the congestion points.

### 4.2 Congestion Influence Scope Based on Space Clustering

Clustering is the method to take objects into the same group according to spatial and temporal similarity to maximize the between-group difference and minimize the within-group difference. The cluster results can discover and analyze the law and essential features of the development and change of geographical phenomena.

This study used the congestion time attributes of the congestion points to cluster the congestion areas. The interpolation method was used to make the clustering more reasonable and close to the real situation in the whole space. The method preprocessed the discrete data, generating
Table 7: Clustering result based on congestion time.

| Classification | Congestion time interval/min | Average congestion time/min |
|----------------|------------------------------|----------------------------|
| 1              | 0–65                         | 41.7                       |
| 2              | 65–95                        | 81.6                       |
| 3              | 95–125                       | 107.2                      |
| 4              | 125–155                      | 146.9                      |
| 5              | 155–185                      | 165.4                      |
| 6              | 185–215                      | 193.6                      |
| 7              | 215–245                      | 232.2                      |

The congestion points in Beijing with the congestion time 65 minutes or longer were clustered. The primary congestion points directly affected the secondary congestion points; the relationship with the proportion of land use was not strong. Hence, the secondary congestion points were not considered, and only the primary congestion points were analyzed in this paper. Figure 6 presents the cluster results about the congestion time of the primary congestion points. Table 7 lists the clustering results using the average congestion time of each area as the dependent variable.

5. Correlation Analysis

This study used a multiple linear regression to identify the relationship between the land use and the traffic congestion:

\[ Y = x_1 E_p + x_2 B_p + x_3 R_p + x_4 C_p + x_5 L_p + \alpha, \]  

where \( \alpha \) was the estimable intercept term (equalling zero), the dependent variable \( Y \) was the average congestion time of congestion region, the independent variables \( E_p, B_p, R_p, C_p, L_p \) were architectural area ratio of five main types of POIs (education, shopping, residential, work, and leisure) within the congestion region, and \( x_1, x_2, x_3, x_4, x_5 \) were the corresponding coefficients.

The congestion time and proportion of land in each cluster region in ArcGis were analyzed. We exclude the unconventional regions, such as Beijing passenger bus station, large parks, and regions that have a large proportion of tourism land. A total of 274 regions that have more than 65-minute congestion time were selected.

The selected regions were in five types of the residential, education, business, company, and leisure land. Since the dependent variable was ordered and multiclassified, IBM SPSS Statistics 19.0 was used to perform an ordered analysis. The results in Table 8 showed that the congestion time was positively correlated with the proportion of company land, commercial land, and educational land. This was because the primary purpose of traveling at the morning and evening peaks of weekdays was commuting to school or workplace. The commercial land had the most intensive congestion because the area within fourth ring road in Beijing had the highest proportion of commercial land. Moreover, the congestion at morning and evening peaks of weekdays was irrelevant to the proportion entertainment land.

Based on the analysis above, the commercial land, the residential land, the educational land, and the company all had a strong correlation with the congestion time. These four types were selected as the independent variables in the linear regression model:

\[ Y = 31.867 E_p + 128.345 B_p + 25.285 R_p + 104.358 C_p + 1.7 x \]  

The fitting degree of the linear regression model is 0.751, which shows that there is a high correlation between the four types of land use and congestion. It was found that as the proportion of commercial land became higher in a particular area, one of other types of land use would be lower and the type of land use would be simpler. Besides, the regions with more commercial land would cause severe traffic
congestion, because residents would have the same start or end time in the morning and evening peak, making traffic flow simultaneously arriving in the morning peak or leaving in the evening peak. On the contrary, regions with a lower proportion of commercial land would have less congestion, because the commuting time varies for different residents.

The above analysis for congestion considered each land use as a single property type, while the land use might have mixed types. For reducing the traffic congestion, the ratio of the four types of land use \( x = E_p/B_p/R_p/C_p \) was regarded as the influencing factor. The linear regression between the congestion time and the ratio was performed, resulting in the following equation:

\[
Y = 89.442 \times x^3 - 340.210 \times x^2 + 452.338 \times x - 28.333
\]

with an \( R^2 \) value of 0.637. Also, different proportions of land use types had a strong correlation with the congestion time. Therefore, this model can be used to predict the congestion time by the ratio of the four types of land use \( x \). Besides, in the process of planning of land use, this model can also be used to coordinate the proportion of different types of land use according to the conditions of different areas. For example, if a resident land is built, there is also need to plan corresponding education land, business land, or company land. Based on the model, the reasonable ratio of the four types of land use can be calculated by the acceptable congestion time. Then, the proportion of the other three types of land use is coordinated according to the residential area.

### 6. Discussions and Conclusions

This study visualized the congestion points in space and showed that the congestion time was a reliable index to describe the congestion states. A cluster analysis on the congestion time data indicated that congestion time could be divided into seven categories according to congestion time intervals. In a multiple linear regression analysis, it was found that the high proportion of commercial land would significantly affect the congestion time. Besides, the linear regression analysis with the ratio of the five types of land use for the congestion time indicated that the reasonable proportions of land use types are more efficient to reduce congestion time and ensure efficient operation of urban traffic. One of the main contributions of this study is that it provides a feasible way to analyze traffic congestions from the online map data. Moreover, this study provides policy guidance for urban reform and gives a reference for the planning of urban land use. The regression models in the study can help us predict and estimate the correlation between land use and congestion intensity without the dependency on complex traffic models (such as traffic four-stage method) and large numbers of data acquisitions.

The traffic data in this study were from Beijing urban land use. Further studies will use traffic data from other regions to determine a general pattern of the congestion and land distribution model. Further, the effect of the mixing ratio of land use on traffic congestion will be studied to obtain an optimized mixing ratio of land use.

### Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

### Acknowledgments

The authors would like to acknowledge the financial support for this study provided by the National Natural Science Foundation of China (no. 51308017), Beijing Nova Program (Grant no. Z141106001814110), Science and Technology Program of Beijing (Grant no. D16100005616001), and Funding Project for Academic Human Resources Development in Institutions of Higher Learning under the Jurisdiction of Beijing Municipality and Ministry of Housing and Urban Construction in 2015 Science and Technology Project Plan (no. 2016K8044).

### References

[1] C. Gridlock, “Peak-period fees to relieve traffic congestion. Volume 2. Commissioned papers. Special report. Transportation Research Board Special Report 1994”.

[2] A. Sarzynski, H. L. Wolman, G. Galster, and R. Hanson, “Testing the conventional wisdom about land use and traffic congestion: the more we sprawl, the less we move?” Urban Studies, vol. 43, no. 3, pp. 601–626, 2006.

[3] R. J. Smeed, “Traffic studies and urban congestion,” Journal of Transport Economics and Policy, vol. 2, no. 1, pp. 33–70, 1968.

[4] N. Shibata, T. Terauchi, T. Kitani, K. Yasumoto, M. Ito, and T. Higashino, “A method for sharing traffic jam information using inter-vehicle communication,” in Proceedings of the 2006...
[5] R. Arnott, A. de Palma, and R. Lindsey, “Economics of a bottleneck,” *Journal of Urban Economics*, vol. 27, no. 1, pp. 111–130, 1990.

[6] S.-I. Mun and M. Yonekawa, “Flextime, traffic congestion and urban productivity,” *Journal of Transport Economics and Policy*, vol. 40, no. 3, pp. 329–358, 2006.

[7] Wingo Lowdon, *Transportation and Land Use*, Resources for the Future, Washington, DC, USA, 1961.

[8] W. Alonso, “Location and land use: toward a general theory of land rent,” *Economic Geography*, vol. 42, no. 3, pp. 11–26, 1966.

[9] O. Izraeli and T. R. McCarthy, “Variations in travel distance, travel time and modal choice among SMSAs,” *Journal of Transport Economics & Policy*, vol. 19, no. 2, pp. 139–160, 1985.

[10] S. Handy, “How land use patterns affect travel patterns: a bibliography,” *CPL Bibliography*, no. 279, 1992.

[11] P. Gordon, A. Kumar, and H. W. Richardson, “Congestion, changing metropolitan structure, and city size in the united states,” *International Regional Science Review*, vol. 12, no. 1, pp. 45–56, 1989.

[12] R. Ewing, R. Pendall, and D. Chen, “Measuring Sprawl and Its Transportation Impacts,” *Transportation Research Record*, no. 1831, pp. 175–183, 2003.

[13] W. Yu and T. Ai, “The visualization and analysis of POI features under network space supported by kernel density estimation,” *Acta Geodaetica et Cartographica Sinica*, vol. 44, no. 1, pp. 82–90, 2015.

[14] X. Ma, Y. Cao, and X. Mu, “The POI of highway transportation and its visual search,” *Applied Mechanics and Materials*, vol. 253-255, no. 1, pp. 1319–1323, 2013.

[15] Y. Z. Liu, L. S. Li, Z. H. Liao, J. X. Liu, and M. D. Tang, “Analyzing POI’s attractiveness variation based on taxis’ trajectories,” *Applied Mechanics and Materials*, vol. 610, pp. 482–486, 2014.
