ISSN:2790-1688 DOI: 10.56028/aetr.4.1.648.2023

Identification Algorithm of Accident Road Section Attribute Based on OBD Data Map Matching

Xingqiang Zhang [1, a], Huihui Li [1, b], Xue Liu [2, c]

[1] Key Laboratory of Transport Industry of Big Data Application Technologies for Comprehensive Transport Ministry of Transport, Beijing Jiaotong University, Beijing, China;

[2] Rizhao Planning Design Research Institute Group Co. LTD, Rizhao, China.

[a] zhangxq@bjtu.edu.cn, b rj_lee_lee@163.com, c 75222036@qq.com

Abstract. Traffic accident not only causes casualties, but also generates traffic congestion, resulting in severe traffic delays and other indirect economic losses. Based on the accurate identification of accident road section attribute, considering the accidental traffic impact is very important to improve the comprehensiveness and accuracy of road location safety evaluation. In this paper, OBD data is used to map matching, the fuzzy clustering algorithm is used to discriminate and classify the road traffic operation status. A kind of identification algorithm of the accident road section traffic attribute impact is proposed, which is used as case analysis of Beijing Fourth Ring Road. The analysis results show that the accident occupying lane is the most important factor of traffic delay, the more lanes are occupied, the more vehicles of single lane are delayed and the delay time is longer.

Keywords: map matching; OBD data; accident attribute; fuzzy clustering; traffic level; traffic delay.

1. Introduction

With the rapid development of urban traffic, the indirect economic losses such as vehicle delays caused by traffic congestion due to traffic accidents are a huge consumption of social resources. With the development of vehicle positioning and Internet technology, it is of great significance to establish a fast and effective identification method of traffic attributes of accident sections based on vehicle real-time trajectory data and Internet accident data to improve the comprehensiveness and accuracy of safety evaluation of urban road locations. The current research on the impact of traffic accidents on road traffic is mainly carried out by 2 means: theoretical modeling and simulation. He Yaqin[1] used the principle of water wave science to establish the traffic impact level threshold model of traffic accidents on intersections by traffic impact coefficient and fractional velocity. Yu[2] analyzed the accumulation and dissipation process of accidents upstream of two-lane highways by using shock wave theory, and established a highway traffic accident impact model based on the number of affected lanes, accident disposal time and upstream traffic flow. Tang Jinjun[3] studied the intermittent flow characteristics under signal control conditions and the diversion characteristics on the inlet lane of the intersection, calculated the single-lane queue length using traffic wave theory to determine the boundary of the accident impact area, and established a calculation model for the vehicle queue length under accident impact conditions. Xiao Huihui[4] analyzed the impact of traffic accidents on the main feature parameters of metric automata based on the traffic flow characteristics of urban expressways, and established a prediction model of the impact range of expressway traffic accidents based on the simulation of metric automata. At present, with the improvement of the accuracy of in-vehicle OBD trajectory data, the Internet accident data are convenient to obtain, and these data have wide coverage and low cost, and can be widely used in accident traffic impact analysis. Therefore, this paper uses OBD trajectory data and Internet accident data for map matching to obtain traffic flow information and accident time-space information. Based on the operation speed, FCM clustering and grading of urban road traffic operation status is carried out to construct the identification algorithm of accident traffic impact on road sections. The delay time and vehicle number are used as indicators to construct the accident indirect loss model. The OBD vehicle trajectory data and Internet accident data in Beijing are also

ISSN:2790-1688 DOI: 10.56028/aetr.4.1.648.2023

used as examples for validation and analysis to provide a theoretical basis for the safety evaluation of urban locations considering the impact of accident traffic.

2. FCM clustering of road section operation state based on map matching

Let $P=\{p_1, p_2, \cdots, p_k, \cdots p_p\}$ be the set of anchor points to be matched according to the matching range of OBD data. The index grid of anchor point p_k is composed of the grid in which it is located and 8 adjacent grids around it, and the index grid of segment $Link_i$ is the "rectangle" area composed of the grid in which two shape points are located as the vertex grid. If p_k coincides with the index grid of $Link_i$, then the matching candidate of p_k in the electronic map is $Q=\{Link_1, Link_2, \cdots, Link_i, Link_n\}$.

Determine whether the is within the matching range of according to the road matching threshold, as shown in equations (1) and (2)

$$d_{ik}=\begin{cases}\min\{d_{ijk}\}, d_{ijk}\le D=\varepsilon_1+\varepsilon_2+\varepsilon_3 \\ p_k \text{ is not in Link}_i\text{'s match, } d_{ijk}>D\end{cases} \tag{1}$$

Where: d_{ik} is the distance between the anchor point p_k and $Link_i$; D is the road matching threshold of the vehicle location point; ε_1 is the anchor drift error; ε_2 is the accuracy of electronic map; ε_3 is other errors; d_{ijk} is the distance between p_k and each $Link_i$ subsegment in set Q.

$$d_{ijk}=\begin{cases}l_{kA}, & \angle p_k AB\ge 90^\circ \\ \dfrac{2\sqrt{l(l-l_{AB})(l-l_{kA})(l-l_{kB})}}{l_{AB}}, & \triangle p_k AB \text{ is an acute triangle, } l=\dfrac{l_{AB}+l_{kA}+l_{kB}}{2} \\ l_{kB}, & \angle p_k BA\ge 90^\circ\end{cases} \tag{2}$$

Where: AB and BC are seg_j and seg_{j+1} subsegments in $Link_i$, M and N are the projections of p_k on seg_j and seg_{j+1} subsegments, respectively. l_{AB} is the distance between A and B; l_{kA} is the distance between p_k and A; l_{kB} is the distance between p_k and B.

Taking the average speed of the road section as the judgment index of the traffic operation status of the road section, according to the sampling time interval of OBD data, the average speed of a single vehicle is calculated by using the location-time interpolation method to calculate the average speed of the road section. Let the start and end points of road section i be M and N, respectively; vehicle j travels at a constant speed between the location points p_k and p_{k+1}; the time t_{Mj} of vehicle j passing through M and the average speed $\bar{v}_{ij}$ passing through road section i are:

$$t_{Mj}=\frac{t_k l_{Mk+1}+t_{k+1} l_{kM}}{l_{Mk+1}+l_{kM}} \tag{3}$$

$$\bar{v}_{ij}=\frac{L_i}{t_{Nj}-t_{Mj}} \tag{4}$$

Where: t_k and t_{k+1} are the location time of vehicle j at p_k and p_{k+1} respectively; l_{kM} and l_{Mk+1} are respectively p_k, the distance from p_{k+1} to M; L_i is the length of road section i; t_{Nj} is the time when vehicle j passes the end point N of road section i.

$$\boldsymbol{u}=\left[\boldsymbol{u_1}, \boldsymbol{u_2}, \cdots, \boldsymbol{u_c}\right]^{\bullet}=\begin{bmatrix}u_{11} & u_{12} & \cdots & u_{1m} \\ u_{21} & u_{22} & \cdots & u_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ u_{c1} & u_{c2} & \cdots & u_{cm}\end{bmatrix} \tag{5}$$

ISSN:2790-1688 DOI: 10.56028/aetr.4.1.648.2023

Where: u is the membership matrix; c is the number of clustering categories of data; u_{ij} is the membership of the i index of the j data sample.

$$\mu_i^{\gamma} = \frac{\sum_{j=1}^{m} u_{ij}^{\gamma-1} \boldsymbol{X}_i}{\sum_{j=1}^{m} u_{ij}^{\gamma-1}} \tag{6}$$

$$d_{ij}^{(\gamma)} = \left\| \boldsymbol{X}_i - \mu_i^{\gamma} \right\| = \sqrt{\sum_{j=1}^{m} (\bar{v}_{ij} - \mu_{ij}^{\gamma})} \tag{7}$$

$$u_{ij}^{\gamma} = \begin{cases} 1, d_{ij}^{(\gamma)} = 0 \,\&\, i = j \\ \sum_{r=1}^{c} \left[\left(\dfrac{d_{ij}^{(\gamma)}}{d_{rj}^{\gamma}} \right)^{\frac{2}{\alpha-1}} \right]^{-1} \\ 0, d_{ij}^{(\gamma)} = 0 \,\&\, i \neq j \end{cases} \tag{8}$$

Where: μ_i^{γ}, $d_{ij}^{(\gamma)}$ and u_{ij}^{γ} are the cluster center, Euclidean distance and fuzzy membership matrix of the γ iteration, respectively. ε is the threshold of the iteration stopping condition; α is the fuzzy weighting index. If $\left\| \mu^{\gamma} - \mu^{\gamma+1} \right\| < \varepsilon$, the iteration is stopped, $c = \arg \max_{i=1,2,\cdots,c} u_{ij}, j(1,2,\cdots m)$.

3. Attribute identification of the accident road section

In this paper, the frequency of occurrence of weekly congestion $f(i,j)$ is used as the discriminant index of $L(i,j)$ in the weekly frequent congestion road section.

$$f(i,j) = \sum_{k=1}^{7} \alpha_{(i,j,k)} \tag{9}$$

$$\alpha(i,j,k) = \begin{cases} 1, & G(i,j,k) > g(i,j) \\ 0, & G(i,j,k) \leq g(i,j) \end{cases} \tag{10}$$

Where: i, j and k are the time period, road section and date respectively; $\alpha(i,j,k)$ is the discriminant coefficient of congestion grade; $g(i,j)$ is the road section congestion grade; $G(i,j,k)$ refers to traffic operation status of 1 hour or more on a single day of the week.

$$V(i,j) = \frac{1}{f(i,j)} \sum_{i=1}^{k} v(i,j,k) \tag{11}$$

Where: $V(i,j)$ is the weekly average speed, the congestion period of the i and the j road section of weekly frequent congestion road section $L(i,j)$, (km/h).

The accident traffic impact time and space range of the accident road section $l[I,J]$ is $R = \{RT, RL\}$, $RT: [I - i_b, I + i_e]$, $RL: [J - n_{ib}, J + n_{ie}]$. RT and RL are the time and space range of traffic impact of the accident, respectively. I and J are the time and location of the accident respectively; i_b and i_e are the start and end time of traffic impact caused by the accident respectively; n_{ie} and n_{ib} are the upper and lower limits of the spatial impact of accidents on road section traffic respectively. Then the indirect loss model of the impact of the accident on road section traffic is

$$d = \sum_{i=i_b}^{i_e} \sum_{j=n_{ie}}^{j=n_{ib}} \left[3600 \times \left(\frac{l(i,j)}{v(i,j)} - \frac{l(i,j)}{V(i,j)} \right) \right] \tag{12}$$

ISSN:2790-1688 DOI: 10.56028/aetr.4.1.648.2023

$$Q = n_1 Q_{ocp} + n_2 Q_c = n_1 \sum_{i=i_b}^{i_h} Q_m t_i + (n - n_1) \sum_{i=i_b}^{i_e} Q_i t_i \tag{13}$$

$$Q_i = \begin{cases} (Q_m - Q_{v_i}) - (Q_m - Q_{V_i}) \\ Q_m - Q_{v_i} \end{cases} \tag{14}$$

Where: d is the delay time caused by the accident, h; $l(i,j)$ is the length of road section j, km; $v(i,j)$ is the average speed of road section j at time period i, km/h; Q is the delayed traffic volume due to an accident, veh; Q_{ocp} and Q_c are respectively the delayed traffic volume caused by lane occupied of accident vehicle and accident congestion, veh/l; n_1 and n_2 are the number of lanes occupied by accidents and the number of remaining lanes; n is the number of one-way lanes of road section; Q_m is the lane capacity, veh/h/l; i_h is the completion time of traffic accident treatment; Q_{v_i} and Q_{V_i} are the traffic volume corresponding to the minimum speed $v_i = \min\{v(i,j)\}$ of the road section in case of accidents and the traffic volume corresponding to the minimum speed $V_i = \min\{V(i,j)\}$ of the road section in case of frequent congestion, veh/h/l, respectively.

4. Case analysis

Taking the Beijing Fourth Ring Expressway as the research object, according to the OBD data of the expressway from 6:00 to 22:00, taking the adjacent entrance (exit) ramps and exit (entrance) ramps as the starting and ending points of the basic sections, the basic sections were divided into $L_i (i = 1, 2, \cdots, n)$ and n as the total number of basic sections. Sub-sections are divided with a critical length of 250m, as shown in Table 1.

Table 1.Division of basic sections and sub-sections of Beijing Fourth Ring Road

Road length, l	(0,250]	(250,500]	(500,750]	(750,1000]	[1000,+∞)
Number of road sections	53	67	40	36	32
Number of sub-sections	53	134	120	144	160

2863 OBD data are used for map matching. Fig.1 shows the ArcGIS map of vehicle tracking location information collected from a randomly selected complete trip. Fig.2 shows the average speed of a road section of the Fourth Ring Road from 6:00 to 22:00 in the directions of the inner and outer rings in a randomly selected day.

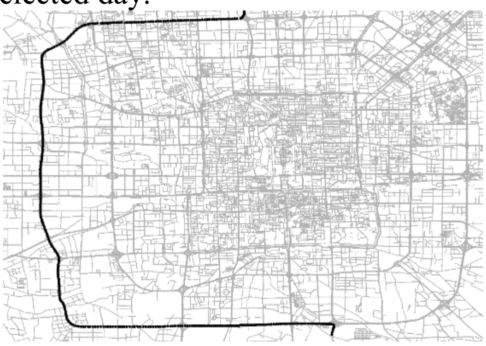

Fig. 1 One travel positioning point matchingeffect of Beijing Fourth Ring Road

ISSN:2790-1688 DOI: 10.56028/aetr.4.1.648.2023

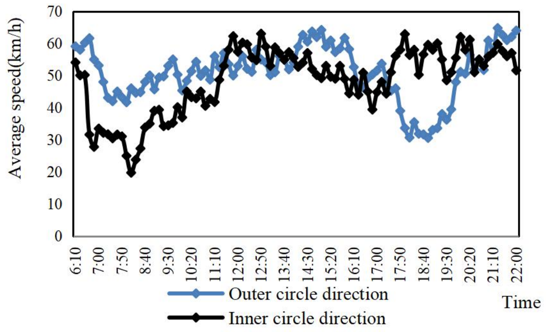

Fig.2 Road section average speed during 6:00-22:00

4.1 Division of traffic operation state.

The FCM clustering algorithm is used to cluster the average speed of road sections with the number of clustering categories $c=6$, fuzzy weighted index $\alpha=2$ [5], and the iterative stopping condition threshold $\varepsilon=10^{-6}$. The traffic operation status levels are classified into 6 levels: very smooth traffic, smooth traffic, stable passing, relative congestion, congestion and serious congestion. The dataset of section average speed , $V=\left\{v_{1,1,1}, v_{1,1,2}, \cdots, v_{i,j,k}, \cdots, v_{a,b,c}\right\}$, i , j , k is the serial number of time, road section and date of v, respectively, is constructed. The final clustering center, $\mu=[26.2970, 65.1431, 41.0423, 51.5677, 58.2267, 10.4564]$, is obtained using FCM. The results of traffic operation status discrimination for randomly selected samples are shown in Table 2.

Table 2. Traffic operation status level fuzzy range

Status level	Ⅰ	Ⅱ	Ⅲ	Ⅳ	Ⅴ	Ⅵ
Fuzzy range	(61,+∞)	(55,61]	(45,55]	(34,45]	(18,34]	[0,18]

4.2 Accident traffic impact attribute identification.

In this paper, the composition conditions of frequent congestion section are defined as: road section congestion level $g(i,j)>3$ and weekly congestion frequency $f(i,j)>3$;The time impact range before and after the accident was demarcated with the measurement unit of 10min, and the accident space impact range was demarcated with the road length of 200m. Traffic impact of the accident $RT:[I-3, I+12]$, $RL:[J-10, J+20]$ [6].

The traffic delay time and the delayed vehicle numbers due to traffic accident of Beijing Fourth Ring Road, respectively, is shown in Fig.4 and Fig.5.

(1) More lanes are occupied by accident, traffic delay time increases. In the peak time, traffic delay time due to the frequent congestion section accident will increase, respectively, 0.93h(1lane occupied), 1.31h(2 lanes occupied) and 1.74h(3 lanes occupied).

(2) Traffic delay caused by the peak time accident is more than the normal time accident. When 1 lane is occupied in the peak time by the non-frequent congestion section accident, the additional 0.28h traffic delay is increased.

(3) Traffic delay caused by the frequent congestion section accident is more than the non-frequent congestion section accident. When 1 lane is occupied by the peak time accident on the frequent congestion section, the additional 0.23h traffic delay is increased.

(4) Accident occupying lane is the most important factor of vehicle delay. In the peak time, the delayed vehicle numbers due to the frequent congestion section accident will increase 931vehicles. However, the delayed vehicle number due to traffic congestion are 163 vehicles.

(5) The delayed vehicle numbers caused by the peak time accident are more than the normal time accident. When 2 lanes are occupied in the peak time by the non-frequent congestion section accident, the additional 1296 vehicles will be delayed.

ISSN:2790-1688 DOI: 10.56028/aetr.4.1.648.2023

(6) The delayed vehicle numbers caused by the frequent congestion section accident are more than the non-frequent congestion section accident. When 1 lane is occupied by the peak time accident on the frequent congestion section, the additional 544 vehicles will be delayed.

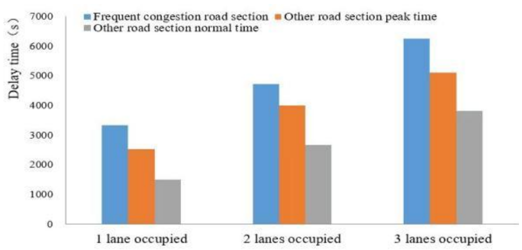

Fig. 3 Average delay time of traffic accident

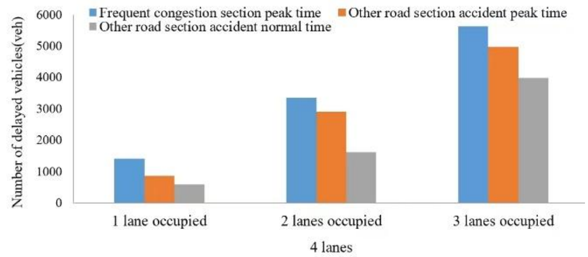

Fig. 4 Delayed vehicles numbers of traffic accident

Conclusion

(1) This paper establishes a map matching algorithm with point-to-lane distance to determine the location of the matching road section of the location point, and uses OBD data for map matching to obtain information such as the instantaneous speed of the location point.

(2) The location-time interpolation method is used to calculate the road section average speed. A kind of FCM clustering discriminant algorithm is constructed to classify the traffic operation status, and used to. The operation status of Beijing Fourth Ring Road, which is classified into very smooth traffic, smooth traffic, stable passing, relative congestion, congestion and serious congestion.

(3) An identification algorithm of the traffic attribute impacts on accident road sections is constructed. Example analysis shows that the accident vehicle occupying lane is the most important factor of vehicle delay. And more lanes are occupied, the delayed vehicle numbers are more and the delay time will increase. Traffic delay and the delayed vehicle numbers caused by the peak time accident are more than the normal time accident. Traffic delay and the delayed vehicle numbers caused by the frequent congestion section accident are more than the non-frequent congestion section accident.

References

[1] He Yaqin, Rong Yulun and Liu Zupeng, “Classification method of urban emergency traffic impact level based on quantile speed,” Transportation System Engineering and Information, vol. 19, Dec, 2019, pp. 163-167+175.

[2] Yu Guizhen, Liu Yumin, Wang Yunpeng, “Analysis of traffic accident temporal and spatial impact based on shockwave theory,” CICTP 2012: Multimodal Transportation Systems-Convenient, 2012, pp. 557-567.

[3] Tang Jinjun, Liu Xinyuan, Ji Ke and Ye Junqing, “Influence Range Estimation of Traffic Accidents in Urban Road Network Considering Lane queuing characteristics,” Journal of Railway Science and Engineering, vol. 19, May, 2022, pp. 2541-2551.

[4] Xiao Huihui, Lu Jian, Qiu Hongtong and Li Yaping, “Prediction model of influence area of Urban Expressway Traffic Events,” Journal of Harbin Institute of Technology, vol. 48, Sep, 2016, pp. 54-59.

[5] Xu Rui and Wunsch li Donald, “Survey of clustering algorithms,” IEEE Transactions on Neural Networks, vol. 16, May, 2005, pp. 645-678.

[6] Hu Xiaojian, Wang Wei and Lu Jian, “A method for Determining the Influence Area of Road Traffic Accidents,” Journal of Southeast University, vol. 37, Sep, 2007, pp. 934-939.