Traffic Congestion Prediction and Assessment Based on Multi-Metric Fuzzy Comprehensive Evaluation (Postprint)

Authors: Yan Yuchan, Bai Lin, Wu Qisheng, Ye Zhen

Date: 2018-10-11T00:00:00+00:00

Abstract

To address the issue that various traffic time periods exert different influences on traffic congestion and that single factors cannot accurately characterize traffic congestion states, this paper proposes a traffic congestion evaluation and prediction method employing multi-index fuzzy comprehensive evaluation. The method utilizes particle swarm optimization to optimize a support vector regression machine for forecasting average road speed and traffic flow, thereby obtaining predicted values for three factor indicators: average speed v, traffic flow density D, and road saturation S. These three factor indicators are subsequently input into a multi-index fuzzy comprehensive evaluation model, which first establishes the factor set and evaluation set for traffic congestion states. The entropy method is employed to determine the weight coefficients of the three factor indicators during morning peak, evening peak, and other time periods, while a trapezoidal membership function is used to determine the membership degree of each indicator across different time periods, ultimately classifying traffic congestion states into six levels. Experimental verification through predictive evaluation on traffic data from the I-405 freeway in the United States PeMS database demonstrates that the traffic congestion states predicted by this method essentially coincide with actual states, achieving high prediction accuracy with a correctness rate of 94.79%.

Full Text

Preamble

Title: Traffic Congestion Prediction and Assessment Based on Multi-Index Fuzzy Comprehensive Evaluation

Authors: Yan Yuchan†, Bai Lin, Wu Qisheng, Ye Zhen
(School of Electronics & Control Engineering, Chang' an University, Xi' an

710064, China)

Abstract: To address the problem that different traffic periods have varying impacts on congestion and that single factors cannot accurately characterize traffic congestion states, this paper proposes a traffic congestion evaluation and prediction method using multi-index fuzzy comprehensive evaluation. The method employs particle swarm optimization to optimize a support vector regression model for predicting road average speed and traffic flow, obtaining predicted values for three factor indicators: average speed v, traffic flow density D, and road saturation S. These three indicators are then input into a multi-index fuzzy comprehensive evaluation model. First, the factor set and evaluation set for traffic congestion states are established. The weight coefficients for the three indicators during morning peak, evening peak, and other periods are determined through the entropy method, and the membership degree of each indicator in each period is determined through a trapezoidal membership function. Finally, traffic congestion states are divided into six levels. Experimental results from predictive evaluation on traffic data from the I405 highway in the U.S. PeMS database demonstrate that the traffic congestion states predicted by this method basically match the actual states, achieving high prediction accuracy with a correct rate of 94.79%.

Keywords: traffic congestion; multi-index fuzzy comprehensive evaluation; factor index; entropy method; trapezoidal membership function

---

0 Introduction

Increasingly severe traffic congestion has significantly impacted citizens' travel time for work and study. According to data analysis from the Beijing Transport Research Center [2], a general correspondence between residents' travel time expenditure and traffic congestion states has been established. Beijing classifies congestion states into five levels, as shown in [Figure 1: see original paper]. The U.S. *Highway Capacity Manual* (HCM) [1] classifies traffic congestion states into six levels (1: very smooth, 2: smooth, 3: mild congestion, 4: moderate congestion, 5: severe congestion, 6: gridlock), where higher values indicate more severe congestion.

When the traffic state is very smooth, it indicates virtually no congestion on the road, and vehicles can travel at the posted speed limit. When the traffic state is basically smooth, residents need to spend 0.3–0.5 times more travel time. For mild congestion, travel time increases by 0.5–0.8 times. For moderate congestion, travel time increases by 0.8–1.0 times. For severe congestion, travel time increases by more than 1.0 times. Evidently, as traffic conditions worsen, the impact on residents' work becomes more severe. By observing changes in traffic congestion levels, residents can understand overall road congestion conditions, anticipate their travel time, and choose alternative routes or modes of transportation to effectively avoid congestion.

1 Selection of Traffic Congestion Evaluation Factor Indicators

For traffic congestion state prediction and evaluation, the first step is to select factor indicators that can accurately and effectively characterize traffic congestion states. The selection principles include overall completeness, objectivity, operability, and comparability [15]. However, traffic congestion conditions cannot be evaluated using only a single traffic flow parameter. While average vehicle speed can intuitively represent traffic congestion states, when vehicles wait at red lights, their speed is very low, which does not necessarily indicate road congestion. Therefore, this paper selects three traffic flow parameters as factor indicators: traffic flow average speed, traffic flow density, and road segment saturation.

Traffic flow average speed refers to the average distance traveled by all vehicles on a road segment within a unit time. This indicator directly reflects current road traffic congestion conditions—generally, higher speed indicates smoother traffic, while lower speed indicates more severe congestion. The calculation formula is:

$$v = \frac{1}{N}\sum_{i=1}^{N} v_i$$

where v is the traffic flow average speed (km/h), N is the total number of vehicles on the road within a unit time, and v_i is the instantaneous speed of the i-th vehicle.

Traffic flow density refers to the total number of vehicles per unit length on a road segment within a unit time. When roads are congested and vehicles are stationary, traffic flow changes are nearly zero, but traffic density becomes very large, playing a decisive role in characterizing congestion states. The calculation method is:

$$D = \frac{f}{v}$$

where D is the traffic flow density (vehicles/km), f is the hourly monitored traffic flow, and v is the average speed.

Road segment saturation refers to the ratio of actual traffic flow to maximum capacity on a road segment, reflecting the actual load capacity of the road. The calculation method is:

$$S = \frac{V}{C}$$

where S is the road saturation (dimensionless), V is the currently collected traffic flow on the road, and C is the maximum road capacity.

Current parameters for measuring traffic congestion states mainly include traffic flow parameters, driver perception parameters, and vehicle queue length. Traffic flow parameters primarily refer to parameters reflecting traffic flow characteristics such as traffic volume, average speed, traffic flow density, or occupancy rate. Traffic congestion state evaluation is divided into single-factor evaluation and multi-factor evaluation. Single-factor evaluation uses only one parameter to assess traffic congestion states. For single-factor evaluation, Shi et al. [3] proposed a spatial average speed estimation method with exponential smoothing for real-time traffic congestion assessment. Castro et al. [4] calculated the capacity of each road segment using historical traffic flow density and speed data extracted from taxi trajectories, then classified congestion density levels by segment capacity for congestion evaluation. Yang [5] applied a weighted improved GM(1,1)-Markov prediction model for speed prediction and compared prediction results with decision thresholds to determine traffic congestion states.

Due to the complexity of road conditions, single traffic flow parameters are often insufficient to characterize the congestion state of an entire road system. To improve evaluation accuracy, scholars have proposed multi-parameter characterization of traffic congestion states. Long et al. [6] extracted three traffic parameters—segment average travel speed, average delay per unit distance, and segment saturation—from traffic data, then applied a fuzzy comprehensive evaluation model for traffic congestion assessment. Tan et al. [7] constructed traffic flow feature vectors from basic data including traffic flow parameters, environmental states, and time periods, then used Softmax regression for multi-state traffic congestion prediction. Li [8] combined the information entropy weighting method with grey system theory to construct an urban road traffic congestion evaluation system including intersection indicators, segment indicators, and regional indicators. Huang et al. [9] established a new road network traffic congestion state characterization model based on multivariate set pair analysis, achieved traffic information fusion through D-S evidence theory, and derived current traffic congestion states.

---

2 Multi-Index Fuzzy Comprehensive Evaluation Model

This paper divides traffic congestion states into six levels: 1 (very smooth), 2 (smooth), 3 (mild congestion), 4 (moderate congestion), 5 (severe congestion), and 6 (gridlock). When analyzing traffic congestion states under multi-factor indicators, there is no clear division boundary for congestion levels corresponding to each factor—this is a fuzzy concept that can be effectively addressed by multi-index fuzzy comprehensive evaluation. Considering the different temporal characteristics of traffic flow parameters during morning peak, evening peak, and other periods, and the problem that single-factor indicators cannot accurately characterize traffic congestion states, this paper first uses a PSO-optimized SVR prediction model [10–14] to predict the required values for three factor indicators. Through multi-index fuzzy comprehensive evaluation, three different time periods are analyzed for traffic congestion, where the entropy method determines the weight coefficients for each indicator in each period, and traffic conditions are finally divided into six levels.

According to the *Highway Capacity Manual* and combined with the road data studied in this paper, the value ranges corresponding to traffic congestion states for the three factor indicators are clearly defined, as shown in .

2.1 Establishing Factor and Evaluation Sets for Traffic Congestion States

The factor set for traffic congestion states is established as $U = \{u_1, u_2, \dots, u_i\}$, where u_i represents the i-th factor indicator. This paper selects three factors: traffic flow average speed, traffic flow density, and road saturation. The evaluation (level) set is determined as $V = \{v_1, v_2, \dots, v_j\}$, where v_j represents the j-th level value. This paper divides traffic congestion states into six levels.

2.2 Determining Weight Coefficients Based on the Entropy Method

The entropy method determines the dispersion degree of each indicator through entropy value calculation. A larger entropy value indicates smaller dispersion of the indicator and less influence on comprehensive analysis, and vice versa. By analyzing various traffic flow parameter data from the California PeMS database, traffic flow during 6:00–8:00 and 17:00–19:00 shows obvious increases over time, then stabilizes, while average speed drops sharply, as shown in [Figure 2: see original paper]. Traffic congestion states basically occur during morning and evening peak periods. Therefore, this paper employs three weight coefficients to analyze overall road traffic congestion conditions: morning peak, evening peak, and normal periods.

The steps for determining weights using the entropy method are as follows. First, calculate the morning peak factor indicator weights:

a) Select n samples representing different congestion indices during the 6:00–8:00 period with m influencing indicators to form the original factor data matrix $X = \{x_{ij}\}_{n\times m}$, where x_{ij} represents the j-th influencing factor of the i-th sample.

b) Normalize the influencing factors. Factors include positive and negative types. For example, speed is a negative factor—higher speed indicates lower congestion index, while lower speed indicates higher congestion index. Different normalization algorithms are applied to different influencing factors. Without normalization, negative factors would yield smaller weights, failing to reflect their influence on comprehensive analysis.

For positive influencing factors:

$$x'_{ij} = \frac{x_{ij} - \min(x_{1j}, \dots, x_{nj})}{\max(x_{1j}, \dots, x_{nj}) - \min(x_{1j}, \dots, x_{nj})}$$

For negative influencing factors:

$$x'_{ij} = \frac{\max(x_{1j}, \dots, x_{nj}) - x_{ij}}{\max(x_{1j}, \dots, x_{nj}) - \min(x_{1j}, \dots, x_{nj})}$$

c) Calculate the proportion of the j-th factor of the i-th traffic sample to the total of that factor:

$$p_{ij} = \frac{x'_{ij}}{\sum_{i=1}^{n} x'_{ij}}$$

d) Calculate the entropy value of the j-th factor:

$$e_j = -k \sum_{i=1}^{n} p_{ij} \ln(p_{ij})$$

where k is related to sample size n, with $k = \frac{1}{\ln(n)} > 0$, and $e_j \geq 0$.

e) Calculate the difference coefficient of the j-th factor. A larger difference coefficient indicates greater influence of that factor on comprehensive analysis:

$$g_j = 1 - e_j$$

f) Calculate the weight of each factor:

$$w_j = \frac{g_j}{\sum_{j=1}^{m} g_j}$$

Similarly, weight coefficients for evening peak and other periods can be calculated, with the sum of weights equaling 1.

2.3 Determining Factor Indicator Membership Degrees

The single-factor evaluation matrix is obtained through membership functions, which calculate the membership degree r_{ij} of each factor u_i to the j-th level value. This step is crucial in the fuzzy comprehensive evaluation method. Since traffic congestion states correspond to each factor within certain ranges, and proximity to range critical values better indicates whether a sample belongs to a state, trapezoidal membership functions are adopted to reduce subjectivity. For the two extreme states (very smooth and gridlock), semi-descending and semi-ascending trapezoidal membership functions are used, respectively. The membership function graph is shown in [Figure 3: see original paper].

The horizontal axis represents the values of each factor indicator for traffic samples, while $k_1, k_2, \ldots, k_{10}$ represent threshold ranges for each factor indicator. The vertical axis uses linear values near critical thresholds of factor indicators, where 0 indicates the factor does not belong to the congestion state and 1 indicates it does. For traffic flow density and road saturation (positive factors), the functions from left to right represent states: very smooth, smooth, to gridlock. Since average speed is a negative indicator, the states from left to right are gridlock, severe congestion, to very smooth.

The calculation formulas for each state are:

State 1 (Very Smooth):

$$f(x) = \begin{cases} 1 & x \le k_1 \\ \frac{k_2 - x}{k_2 - k_1} & k_1 < x < k_2 \\ 0 & x \ge k_2 \end{cases}$$

States 2, 3, 4, 5:

$$f_i(x) = \begin{cases} 0 & x \le k_{2i-3} \text{ or } x \ge k_{2i+2} \\ \frac{x - k_{2i-3}}{k_{2i-2} - k_{2i-3}} & k_{2i-3} < x < k_{2i-2} \\ 1 & k_{2i-2} \le x \le k_{2i+1} \\ \frac{k_{2i+2} - x}{k_{2i+2} - k_{2i+1}} & k_{2i+1} < x < k_{2i+2} \end{cases}$$

for $i = 2, 3, 4, 5$.

State 6 (Gridlock):

$$f(x) = \begin{cases} 0 & x \le k_9 \\ \frac{x - k_9}{k_{10} - k_9} & k_9 < x < k_{10} \\ 1 & x \ge k_{10} \end{cases}$$

2.4 Determining Congestion Level Under Multi-Factor Indicators

The fuzzy comprehensive evaluation matrix is obtained through fuzzy composition operations:

$$B = W \circ R = (w_1, w_2, w_3) \circ \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1j} \\ r_{21} & r_{22} & \cdots & r_{2j} \\ r_{31} & r_{32} & \cdots & r_{3j} \end{pmatrix}$$

where W_a, W_b, and W_c represent the weight coefficients for morning peak (6:00-8:00), evening peak (17:00-19:00), and other periods, respectively.

According to the maximum membership degree principle, the level corresponding to the maximum membership value in the fuzzy comprehensive evaluation matrix is selected as the final congestion level:

$$b_j = \max\{b_1, b_2, \ldots, b_j\}$$

If the final maximum membership value is b_j, then the traffic sample at that moment belongs to traffic level j.

---

3 Traffic Congestion State Prediction Based on Multi-Index Fuzzy Comprehensive Evaluation

The traffic congestion prediction system based on fuzzy comprehensive evaluation consists of three components: data parameter acquisition, traffic congestion prediction, and traffic congestion index evaluation. The data parameter acquisition component mainly obtains traffic flow at time t and *t-i* $(F(t), F(t-i))$ and average speed at time t and *t-i$* $(V(t), V(t-i))$*. The traffic congestion prediction component primarily includes traffic flow parameter prediction. This paper selects the method of using Particle Swarm Optimization (PSO) to optimize Support Vector Regression (SVR) for predicting traffic flow and average speed parameters. This method uses current time* t* and historical time *t-i* data to predict future time *t+i* data. According to the formulas mentioned, the predicted traffic flow parameters are used to calculate the density D and road saturation S required for the fuzzy comprehensive evaluation. The traffic congestion index evaluation component obtains the final traffic congestion state through the entropy method and fuzzy comprehensive evaluation method. The system framework is shown in [Figure 4: see original paper].

---

4 Experimental Results and Analysis

4.1 Traffic Flow Parameter Prediction

This paper selects traffic flow data from highway I405 in the California PeMS system database for experimental analysis, with the selected period from March 1, 2018 to March 4, 2018. The data sampling period is 5 minutes, totaling 1,152 data points. The first three days of traffic flow serve as training data for the PSO-SVR prediction model, and the last day serves as test data, resulting in 864 training samples and 288 test samples. Before prediction, data normalization to [0,1] is required, followed by denormalization after prediction. This effectively reduces computation time and improves standardization. This paper adopts extremum-based normalization:

$$X' = \frac{X - X_{\min}}{X_{\max} - X_{\min}}$$

where $X_{\min}$ and $X_{\max}$ are the minimum and maximum values of the training samples.

For establishing the PSO-SVR traffic flow prediction model, current time t is determined. The training sample inputs are traffic flow and average speed at times *t-1* and *t-2* ($F(t-1), F(t-2), V(t-1), V(t-2)$), and the training set output is traffic flow at time *t+1* ($F(t+1)$). Through PSO optimization, the optimal penalty factor c and kernel function parameter g for SVR are obtained, thereby finding support vectors and the prediction function. Finally, test set inputs at times *t-1* and *t-2* are used to predict traffic flow at time *t+1$. Similarly, for average speed prediction, only the model output needs to be changed to average speed.

The traffic flow and average speed prediction results obtained through PSO-SVR are shown in Figure 5: see original paper and Figure 6: see original paper. To more intuitively demonstrate prediction accuracy, evaluation metrics of Mean Square Error (MSE) and Mean Absolute Percentage Error (MAPE) are selected for analysis:

$$\mathrm{MSE} = \mathrm{mean}(e_t^2), \quad e_t = y_t - \hat{y}_t$$

$$\mathrm{MAPE} = \mathrm{mean}(p_t), \quad p_t = \left|\frac{e_t}{y_t}\right| \times 100\%$$

where e_t is model error, y_t is actual value, and $\hat{y}_t$ is predicted value.

The corresponding MSE error histograms are shown in Figure 5: see original paper and Figure 6: see original paper. The MSE and MAPE for traffic flow and average speed are 36.8951, 14.4882% and 21.6691, 9.7387%, respectively.

According to the U.S. *Highway Capacity Manual*, the maximum capacity of a multi-lane highway with a design speed of 75 km/h is 2,200 pcu/h/lane. Since the data sampling period is 5 minutes, the maximum capacity per lane is 180 PCU. Therefore, the required factor indicators—traffic flow density D and road saturation S—for fuzzy comprehensive evaluation can be calculated from the predicted traffic flow and average speed using equations (2) and (3).

4.2 Traffic Congestion Level Evaluation

According to the fuzzy comprehensive evaluation method steps, the factor set and evaluation (level) set for the traffic congestion index are first established. The entropy method determines the weight coefficients for the three factor indicators corresponding to morning peak, evening peak, and other periods. The calculated weights are:

- Morning peak: $W_a = [0.2273, 0.5788, 0.1938]$
- Evening peak: $W_b = [0.2926, 0.5351, 0.1723]$
- Other periods: $W_c = [0.1676, 0.4530, 0.3794]$

The trapezoidal membership function is determined based on the value ranges of the three traffic factor indicators shown in Table 1 to obtain the membership degree of each factor indicator. Finally, according to the maximum membership degree principle, the traffic congestion level is determined to obtain the predicted traffic congestion state. The comparison between predicted and actual traffic congestion levels is shown in [Figure 7: see original paper].

In the figure, each level represents: 1 (very smooth), 2 (smooth), 3 (mild congestion), 4 (moderate congestion), 5 (severe congestion), and 6 (gridlock). The results demonstrate that the proposed multi-index fuzzy comprehensive evaluation method can accurately calculate the final congestion level based on multiple indicators and their different impacts on traffic congestion states. The traffic congestion levels analyzed from predicted values basically match actual congestion levels with similar trends. During morning peak periods, congestion levels rise significantly, with gridlock occurring; after morning peak, there is obvious relief. Evening peak congestion is generally slightly less severe than morning peak. Among 288 predicted data points, only 15 time points have mismatched congestion levels, achieving an accuracy rate of 94.79%.

---

5 Conclusion

This paper proposes a method for traffic congestion state prediction and assessment using multi-index fuzzy comprehensive evaluation. The method addresses the problem that traffic flow parameters have different impacts on congestion states during morning peak, evening peak, and other periods, and that single-factor indicators cannot accurately characterize traffic congestion states. The entropy method determines the weight coefficients of each factor indicator in different periods, and the three factor indicators are input into a multi-index fuzzy comprehensive evaluation model to make the fuzzy evaluation results more accurate. For prediction, PSO is used to optimize SVR, reducing the error between predicted and actual values of the two traffic flow parameters. Through predictive evaluation experiments on California traffic data, the predicted average speed and traffic flow are used in the multi-index fuzzy comprehensive evaluation method for traffic congestion level prediction. The results show that the proposed method' s predicted traffic congestion states match well with actual states, achieving an accuracy rate of 94.79%.

References

[1] Transportation Research Board. *Highway capacity manual* [M]. Trans. Ren Futian, Liu Xiaoming, Rong Jian, et al. Beijing: China Communication Press, 2007: 116-189.

[2] Beijing Transport Institute [EB/OL]. [2018-06-01]. http://www.bjtrc.org.cn/S_IndexPage.aspx.

[3] Shi W, Liu Y. Real-time urban traffic monitoring with global positioning system-equipped vehicles [J]. *IET Intelligent Transport Systems*, 2010, 4(2):

113-120.

[4] Castro P S, Zhang Daqing, Li Shijian. Urban traffic modeling and prediction using large scale taxi GPS traces [C]// Proc of the 10th International Conference on Pervasive Computing. Newcastle, UK: [s.n.], 2012: 57-72.

[5] Yang Junying. Study on the identification and prediction of the traffic congestion state for urban road [D]. Chengtu: Southwest Jiaotong University, 2014.

[6] Long Xiaoqiang, Tan Yunlong. Urban traffic congestion evaluation based on fuzzy comprehensive evaluation [J]. *Transport Standardize*, 2011(11): 114-117.

[7] Tan Juan, Wang Shengchun. Research on prediction model for traffic congestion based on deep learning [J]. *Application Research of Computers*, 2015(10): 2951-2954.

[8] Li Fuzhong. Urban road traffic congestion evaluation based on grey correlation model [J]. *Chongqing Architecture*, 2018(17): 38-40.

[9] Huang Darong, Chai Yanchong, Zhao Ling, et al. Traffic congestion status identification method for road network with multi-source uncertain information [J]. *Acta Automatica Sinica*, 2018, 44(3): 533-544.

[10] Sun Zhanquan, Geoffrey Fox. Traffic flow forecasting based on combination of multidimensional scaling and SVM [J]. *International Journal of Intelligent Transportation Systems Research*, 2014, 12(1): 1-13.

[11] Tian Ruijie, Zhang Weishi, Zhai Huawei. Short-term traffic flow velocity prediction model based on time series and BP-ANN [J/OL]. *Application Research of Computers*, 2019, 36(11). [2018-08-10]. http://www.arocmag.com/article/02.

[12] Hong W C, Dong Yucheng, Zheng Feifeng, et al. Forecasting urban traffic flow by SVR with continuous ACO [J]. *Applied Mathematical Modelling*, 2011, 35(3): 1282-1291.

[13] Luo Xianglong, Jiao Qinqin, Zhu Liyao, et al. Short-term traffic flow prediction based on deep learning [J]. *Application Research of Computers*, 2017, 34(1): 91-97.

[14] Chen Xiaobo, Yang Jian, Liang Jun. A flexible support vector machine for regression [J]. *Neural Computing and Applications*, 2012, 21(8): 1891-1898.

[15] Zhu Fuling. Study on the evaluation index system of urban road traffic congestion [D]. Nanjing: Southeast University, 2007.

Note: Figure translations are in progress. See original paper for figures.