Journal of Engineering, Project, and Production Management
2022, 12(2), 108-115

Road Traffic Status Prediction Approach Based on Kmeans-Decision Tree Model

Xinghua Hu[1], Xinghui Chen[2], Wei Liu[3], and Gao Dai[4]
[1]Professor, College of Traffic & Transportation, Chongqing Jiaotong University, Nan'an District, Chongqing, China. E-mail: xhhoo@cqjtu.edu.cn (corresponding author).
[2]Graduate Student, College of Traffic & Transportation, Chongqing Jiaotong University, Chongqing, China. E-mail: Hui981008@163.com
[3]Professor, College of Traffic & Transportation, Chongqing Jiaotong University, Chongqing, China. E-mail: neway119@qq.com
[4]Senior Engineer, Chongqing Ulit Science & Technology Co., Ltd, Chongqing. China, E-mail: 1090685519@qq.com

Engineering Management
Received September 10, 2021; revised December 9, 202; accepted December 12, 2021
Available online December 27, 2021

Abstract: An effective way to solve the problem of urban traffic congestion is to predict the road traffic status accurately and take effective traffic control measures in time. Considering the impact of visibility on traffic, the pavement status and time characteristics were finely divided, and a regression decision tree was used to establish the traffic flow velocity prediction model with pavement status, time characteristics, and working day characteristics as characteristic parameters. Furthermore, based on the perspective of avoiding using velocity as a single parameter to classify the road traffic status levels, the Kmeans clustering algorithm was used to obtain the classification label results. Moreover, the traffic flow velocity and pavement status were used as characteristic parameters of the classification decision tree to establish the multi-parameter road traffic status prediction model. The experimental result showed that the prediction accuracy of the proposed road traffic status prediction model was 81.31%, and this method has good applicability and certain application value for road traffic status prediction.

Keywords: Traffic engineering, traffic flow velocity, road traffic status prediction, Kmeans clustering, decision tree.

DOI 10.32738/JEPPM-2022-0010

1. Introduction

In recent years, with the increase of motor vehicle ownership in China, the existing road resources can no longer meet the growth of the number of motor vehicles. Traffic congestion has become a serious problem in the majority of urban developments in China. How to accurately predict the traffic status, guide drivers to choose appropriate travel routes, and avoid the occurrence of traffic congestion has become the key to solve the problem. Aifadopoulou et al. (2018) used flow and speed as index parameters and used a neural network method to predict short-term traffic status. Huang et al. (2020) developed a traffic congestion prediction system based on road conditions using flow and velocity as index parameters. Nguyen et al. (2020) proposed a deep learning method based on a 3D Convolutional Neural Network, which used a large number of urban remote sensing data sources in three-dimensional raster images to predict traffic congestion. Alghamdi et al. (2019) proposed a traffic status prediction method by establishing the ARIMA model. Zaki et al. (2020) established a traffic status prediction model based on a Hidden Markov Model with average velocity and contrast as index parameters. Minh et al. (2019) identified and predicted traffic conditions by using data shared by mobile devices and traffic flow data collected by fixed sensor systems. Sunindyo and Satria (2020) proposed a traffic status prediction method based on the multi-layer perceptron and long short-term memory (LSTM) neural network. Kumar and Sivanandan (2019) established a regression model to predict the congestion index (CI) of various types of vehicles by considering the bus congestion index, lane width, and signalized intersection as independent variables. Hao et al. (2020) established an urban road traffic status prediction model based on a deep recursive Q-learning method, which was based on an optimized LSTM neural network. Ji and Hong (2019) used a deep learning method to predict the velocity of traffic flow based on real-time long-term evolution (LTE) data. Miglani and Kumar (2019) studied the prediction of traffic flow parameters using various deep learning models in the field of autonomous driving and compared the applicability of these models in intelligent transportation systems. Guo et al. (2018) combined three prediction methods, i.e., neural network, support vector regression, and random forest, to predict short-term traffic flow under normal conditions and traffic incidents. Liu et al. (2018) used k-nearest neighbor and support vector regression (KNN-SVR) to establish a traffic flow prediction model. Zhang et al. (2019) used a neural network model with a deep autocoder to achieve a short-time prediction of traffic status. Quang and Bae (2021) propose a hybrid depth convolution neural network (CNN) method to predict the short-term traffic congestion index in the urban network. It can be observed from the above-mentioned studies that in the selection of traffic status prediction models, the majority of studies used autoregressive integrated moving average model (ARIMA), random forest, and LSTM neural networks. However, owing to the large amount of data required and the complexity of the models, the computer requires a large amount of memory. Furthermore, the processing speed is relatively low. In the selection of traffic status prediction parameters, the majority of studies consider the traffic flow velocity as the judgment standard. Few studies considered parameter indicators, such as pavement status and time characteristics, but these indicators often have a certain impact on the discrimination of traffic status. If the road traffic status is classified based on velocity only, the classification result will be inaccurate. However, due to the low complexity of the decision tree model, the amount of data required is not particularly large, which makes the processing speed of the computer faster, and the prediction accuracy of the model is also high. Compared with these models proposed above, this model has a better application value. Therefore, this study intends to use the road traffic status prediction method of the Kmeans-decision tree model to classify and predict the traffic status of urban roads.

2. System framework of road traffic status prediction approach

Accurate and efficient prediction of road traffic status is an effective way to improve the commuting efficiency of urban residents. By using a road traffic status prediction approach based on the Kmeans-decision tree model, this study firstly introduced characteristic parameter data, such as pavement status, time characteristics, and working day characteristics., and a regression decision tree was used to establish the traffic flow velocity prediction model. Then, based on the two characteristic parameters of velocity and pavement status, the Kmeans clustering algorithm was used to cluster them to obtain the corresponding category labels. Finally, the velocity and pavement status were used as the characteristic parameters of the model. The road traffic status label obtained using the Kmeans clustering algorithm was used as the output result of the model to establish the classification decision tree model to realize the classification prediction of the road traffic status levels. This specific process is illustrated in Fig. 1.

3. Traffic flow velocity prediction

3.1. Characteristic parameter selection

(1) Pavement status

In rainy weather, owing to the change of the pavement status, the anti-skid coefficient of each status is different, resulting in inconsistent traffic flow velocity. When a vehicle drives over flooded pavements, a water film is formed between the road and the tire. This causes the tire to float, thereby reducing the friction between the tire and the road. Furthermore, rainy weather reduces visibility and interferes with the field of vision of the drivers, thus reducing the overall driving velocity of vehicles.

In this study, the pavement status was divided into three conditions depending on the rainfall intensity: dry pavement, wet pavement, and flooded pavement. When the rainfall was less than 5 mm/h, the pavement was considered to be wet; similarly, when the rainfall was 5 mm/h or more and when there was no rainfall, the pavement was considered to be flooded and dry, respectively (Zhao and Ren, 2017). For example, when vehicles are driving on flooded pavement, the overall velocity of vehicles is reduced compared to dry pavement. At a certain velocity, vehicles driving on flooded pavement will not be congested, while vehicles driving on dry pavement will be congested. Therefore, if road traffic status is predicted only according to the traffic flow velocity, the prediction result will not be accurate enough.

(2) Time characteristics

Owing to the commuting needs of urban residents, the overall velocity of vehicles on some road sections will be different during the morning, evening, and flat peak hours of each working day. Based on the peak and flat peak hours proposed by existing studies, we further added night hours in this study to reflect the impact of visibility on velocity.

(3) Working day characteristics

As the lives of urban residents have a strong regularity, urban residents primarily travel to go to school or work on working days. This is more concentrated in time distribution, and the peak passenger flow is obvious. Similarly, residents primarily travel for leisure and entertainment on non-working days. Thus the peak passenger flow is low. The traffic flow velocity will change correspondingly during the peak hours of the working days.

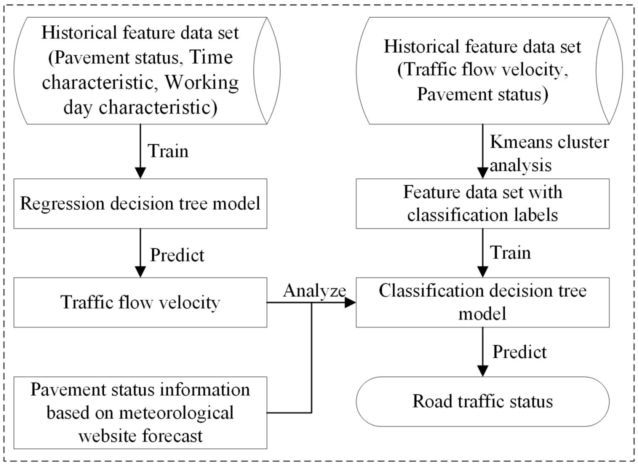

Fig. 1. Model framework of the road traffic status prediction

3.2. Establishment of traffic flow velocity prediction model

The idea of decision tree learning mainly comes from the ID3 algorithm proposed in 1986, the C4.5 algorithm proposed by Quinlan in 1993, and the classification and regression tree (CART) algorithm proposed by Breiman et al. in 1984 (Li, 2019). However, because the ID3 and C4.5 algorithms can only classify and predict data, this study is based on the CART algorithm. Additionally, the aforementioned three characteristic parameters were introduced to establish the regression prediction model of traffic flow velocity. The steps to establish the proposed model are as follows.

(1) Generation of regression decision tree

The process of using the CART algorithm to generate a regression decision tree is the process of recursively constructing a binary decision tree. Because this study is a regression prediction of traffic flow velocity, the regression decision tree is generated with the goal of minimizing the square error, which is usually called the least squares regression tree. The process of generating a regression decision tree is as follows.

Step (1): Choose the optimal segmentation variable j and segmentation point s from the three characteristic parameters of pavement status, time characteristics, and working day characteristics to solve Eq. (1). Browse the variable j, check the segmentation point s for the fixed segmentation variable j, and choose the pair (j, s) that makes Eq. (1) attain the minimum value.

$$\min_{j,s}\left[\min_{c_1}\sum_{x_i\in R_1(j,s)}\left(v_i-c_1\right)^2+\min_{c_2}\sum_{x_i\in R_2(j,s)}\left(v_i-c_2\right)^2\right] \quad (1)$$

In Eq. (1), v_i represents the velocity of each sample, and c_1 and c_2 represent the mean value of the sample velocity on the left and right sides of the corresponding segmentation point s in each segmentation variable j, respectively.

Step (2): Use the selected pair (j, s) to divide the area according to Eq. (2), and then obtain the corresponding velocity value according to Eq. (3).

$$R_1(j,s)=\left\{x \mid x^{(j)}\le s\right\},\ R_2(j,s)=\left\{x \mid x^{(j)}>s\right\} \quad (2)$$

$$\overline{V}=\frac{1}{N_m}\sum_{x_i\in R_m(j,s)} v_i,\ x\in R_m,\ m=1,\ 2 \quad (3)$$

Step (3): Repeat Steps (1) and (2) for the two subregions until the stop condition is met.

Step (4): Divide the input space into M regions, i.e., R_1, R_2, R_M, to generate the traffic flow velocity model:

$$f(x)=\sum_{m=1}^{M}\overline{V}I\left(x\in R_m\right) \quad (4)$$

Where $x \in R_m$, the value of I is 1, otherwise, it is 0.

(2) Regression decision tree pruning

The CART pruning algorithm was used to cut some subtrees from the bottom of the fully grown regression tree to reduce the complexity of the traffic flow velocity prediction model to avoid overfitting. The pruning algorithm is as follows.

Step (1): let k=0, T=$f(x)$.

Step (2): Let α=+∞.

Step (3): Calculate Eq. (5) and (6) for each internal node t from bottom to top:

$$g(t)=\frac{C(t)-C(T_t)}{|T_t|-1} \quad (5)$$

$$\alpha=\min(\alpha,\mathrm{g(t)}) \quad (6)$$

Where T_t represents the subtree with t as the root node, $C(T_t)$ is the prediction error of the training data, and $|T_t|$ is the number of leaf nodes of T_t.

Step (4): Prune the internal node t with $g(t) = \alpha$, and decide the class of the leaf node t using the majority voting method to obtain the tree T.

Step (5): Let k=k+1, α_k=α, and T_k=T.

Step (6): If T_k is not a tree composed of the root node and leaf nodes, repeat from Step (2); otherwise, let $T_k = T_n$.

Step (7): The cross validation method is used to choose the optimal subtree T_α from the subtree sequence T_0, T_1, …, T_n to obtain the pruned traffic flow velocity prediction model T_α.

4. Road Traffic Status Prediction Based on Kmeans-Decision Tree Model

In this study, we adopted the method of the combined model, considered the result of the Kmeans clustering algorithm as the classification label of the road traffic status, and established the classification decision tree model to classify and predict the road traffic status.

4.1. Selection of Characteristic Parameters

When vehicles encounter rainy weather while driving, the pavement status changes with different rainfall intensities. Simultaneously, rainy weather will also interfere with the field of vision of the driver, thereby reducing the overall velocity of the vehicle. Therefore, if the road traffic status level is divided only according to the traffic flow velocity, there will be some deviation. Considering the characteristic parameters of the pavement status, the division results of the road traffic status levels will be more accurate.

4.2. Road Traffic Status Division Based on the Kmeans Clustering Algorithm

According to the provisions of the road traffic operation safety status index in the Urban Road Traffic Operation Evaluation Index System and combined with the established mathematical model, we divided the road traffic status into four categories in this study: no congestion, mild congestion, moderate congestion, and serious congestion. Before Kmeans clustering, the sample data should be normalized to improve the convergence rate of the algorithm and make the clustering result more accurate. The normalization formula is shown in Eq. (7).

$$x_{ij}^{'}=\frac{x_{ij}-x_{j,\min}}{x_{j,\max}-x_{j,\min}} \quad (7)$$

where $x_{j,max}$ and $x_{j,min}$ represent the maximum and minimum values of the j^{th} characteristic parameter, respectively, and x'_{ij} represents the normalized value of the j^{th} characteristic parameter of the i^{th} element, respectively.

The Kmeans clustering algorithm is an iterative process. First, the centers of K classes were selected, and the samples were assigned to the nearest center individually to obtain a clustering result. Then, the mean value of the samples of each class was updated as the new center of the class. The aforementioned steps were repeated until convergence. The specific process is as follows:

Step (1): Initialization. Let t = 0. Randomly select k sample points as the initial clustering center $m^{(0)}=(m_1^{(0)}, \ldots, m_l^{(0)}, \ldots, m_k^{(0)})$, where each sample data is a twodimensional feature vector composed of the velocity and pavement status.

Step (2): Cluster the samples. For the fixed class center $m^{(t)} = (m_1^{(t)}, \ldots, m_l^{(t)}, \ldots, m_k^{(t)})$, calculate the distance from each sample to the class center, assign each sample to the class with the nearest center, and form the clustering result $C^{(t)}$, where $m_l^{(t)}$ is the class center of G_l. The calculation formula is shown in Eq. (8)

$$C^{(t)} = \min_{m_1,\ldots,m_k} \sum_{l=1}^{k} \sum_{C(i)=l} \|x_i - m_l\|^2 \tag{8}$$

Where k represents the number of categories.

Step (3): Calculate the new class center. For the clustering result $C^{(t)}$, calculate the average value of the current velocity and pavement status in each class as the new class center $m^{(t+1)} = (m_1^{(t+1)}, \ldots, m_l^{(t+1)}, \ldots, m_k^{(t+1)})$.

Step (4): If the iteration converges or meets the stop condition, the sample clustering result $C^* = C^{(t)}$ will be the output. Otherwise, let $t = t + 1$ and return to Step (2).

4.3. Road Traffic Status Prediction Based on the Classification Decision Tree Model

(1) Feature selection

Generally, the criterion of feature selection is the information gain or information gain ratio. However, when information gain is used as a feature to divide the training dataset, it tends to be a problem of selecting features with more values (Saracoglu and Ozen, 2020). Therefore, in this study, we used the information gain ratio to address this problem. The algorithm used to calculate the information gain ratio is as follows.

Step (1): Calculate the empirical entropy $H(D)$ of the data set D. The calculation formula is shown in Eq. (9)

$$H(D) = -\sum_{k=1}^{K} \frac{|C_K|}{|D|} \log_2 \frac{|C_K|}{|D|} \tag{9}$$

Step (2): Calculate the empirical conditional entropy $H(D|A)$ of feature A to data set D. The calculation formula is shown in Eq. (10)

$$H(D|A) = \sum_{i=1}^{n} \frac{|D_i|}{|D|} H(D_i) = -\sum_{i=1}^{n} \frac{|D_i|}{|D|} \sum_{k=1}^{K} \frac{|D_{ik}|}{|D_i|} \log_2 \frac{D_{ik}}{D_i} \tag{10}$$

Step (3): Calculate information gain ratio $gR(D, A)$. The calculation formula is shown in Eq. (11)

$$gR(D, A) = \frac{H(D) - H(D|A)}{H_A(D)} \tag{11}$$

Where $H_A(D) = -\sum_{i=1}^{n} \frac{|D_i|}{|D|} \log_2 \frac{|D_i|}{|D|}$, and n represents the number of values of feature A.

(2) Generation of the classification decision tree

The C4.5 algorithm was used to select characteristic parameters on each node of the classification decision tree by using the information gain ratio criterion. The classification decision tree was constructed recursively. The construction process is shown in Table 1.

The classification decision tree generated by the C4.5 algorithm often has high classification accuracy for model training data, but it is not accurate for unknown test data. This is generally called as the over-fitting phenomenon (Malakis et al., 2020). Therefore, it is necessary to reduce the complexity of the road traffic status prediction model via decision tree pruning. The pruning process for the classification decision tree is presented in Table 2.

5. Case Analysis

5.1. Data Source

Based on velocity data collected by the road section detectors in Nan'an District, Chongqing from June 1 to June 30, 2021, and the Chongqing meteorological data crawled from the website www.wunderground.com, this study demonstrated and analyzed the established road traffic state prediction model. As the time interval for updating the weather data was one h, the traffic flow velocity data were divided at an interval of 1 hour to calculate the average velocity. Thus, 720 groups of characteristic parameter data, such as pavement status, time characteristics, working day characteristics, and traffic flow velocity can be obtained.

Tabel 1. Classification decision tree generated using the C4.5 algorithm

Input: training data set D and feature A of the road traffic status prediction model, where the values of data set D and feature A are the same as those in Table 3. Output: the road traffic status prediction model.
(1) If all instances in D belong to the same class C_k, set it as a single node tree, consider C_k as the class of the node, and return T'.
(2) If $A=\emptyset$, set T' as a single node tree, consider the class C_k with the largest number of instances in D as the class of the node, and return T'.
(3) Otherwise, calculate the information gain ratio of each feature in A to D according to Equation (10) and select the feature A_g with the largest information gain ratio.
(4) If the information gain ratio of A_g is less than the threshold ε, set s as a single node tree, consider the class C_k with the largest number of instances in D as the class of the node, and return T'.
(5) Otherwise, for each possible value a_i of A_g, D is divided into several subsets of non-empty D_i according to $A_g = a_i$, and the class with the largest number of instances in D_i is used as a marker to construct sub-nodes. The tree T' is composed of the node and subnodes, and T' was returned.
(6) For node i, take D_i as the training set and $A - \{A_g\}$ as the feature set, and recursively call (1) – (5) to obtain the subtree T_i, which is the road traffic status prediction model. In the equation, the attribute values of the class C_k are no congestion, mild congestion, moderate congestion, and serious congestion.

Tabel 2. Pruning algorithm of classification decision tree

Input: the road traffic status prediction model T' before pruning.

Output: the road traffic status prediction model T_β after pruning.

(1) Calculate the empirical entropy of each node.

(2) Recursively retract upward from the leaf nodes of the tree.

(3) Let the entire trees be T_A and T_B before and after a group of leaf nodes retract to their parent nodes, respectively. The corresponding loss function values are $C_\alpha(T_A)$ and $C_\alpha(T_B)$, respectively. If $C_\alpha(T_A) < C_\alpha(T_B)$, then pruning is performed to change the parent node into a new leaf node.

(4) Return to Steps (2) and (3) until it cannot continue and obtain the subtree with the smallest loss function, which is the pruned road traffic status prediction model T_β.

5.2. Evaluation Index

To test the reliability of the model, this study used the mean absolute error (*MAE*) and root mean square error (*RMSE*) to evaluate the accuracy of the regression decision tree model for velocity prediction. Moreover, the accuracy rate (*Acc*) was used to evaluate the accuracy of the classification decision tree model for road traffic status prediction. The calculation formulas are shown in Eq. (12)-(14).

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|p_i - q_i| \tag{12}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(p_i - q_i)^2} \tag{13}$$

$$Acc = \frac{1}{N}\sum_{i=1}^{N} o_i * 100\%, \qquad o_i = \begin{cases} 1, u_i = z_i \\ 0, u_i \neq z_i \end{cases} \tag{14}$$

Where N represents the number of samples, p_i represents the true value of velocity, q_i represents the predicted value of velocity, u_i represents the true value of road traffic status levels, and z_i represents the predicted value of road traffic status levels.

5.3. Result Analysis

5.3.1. Analysis of traffic flow velocity prediction results

According to the traffic travel characteristics of urban roads in Chongqing, we considered 7:00 – 9:00 and 17:00 – 19:00 as the peak hours, 6:00 – 7:00, 9:00 – 17:00, and 19:00 – 23:00 as the flat peak hours, and 23:00 – 6:00 as the night hours. Among all the sample data collected, 520 datasets were selected as the training set and the remaining 200 datasets were selected as the test set to train and test the traffic flow velocity prediction model. The test results of the model are shown in Fig. 2. As shown in Fig. 2, the traffic flow velocity prediction model had a total of 14 velocities with different values. Furthermore, we can observe that the model has a good prediction effect on the traffic flow velocity. By substituting the predicted and real velocity values obtained from the sample test set into Eq. (12) and (13) for calculation, it can be concluded that the *MAE* and *RMSE* values of the model are 1.69 and 2.38, respectively.

5.3.2. Analysis of the road traffic status prediction results

Before generating the road traffic status prediction model, some parameters in the model need to be calibrated. Set the value of the hyperparameter k in the Kmeans clustering algorithm as 4. In the classification decision tree model, the values of the pavement status attributes are 1, 2, and 3 to represent dry pavement, wet pavement, and flooded pavement, respectively. The values of the traffic flow velocity attributes are listed in Table 3. In the classification results, the values of road traffic status attributes 1, 2, 3, and 4 represent no congestion, mild congestion, moderate congestion, and serious congestion, respectively.

Table 3. Traffic flow velocity attribute values

Velocity	$v\in$ (0,15)	$v\in$ [15,20)	$v\in$ [20,30)	$v \geq 30$
Value	1	2	3	4

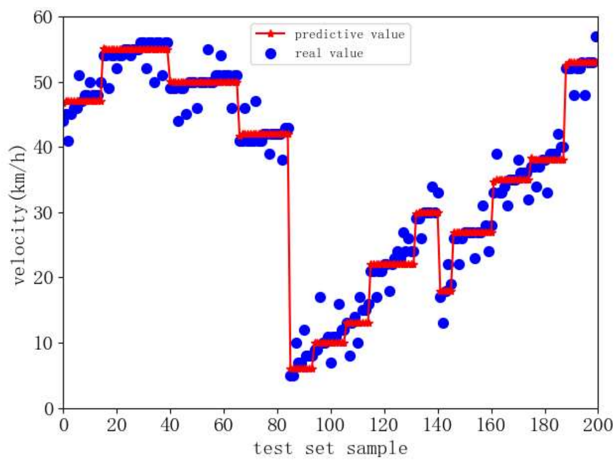

Fig. 2. Prediction results of traffic flow velocity

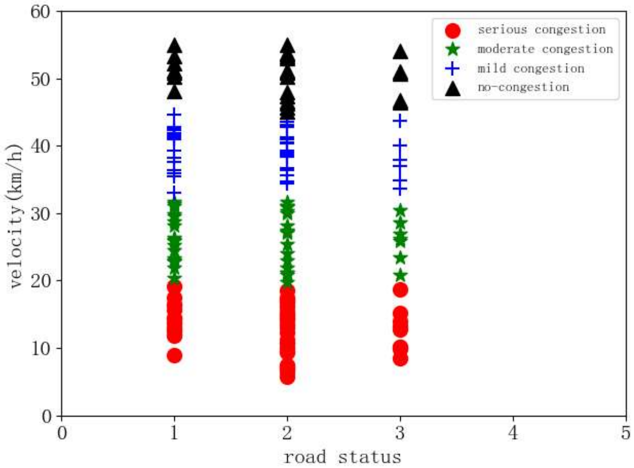

Fig. 3. Results of Kmeans clustering algorithm

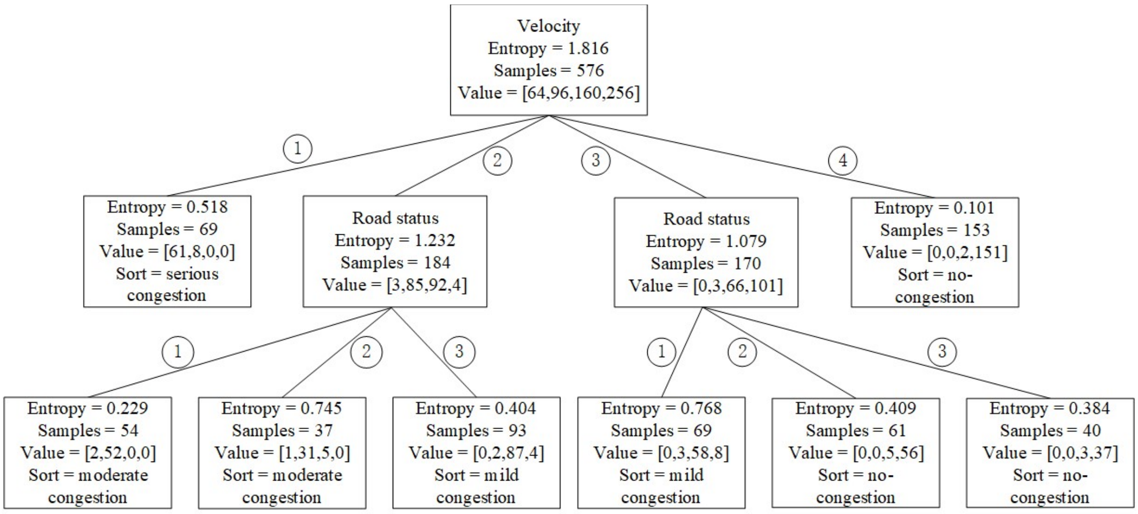

Fig. 5. Classification decision tree model for road traffic status prediction

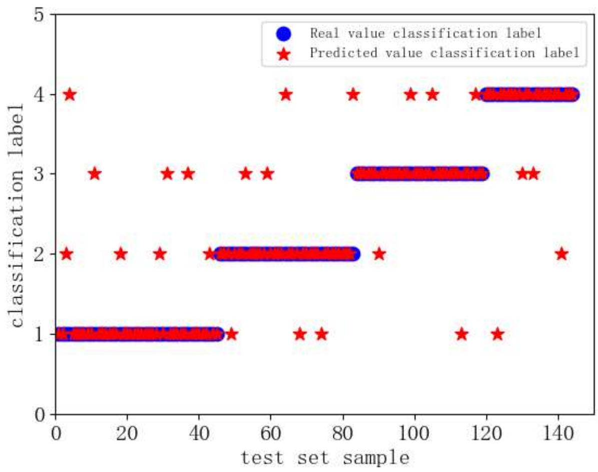

Fig. 4. Results of road traffic status prediction

After calibrating the values of parameter attributes, all the sample data were substituted into the Kmeans clustering algorithm for clustering analysis. The clustering results obtained are shown in Fig. 3. After obtaining the traffic status labels of each sample by using the Kmeans clustering algorithm, 80% of the sample data were selected as the training set of classification decision tree model, and the remaining 20% of the sample data was used as the test set to train and test the model, respectively. The results are shown in Fig. 4 and 5, respectively.

After substituting the real value of the sample classification label in the test set and the predicted value of the sample label obtained by the classification decision tree model into Eq. (14), it can be concluded that the classification accuracy (*Acc*) of the model is 81.31%, indicating that the classification result is good. Therefore, it can be considered that the model has good applicability to the problem of road traffic status level prediction.

6. Conclusion

In this study, based on the respective advantages of unsupervised and supervised learning algorithms in machine learning, a road traffic status prediction model based on the Kmeans decision tree is established. Through the empirical analysis, the validity of the model is verified, and the classification prediction of the urban road traffic status is realized, and the prediction accuracy is also relatively high. The traffic management department can predict the road traffic status in advance and take effective control measures in time, so as to avoid the occurrence of traffic congestion to the greatest extent and reduce the losses caused by traffic congestion. Since this method can only be applied to traffic status prediction under a specific road section, we can try to use different models to predict the traffic status of various regions and roads to find a suitable prediction method for different regions and roads in the subsequent research, thereby making it more universal.

7. Acknowledgments

This research was supported by the Chongqing Postgraduate Joint Training Base Project (Chongqing Jiaotong University-Chongqing YouLiang Science & Technology Co., Ltd Joint Training Base for Postgraduates in Transportation).

References

Aifadopoulou, G., Bratsas, C., Koupidis, K., Chatzopoulou, A., Salanova, J. M., and Tzenos, P. (2018). Short-Term Prediction of the Traffic Status in Urban Places Using Neural Network Models. *Proceedings of The 4th Conference on Sustainable Urban Mobility*, Springer, Cham, 181-188.

Alghamdi, T., Elgazzar, K., Bayoumi, M., Sharaf, T., and Shah, S. (2019). Forecasting traffic congestion using ARIMA modeling. *Proceedings of 2019 15th International Wireless Communications & Mobile Computing Conference* (*IWCMC*), IEEE, 1227-1232.

Guo, F., Polak, J. W., and Krishnan, R. (2018). Predictor fusion for short-term traffic forecasting. *Transportation research part C: emerging technologies*, 92, 90-100. doi: 10. 1016/j. trc. 2018. 04. 025

Hao, W., Rong, D., Yi, K., Zeng, Q., Gao, Z., Wu, W., and Scepanovic, B. (2020). Traffic Status Prediction of Arterial Roads Based on the Deep Recurrent Q-Learning. *Journal of Advanced Transportation*, 2020(1). doi: 10. 1155/2020/8831521

Huang, F. R., Wang, C. X., and Chao, C. M. (2020). Traffic Congestion Level Prediction Based on Recurrent Neural Networks. *Proceedings of 2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, IEEE, 248-252.

Ji, B. and Hong, E. J. (2019). Deep-learning-based real-time road traffic prediction using long-term evolution access data. *Sensors*, 19(23), 53-65. doi: 10. 3390/s19235327

Kumar, S. V. and Sivanandan, R. (2019). Traffic congestion quantification for urban heterogeneous traffic using public transit buses as probes. *Periodica Polytechnica Transportation Engineering*, 47(4), 257-267. doi: 10. 3311/PPtr. 9218

Li, H. (2019). *Statistical Learning Methods*. Beijing: Tsinghua University Press, 67-89.

Liu, Z., Du, W., Yan, D. M., Chai, G., and Guo, J. H. (2018). Short-term traffic flow forecasting based on combination of k-nearest neighbor and support vector regression. *Journal of Highway and Transportation Research and Development* (*English Edition*), 12(1), 89-96. doi: 10. 1061/JHTRCQ. 0000615

Malakis, S., Psaros, P., Kontogiannis, T., and Malaki, C. (2020). Classification of air traffic control scenarios using decision trees: insights from a field study in terminal approach radar environment. *Cognition, Technology & Work*, 22(1), 159-179. doi: 10. 1007/s10111-019-00562-7

Miglani, A. and Kumar, N. (2019). Deep learning models for traffic flow prediction in autonomous vehicles: A review, solutions, and challenges. *Vehicular Communications*, 20(2), 100-114. doi: 10. 1016/j. vehcom. 2019. 100184

Minh, Q. T., Pham-Nguyen, H. N., Tan, H. M., and Long, N. X. (2019). Traffic congestion estimation based on crowd-sourced data. *Proceedings of 2019 International Conference on Advanced Computing and Applications (ACOMP)*. IEEE, 119-126.

Nguyen, N. T., Dao, M. S., and Zettsu, K. (2020). Leveraging 3D-Raster-Images and Deep CNN with Multi-source Urban Sensing Data for Traffic Congestion Prediction. *Proceedings of International Conference on Database and Expert Systems Applications*, Springer, Cham, 396-406.

Quang, D. T., and Bae, S. H. (2021). A Hybrid Deep Convolutional Neural Network Approach for Predicting the Traffic Congestion Index. *Promet-Traffic&Transportation*, 33(3), 373-385. doi:10. 7307/ptt.v33i3.3657

Saracoglu, A. and Ozen, H. (2020). Estimation of traffic incident duration: a comparative study of decision tree models. *Arabian Journal for Science and Engineering*, 45, 8099-8110. doi: 10. 1007/s13369-020-04615-2

Sunindyo, W. D. and Satria, A. S. M. (2020). Traffic Congestion Prediction Using Multi-Layer Perceptrons And Long Short-Term Memory. *Proceedings of 2020 10th Electrical Power, Electronics, Communications, Controls and Informatics Seminar (EECCIS)*, IEEE, 209-212 .

Zaki, J. F., Ali-Eldin, A., Hussein, S. E., Saraya, S. F., and Areed, F. F. (2020). Traffic congestion prediction based on Hidden Markov Models and contrast measure. *Ain Shams Engineering Journal*, 11(3), 535-551. doi: 10. 1016/j. asej. 2019. 10. 006

Zhang, S., Yao, Y., Hu, J., Zhao, Y., Li, S., and Hu, J. (2019). Deep autoencoder neural networks for short-term traffic congestion prediction of transportation networks. *Sensors*, 19(10), 92-105. doi:10. 3390/s191 02229

Zhao, X. H.and Ren, G. C. (2017). Summary of Research on Driving Behavior in Bad Weather. *Traffic Information and Safety*, 35(5), 70-75. doi: 10.3963/j. issn. 1674-4861. 2017. 05. 009

Xinghua Hu was born in China in 1981. He received the Ph. D. degree from Beijing Jiaotong University in 2016. His current research interests include urban and regional transportation planning, green transportation, intelligent transportation. He is currently a professor of Chongqing Jiaotong University, deputy director of Chongqing Key Laboratory of transportation engineering, green transportation expert of the Ministry of transportation and Chongqing comprehensive bid evaluation expert. From 2016 to 2021, he presided over and participated in more than ten national, provincial and ministerial scientific research projects, presided over and participated in the preparation of 11 national transportation industry energy conservation and emission reduction project guidelines, local industry standards and so on. He published more than 30 papers and six monographs and won five provincial and ministerial awards such as Chongqing Development Research Award, Chongqing Social Science Outstanding Achievement Award, and China Highway Society Science and Technology Award.

Xinghui Chen was born in Chongqing, China, in 1998. He received a B. E. degree from the School of Traffic and Transportation, Chongqing Jiaotong University in 2018, where he is currently pursuing the master's degree. His current research interests include green energy transportation and transportation planning and management.

Wei Liu was born in China in 1978. He received a Ph.D. from Southwest Jiaotong University in 2011. His current research interests include traffic system planning and intelligent management, road traffic safety and improvement design, and other fields. He is currently a professor of Chongqing Jiaotong University, director of the Institute of transportation engineering, School of transportation, Chongqing Jiaotong University, chairman of the Technical Committee for traffic organization and operation management of the Transportation Engineering Department of the world transportation Congress, director of China Society of transportation engineering, expert of China Road Traffic safety think tank, vice president and Secretary General of Chongqing safety production accident research association Deputy Secretary General of urban transportation special committee of Chongqing Planning Society, expert of Chongqing automatic driving on the road expert committee, expert of Chongqing vehicle management expert advisory committee, and expert of Chongqing Comprehensive Transportation Planning Advisory Committee. From 2011 to 2021, his scientific research projects and achievements have mainly completed more than 110 scientific research projects, won the third prize of provincial and ministerial scientific research progress twice, and the third prize of social development research award. He has guided students to win the National University Student Transportation

Science and Technology Competition twice. He completed three innovation and entrepreneurship training projects, instructed students to apply for one invention patent and two utility model invention patents, published more than 30 papers.

Gao Dai was born in China in 1966. He received the master degree from Huazhong University of Science and Technology, 1989. He is currently the general manager of Chongqing Ulit Science & Technology Co., Ltd. His current research interests include traffic behavior and big data modeling application, intelligent traffic control, and construction of road traffic facilities.