ROAD NETWORK PARTITIONING METHOD BASED ON CANOPY-KMEANS CLUSTERING ALGORITHM

Xiaohui LIN[1], Jianmin XU[2]

[1]*Institute of Rail Traffic, Guangdong Communication Polytechnic, Guangzhou, China*

[2]*School of Civil Engineering and Transportation, South China University of Technology, Guangzhou, China*

Abstract:

With the increasing scope of traffic signal control, in order to improve the stability and flexibility of the traffic control system, it is necessary to rationally divide the road network according to the structure of the road network and the characteristics of traffic flow. However, road network partition can be regarded as a clustering process of the division of road segments with similar attributes, and thus, the clustering algorithm can be used to divide the sub-areas of road network, but when Kmeans clustering algorithm is used in road network partitioning, it is easy to fall into the local optimal solution. Therefore, we proposed a road network partitioning method based on the Canopy-Kmeans clustering algorithm based on the real-time data collected from the central longitude and latitude of a road segment, average speed of a road segment, and average density of a road segment. Moreover, a vehicle network simulation platform based on Vissim simulation software is constructed by taking the real-time collected data of central longitude and latitude, average speed and average density of road segments as sample data. Kmeans and Canopy-Kmeans algorithms are used to partition the platform road network. Finally, the quantitative evaluation method of road network partition based on macroscopic fundamental diagram is used to evaluate the results of road network partition, so as to determine the optimal road network partition algorithm. Results show that these two algorithms have divided the road network into four sub-areas, but the sections contained in each sub-area are slightly different. Determining the optimal algorithm on the surface is impossible. However, Canopy-Kmeans clustering algorithm is superior to Kmeans clustering algorithm based on the quantitative evaluation index (e.g. the sum of squares for error and the R-Square) of the results of the subareas. Canopy-Kmeans clustering algorithm can effectively partition the road network, thereby laying a foundation for the subsequent road network boundary control.

Keywords: *traffic engineering; road network partition; canopy clustering algorithm; kmeans clustering algorithm; macroscopic fundamental diagram*

To cite this article:

Contact:
1) linxh1981@163.com [https://orcid.org/0000-0002-4126-5430], 2) [https://orcid.org/0000-0002-8922-8815]

1. Introduction

The number of vehicles is quickly increasing, urban traffic network is becoming increasingly complex and the scope of traffic signal control widens with the rapid development of social economy. the traffic density of each section of urban road network, which is not conducive to urban traffic management and control, varies greatly with the uneven distribution of road network congestion and various types of roads(Zhao et al., 2018; Klos and Sobota, 2019). Therefore, the urban road network needs to be divided. Walinchus(1971), an American scholar, first proposed the concept of road network partition. He pointed out that a complex and huge road network can be divided into several independent sub areas according to certain principle indicators, and then appropriate control optimization strategies can be implemented according to the holding of sub areas, so as to lower the control power of the road network level by level and make the whole road network system more flexible, efficient and reliable. It can be seen that reasonable road network partition can improve the stability and flexibility of the traffic control system.

Initially, the static division method was used to partition the road network.This method means that researchers divide the road network into sub regions according to the historical data of the road network (such as traffic flow, traffic density, road network structure and road network size). The static division method is easy to realize and is feasible for the road network with little traffic flow change. However, once the random change of traffic flow is large, a lot of manpower and material resources need to be invested to get the traffic data again. Therefore, some scholars have studied the method of dynamic subarea division. For example, Yun et al.(2004) increased the constraints of travel time, improved the road network zoning method in the accident management system, solved the improved model with genetic algorithm, and finally evaluated the traffic benefit. Zhou et al.(2013) proposed a new calculation method of correlation degree to determine whether two adjacent intersections need to be divided in the same sub area. Xu et al.(2017) proposed a dynamic sub area division method of road network based on different congestion levels according to the homogeneity and relevance of intersections in different states of road network.

Recent research shows that urban road network traffic flow has objective regularity, which is called macroscopic fundamental diagrams (MFD)(Daganzo, 2007; Geroliminis and Sun, 2011). Godfrey(1969) first proposed the concept of MFD, but it was not until 2007 that Daganzo And Geroliminis revealed the theoretical principle of MFD (Daganzo, 2007; Geroliminis and Daganzo, 2008). These two scholars proposed that MFD is the internal objective law of the road network, which objectively reflects the general relationship between the weighted traffic flow (q^w) and the weighted traffic density (k^w) in the road network. Some scholars use MFD theory to study the method of road network subarea division. For example, Ji et al. (2012) firstly used Ncut method to initially divide the road network and then used merge algorithm to roughly divide the initial division. Finally, the road network was divided into several homogeneous road network partitions by using boundary adjustment algorithm to reduce the density variance of the section. Haddad et al.(2014) designed robust peripheral controller based on MFD at the boundary of two sub regions, solved the flow regulation problem of boundary control of two sub regions, and then proposed a new sub region division model. Saeedmanesh et al.(2015) searched for the road section with the lowest density variance in each road section of the road network and its adjacent road sections, combined to form a set of road sections with similar density, which was defined as snake set. Then, the snake sets with similar similarity were merged and fine-tuned, and finally the heterogeneous road network was divided into several homo-proton areas. Later, they optimized three models and verified them in the actual large-scale road network. The results show that the algorithm is better than Ji's method(Saeedmanesh and Geroliminis, 2016). An et al.(2017) proposed a robust and efficient road network subarea division algorithm by using the link connectivity and regional growth technology, which was implemented and tested on the regional planning road network in Arizona, USA.

In fact, the road network partition can be regarded as the clustering process of the segments with similar attributes, and thus, clustering algorithm can be used to partition the sub-areas of the road network. Clustering divides the data set into several clusters (classes) with similar attributes and features. The attributes or features of objects in the same cluster are similar to one another but are quite different from

those in other clusters. Clustering algorithm has been applied to traffic condition identification; road network, traffic control period, accident point and flow partition; and other traffic fields.

However, the application of clustering algorithm in road network partition is still in its infancy. Scholars used clustering algorithm to investigate road network partition. For example, Li et al.(2009) proposed an automatic sub-area division method of road network based on spatial statistical clustering algorithm and used the floating car data of Shanghai actual road network to realise the automatic sub-area division of road network. Dai et al. (2010) established a weighted fuzzy clustering analysis method based on fuzzy C-means clustering algorithm (FCM) and used the analytic hierarchy process to optimise the weights. Finally, the feasibility of the method was verified by taking the West Bank Economic Zone of the Straits as the experimental area. Yin et al. (2010) proposed dynamic road network partition based on spectral graph theory and spectral clustering algorithm according to real-time traffic flow data and the topological structure attributes of road intersections. Du et al. (2014) proposed a traffic area partition method framework based on expressway traffic connection by using and converting weighted average distance clustering analysis method, which is based on OD traffic volume data between toll stations of expressway network toll collection, to calculate the traffic connection volume. Feng et al. (2015) proposed a sub-area merging model based on two-dimensional graph theory clustering algorithm for road network merging and used F test to determine the optimal merging results. Pascale et al.(2015) proposed a homogeneous sub-area detection method based on spatial clustering method, which was validated by the actual London road network. Lopez et al.(2017) proposed density-based spatial and data point clustering post-processing methods by using normalised cutting, density-based noise spatial clustering and growth nerve gas. The method was validated by a large urban network in Amsterdam, the Netherlands. Wang (2017) proposed road network partition methods based on Kmeans clustering and improved FCM algorithm and compared their advantages and disadvantages through actual road network analysis. However, the K value of Kmeans algorithm is pre-set, which is difficult to estimate and does not have universality. At

the same time, random selection of clustering centres may lead to different clustering results each time, and the obtained results may not be the optimal solution.

Therefore, we proposed a road network partitioning method based on Canopy-Kmeans clustering algorithm to make up for the shortcomings of the Kmeans algorithm. We considered the longitude and latitude of the central section, average speed and density of the section collected in real time as sample data and constructed a vehicle-connected network simulation model. Then, Kmeans and Canopy-Kmeans clustering algorithms are used to partition the road network. Finally, the quantitative evaluation method of road network partition based on MFD is used to evaluate the results of road network partition and the optimal road network partition algorithm is determined.

2. Brief introduction of kmeans and canopy algorithms

2.1. Introduction of kmeans algorithm

Kmeans algorithm is a classical unsupervised learning clustering method based on distance. The algorithm divides the samples that have eigenvalues close to one another to form multiple clusters. The distances between two objects is considered close, and the similarity degree is higher.

The algorithm randomly selects K data objects from N data objects as initial clustering centres, calculates the distance between the remaining data objects and K clustering centres, divides the remaining data objects into corresponding clusters with the smallest distance and recalculates the pairwise mean of all data objects in the clustering of each new sample and considers them new clustering centres. This process is repeated until the sum of mean square deviation is minimised.

The advantages of Kmeans algorithm are presented as follows: 1) simple idea and rapid running speed; 2) high computing efficiency and scalability for large data sets and 3) low time complexity and suitable for data mining of large data sets.\

However, the algorithm presents the following shortcomings: 1) pre-selected K value is difficult to estimate and 2) initial clustering centre is randomly selected. Different results may appear in each calculation. If the improper initial value is selected the clustering result may not be the optimal clustering result.

2.2. Introduction of canopy algorithm

Canopy algorithm is a kind of rough clustering algorithm that does not need to specify K value beforehand. It has rapid execution speed, indicating that it has great practical application value. The main idea is to randomly select one data object as the initial clustering centre for arbitrary data set V, set two concentric region radii (e.g. T1 and T2), calculate the similarity of all data objects in the data set by rough distance calculation method and divide the data set into several overlapping small datasets according to the similarity of each data object (defined as Canopy). After many iterations, all data objects can eventually fall within the scope of Canopy coverage(Liu and Su, 2016) (Fig 1).

Aiming at the problems of radii T1 and T2 of the Canopy region and random initial clustering centres, Mao et al.(2012) proposed a Canopy optimization method based on the 'minimum maximum principle' to improve the clustering accuracy of the Canopy algorithm.

Canopy algorithm can effectively overcome the shortcomings of Kmeans algorithm, but with low accuracy. It can combine the two aforementioned algorithms to improve clustering accuracy. Firstly, the Canopy algorithm is used to divide data objects into several canopies, and Kmeans algorithm is used to calculate the distances of all data vectors in the same canopy.
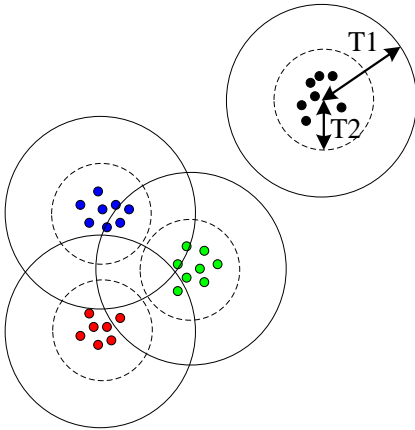

Fig. 1. Schematic of canopy algorithm

3. Road network partition method based on Canopy-Kmeans clustering

The subarea of road network is divided by Canopy-Kmeans clustering based on real-time sample data, such as longitude and latitude and average speed and density. The algorithm flow is presented as follows:

Phase 1 (Data Pre-processing): Several Canopy and Canopy centres are determined by Canopy algorithm based on 'minimum and maximum principle'(Mao, 2012).

(1) Determine Canopy: The road network dataset is expressed as:

$$X = \{x_i | i = 1,2,\cdots,n\} \quad (1)$$

where x_i is the partition parameter of section i in the road network and:

$$x_i = \{x_{i1}, x_{i2}, x_{i3}, x_{i4}\} \quad (2)$$

where x_{i1} is the longitude of section centre, x_{i2} is the latitude of section centre, x_{i3} is the average speed of section and x_{i4} is the average density of section. If $\forall x_i \in X$ satisfies the following formula:

$$\{C_j | \exists ||x_i - C_j|| \leq Dist_{\min}(i), C_j \subseteq X, \quad i \neq j\} \quad (3)$$

Then x_i is called a Canopy set, C_j is the Canopy central point and $\forall C_j \in X$. Assuming that the first m Canopy centres are known, $Dist_{\min}(i)$ denotes the largest minimum distance between the second candidate data point x_i and the first m Canopy centres. The formula is presented as follows:

$$\begin{cases} DistC(i) = \min\{d(x_i, x_r), \\ \qquad r = 1,2,\cdots,m\} \\ Dist_{\min}(i) = \max\{\min[d(x_i, x_{j_r})], \\ \qquad i \neq j_1, j_2, \cdots, j_r\} \end{cases} \quad (4)$$

where $\min[d(x_i, x_{j_r})]$ is the smallest distance between i-th Canopy central point to be determined and the first m determined Canopy centres.

(2) Determine the Canopy central point: The road network data set is expressed as:

$$X = \{x_i | i = 1,2,\cdots,n\} \tag{5}$$

if it satisfies the following formula:

$$\{C_m | \exists ||x_i - C_m|| \leq Dist_{\min}(i), C_j \subseteq X, \\ i \neq m\} \tag{6}$$

where C_m is a set of non-Canopy candidate centres.

Phase 2 (Kmeans clustering): The Canopy centres are used as the clustering centres of Kmeans algorithm. The number of central points of Canopy corresponds to the K value of Kmeans algorithm, and the Kmeans algorithm is used for clustering analysis.

(1) Take each Canopy central point as the K-cluster central point, and the number of Canopy central points is K.

(2) Calculate the Euclidean distances from each data point to K clustering centres and divide them into clusters with the smallest distance to form a new cluster.

(3) Calculate the mean value of each data point in K new clustering and regard this value as a new clustering centre. The formula is expressed as follows:

$$u_k = \frac{\sum_n r_{nk} x_n}{\sum_n r_{nk}} \tag{7}$$

where r_{nk} denotes whether the coefficients belong to class K. If x_n belongs to the range of class K, then $r_{nk}=1$; otherwise, $r_{nk}=0$. The formula is expressed as follows:

$$r_{nk} = \begin{cases} 1 & k = \text{argmin}||x_n - x_j|| \\ 0 & \text{others} \end{cases} \tag{8}$$

(4) Repeat Steps 2 to 3 until the cluster centre remained unchanged.

(5) K clusters are obtained by outputting the results. Fig 2 shows the flow chart of the aforementioned Canopy-Kmeans clustering algorithm.

4. Quantitative evaluation method of road network partition based on MFD

Nagle et al.(2014) proposed that if the floating cars are evenly distributed in the road network and the proportion of floating cars is known, then the MFD of the road network can be estimated by using the traffic trajectory estimation method(Edie,1963), which is called floating car data(FCD) estimation method. The formula is presented as follows :

$$k^w = \frac{\sum_{j=1}^{m'} t'_j}{\rho^* T * \sum_{i=1}^{n} l_i} \tag{9}$$

$$q^w = \frac{\sum_{j=1}^{m'} d'_j}{\rho^* T * \sum_{i=1}^{n} l_i} \tag{10}$$

where k^w and q^w are the weighted traffic density (*veh/km*) and the weighted traffic flow (*veh/h*) respectively; ρ is the ratio of floating cars; T is the acquisition cycle(s); n is the total number of road network sections; l_i is the road section i length (*km*); m' is the number of floating cars during the acquisition cycle T (veh); t'_j is the driving time of j-th floating car during the acquisition cycle T(s) and d'_j is the driving distance of j-th floating car during the acquisition cycle T(m).

According to Daganzo and Geroliminis, the MFD of road network presents a cubic function curve of one variable, which can be expressed as follows:

$$q^w(k^w(t)) = ak^w(t)^3 + bk^w(t)^2 \\ + ck^w(t) + d \tag{11}$$

and MFD is in homogeneous road network. Thus, the rationality of road network partition can be evaluated according to the fitting degree of MFD scatters. However, to accurately judge the fitting degree of MFD scatters in road network only from the perspective of vision is difficult. Therefore, a quantitative analysis method of road network's MFD based on the sum of squares for error (SSE) and the R-Square is used to evaluate the rationality of the aforementioned road network partition clustering method(Wang, 2017), and its flow chart is shown in Fig 3.

(1) Drawing MFDs of each partition and determining its fitting function

The MFDs of each partition are drawn by using the relevant theory of MFD, and the curve function is fitted on the scatter plot of MFD.

(2) Calculating the SSE and the R-Square

The SSE and the R-Square between scatter points and fitting functions are calculated to

quantitatively evaluate road network partitions. SSE is the deviation between the fitting and the actual data. The smaller the value of SSE, the smaller the deviation between the fitting and the actual data. The formula is as follows:

$$SSE = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \qquad (12)$$

Where, y_i is the actual value; $\hat{y}_i$ is the fitting value; i is i-th data and n is the total data.

The R-Square is determined by the sum of squares for regression (SSR) and the sum of squares for total (SST). SST is the sum of squares of the difference between each actual and average value, which reflects the overall fluctuation of the scatter point. SSR is the sum of squares of the difference between fitting value and mean value. Their formulas are as follows:

$$SST = \sum_{i=1}^{n}(y_i - \bar{Y})^2 \qquad (13)$$

$$SSR = \sum_{i=1}^{n}(\hat{y}_i - \bar{Y})^2 \qquad (14)$$

$$SR - square = \frac{SSR}{SST} = \frac{SST - SSE}{SST}$$
$$= 1 - \frac{SSE}{SST} \qquad (15)$$

where the value of R-square is between 0 and 1. If the value of R-square is closer to 1, then the effect of curve fitting is better.

(3) Analysing the fitting degree of MDF in road network partition SSE and R-square were used to describe the fitting degree of MFD. When SSE is smaller and R-square is closer to 1, the fitting degree of MFD is higher. The MFD scattering points are more centralised, scattering is lower, image is clearer, critical traffic density and maximum traffic flow of road network are easier to determine, traffic state of road network is easier to distinguish from macroscopic level and traffic flow density of whole road network is more uniform. On the contrary, if SSE is larger and R-square is smaller, then the fitting degree of MFD

is worse. If the scatter points of MFD are scattered, then the fitting curve is not obvious, and determining the maximum flow is difficult. Moreover, critical traffic density and state of road network are difficult to judge.
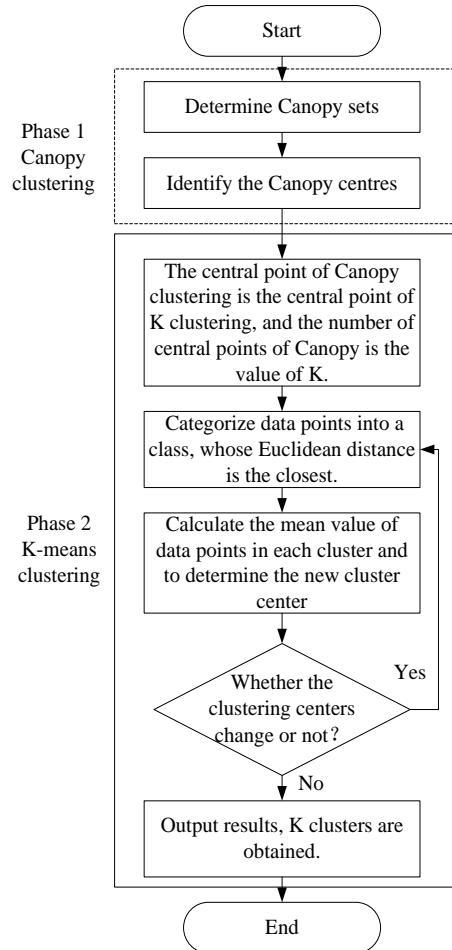

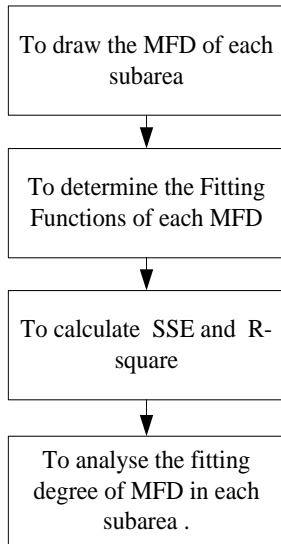Fig. 2 Flow chart of road network partition based on Canopy-Kmeans clustering

Fig. 3 Evaluation method of road network partition based on MFD

5. Empirical Analysis

5.1. Construction of vehicle-connected network simulation platform

To verify the effectiveness of the proposed algorithm, the core road network of Tianhe District in Guangzhou is selected as the research object(Lin et al., 2019), and a vehicle-connected network simulation platform based on Vissim simulation software is constructed (Fig 4). The road network consists of 8 three-dimensional intersections, more than 60 plane intersections and more than 100 entrances and exits.

The simulation results show that when the proportion of connected vehicles (i.e. floating cars) in the traffic volume of the road network (i.e. coverage) reaches 42%, the MFD estimation accuracy of the road network (i.e. the accuracy of weighted traffic density and weighted traffic flow) can reach 97%. Therefore, the coverage rate of networked vehicles is set to 42%. The networked vehicles upload the trajectory data every 15 seconds for a total of 32400 seconds. Then, the networked vehicle data file (*.fzp) of the simulation results is imported into EXCEL file. The FCD estimation method is realised by VBA macroprogramming. The statistical period of the data is 120 seconds. Finally, the weighted traffic

flow q^w and the weighted traffic density k^w of the road network are obtained.

5.2. Result and evaluation of road network partition

5.2.1. Result of road network partition

The spectral clustering algorithm is programmed in MATLAB software by taking simulation time and weighted traffic density of road network as sample data. The simulation time of road network is divided into four stages, namely, low, peak flat peak time, peak time and the over-saturation state. The results of simulation time division of the vehicle network simulation platform are calculated (Table 1).

The results of road network partitioning using Kmeans and Canopy-Kmeans algorithms under over-saturation state are analyzed by taking road network partitioning under over-saturation state as an example. The simulation result section evaluation data file (*. str) is imported into EXCEL. The coordinates of the section centre point are calculated according to the coordinates of the starting and ending points of the section.

The average speed and density of the section under the supersaturated state are calculated according to the time division results in Table 1. Taking the X-coordinate of the central point, Y-coordinate of the central point, average velocity and density of each section as sample data, Kmeans and Canopy-Kmeans algorithms are programmed separately in MATLAB software to partition the road network (Figs. 5 and 6).
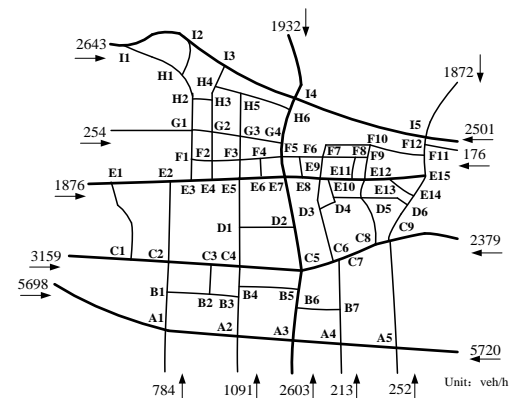

Fig. 4 Layout of the simulation experiment area

The whole road network is divided into four sub-areas (Figs.5 and 6). The sections in subareas 1, 2 and 4 by Kmeans and Canopy-Kmeans algorithms, which are comparable, are slightly different. Sections in sub-area 3 by two algorithms are quite different and not comparable. On the surface, evaluating the results of the two algorithms is impossible. Hence, the quantitative evaluation method of road network partition based on MFD is needed to make further comparison.

Table 1 Results of simulation period division of vehicle networking simulation platform

Simulation period	Weighted traffic density	Traffic state
[0,6960]	[0,12.5]	Low peak time
[6960,16080]	[12.5,20.84]	Flat peak time
[16080,24240]	[20.84,40.88]	Peak time
[24240,32400]	[40.88, -]	Over-saturation state

5.2.2. Quantitative evaluation of road network zoning results

The sections in each sub-area of the road network are screened out according to the results of the aforementioned road network partition (Table 2).

Fig 14 MFD of subarea 4 obtained by Canopy-Kmeans algorithm

Table 3 shows that the SSE and the R-square of MFD in each subarea of the road network are better than those of MFD in the whole network when two clustering algorithms are used for road network partition. The MFDs of subareas 2 and 4 obtained by the two algorithms have the best fitting effect (R-square is more than 0.9). To compare the road network partition results of Kmeans, Canopy-Kmeans more clearly, Table 3 is combined again according to the SSE and the R-square, as shown in Tables 4 and 5.

The SSE and the R-square of MFD in subareas 1, 2 and 4 by Canopy-Kmeans algorithm are superior to those by Kmeans algorithm (Tables 4 and 5). Subarea 3 is not comparable because of the great difference of the road sections. Canopy-Kmeans clustering algorithm is superior to Kmeans clustering algorithm in terms of the quantitative evaluation index of the partition results.

Table 2. Sections included in subarea of road network

Clustering algorithm	Subrea	Section numer
Kmeans	1	1–56, 59–62, 65–75, 239–240
	2	76, 81–84, 86–89, 93–99, 141–174, 191–198, 220–223
	3	114–132, 134–140, 179–190, 206–209, 212–213, 230–231
	4	57–58, 100–113, 133, 175–178, 199–205, 210–211, 214–219, 236
Canopy-Kmeans	1	21–55, 71–81, 90–93, 239–240
	2	83–84, 86–89, 94–99, 127, 130–132, 134, 136, 139–174, 187–198, 220–223
	3	1–20, 56–70, 100–101, 104–105, 116–117
	4	102–103, 106–115, 118–126, 128–129, 133, 135, 138, 175–186, 199–219, 224–238

Table 3 Fitting function of road network partitions' MFD based on Kmeans and Canopy-Kmeans algorithms

Clustering algorithm	Subarea	Fitting function	SSE	R-square
--	whole road network	$y = 0.0107x^3 - 1.55x^2 + 68.585x - 161.41$	9.825E+05	0.8377
Kmeans	subarea 1	$y = 0.0055x^3 - 1.2945x^2 + 88.418x - 272$	8.371E+05	0.8759
	subarea 2	$y = 0.0061x^3 - 0.9491x^2 + 45.494x + 80.979$	4.468E+05	0.9122
	Subarea 3	$y = 0.0037x^3 - 0.6775x^2 + 38.092x + 95.098$	5.088E+05	0.8947
	Subarea 4	$y = 0.0251x^3 - 2.7011x^2 + 86.708x - 161.24$	3.983E+05	0.9163
Canopy-Kmeans	Subarea 1	$y = 0.0151x^3 - 2.4614x^2 + 117.34x - 465.6$	7.202E+05	0.8927
	Subarea 2	$y = 0.0145x^3 - 1.8487x^2 + 73.173x - 110.43$	3.782E+05	0.9344
	Subarea 3	$y = 0.0069x^3 - 1.4005x^2 + 82.683x - 194.6$	9.622E+05	0.8514
	Subarea 4	$y = 0.0152x^3 - 1.892x^2 + 75.561x - 169.65$	2.658E+05	0.9528

Table 4 SSE of MFD fitting data for road network partition based on two clustering algorithms

Algorithm / Subarea	Kmeans	Canopy-Kmeans
Subarea 1	8.37E+05	7.20E+05
Subarea 2	4.47E+05	3.78E+05
Subarea 3	5.09E+05	9.62E+05
Subarea 4	3.98E+05	2.66E+05

Table 5 R-square of MFD fitting data for road network partition based on two clustering algorithms

Algorithm / Subarea	Kmeans	Canopy-Kmeans
Subarea 1	0.8759	0.8927
Subarea 2	0.9122	0.9344
Subarea 3	0.8947	0.8514
Subarea 4	0.9163	0.9528
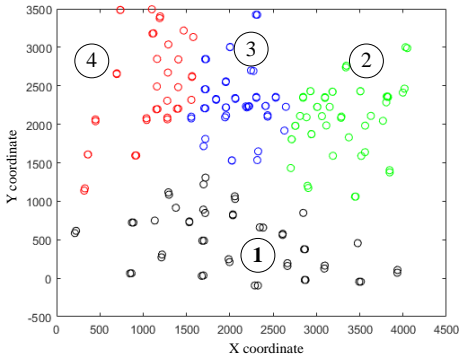
Fig. 5 Result of road network partition based on Kmeans under over-saturation
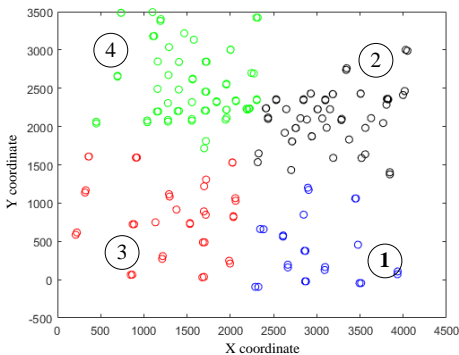
Fig. 6 Result of road network partition based on Canopy-Kmeans algorithm under algorithm over-saturation
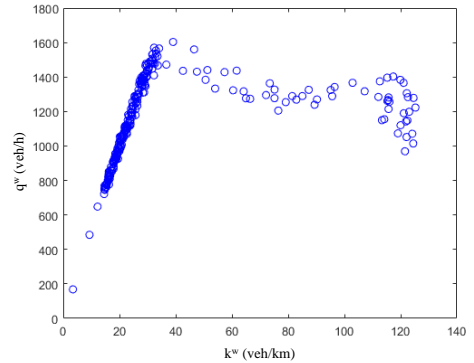
Fig. 7 MFD of subarea 1 obtained by Kmeans algorithm
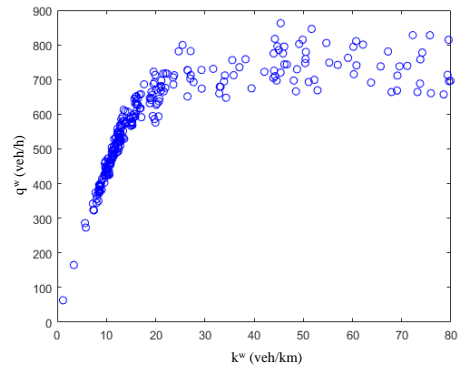
Fig. 8 MFD of subarea 2 obtained by Kmeans algorithm
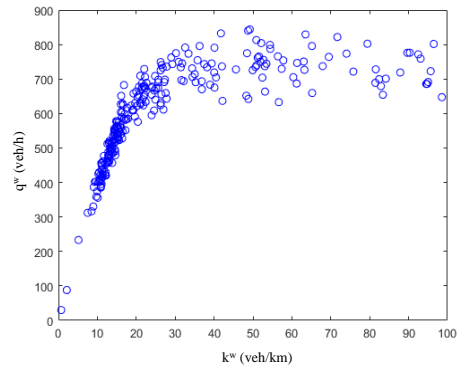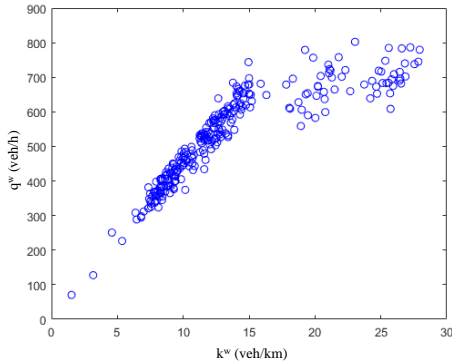

Fig. 9 MFD of subarea 3 obtained by Kmeans algorithm
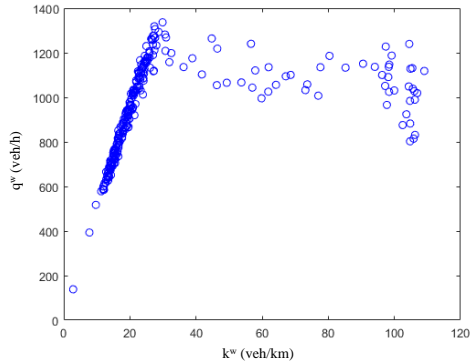
Fig. 10 MFD of subarea 4 obtained by Kmeans algorithm

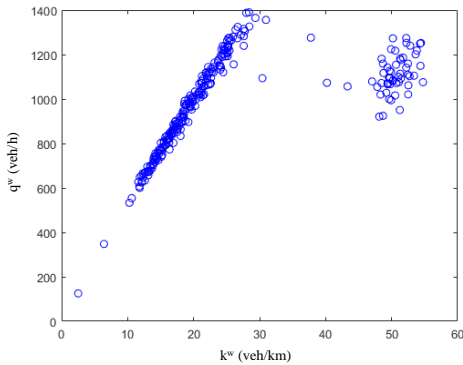Fig. 11 MFD of subarea 1 obtained by Canopy-Kmeans algorithm

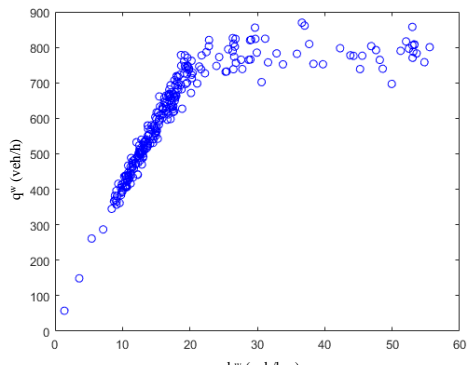Fig. 12 MFD of subarea 2 obtained by Canopy-Kmeans algorithm

Fig. 13 MFD of subarea 3 obtained by Canopy-Kmeans algorithm

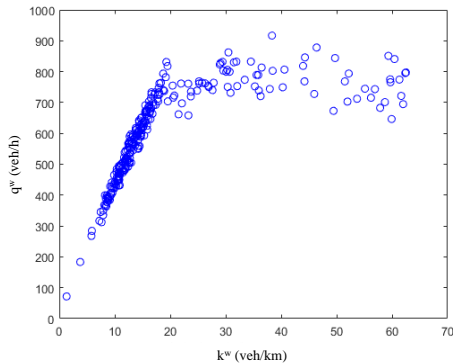Fig. 14 MFD of subarea 4 obtained by Canopy-Kmeans algorithm

6. Conclusion

With the increasing scope of traffic signal control, dividing the road network according to its structure and traffic flow characteristics is necessary to improve the operational efficiency of road network. We proposed a road network partition method based on Canopy-Kmeans clustering to solve the problem in which the Kmeans clustering algorithm is easy to fall into local optimal solution.

The following conclusions are drawn according to the results of empirical analysis:

(1) Kmeans and Canopy-Kmeans clustering algorithms can divide the whole road network into four sub-areas. Slight differences are observed among sub-areas 1, 2 and 4, and great differences are observed in sub-area 3. On the surface,

evaluating the advantages and disadvantages of the two algorithms is impossible.

(2) Canopy-Kmeans clustering algorithm is better than Kmeans clustering algorithm according to the quantitative evaluation index of partition results (SSE and the R-square).

(3) Simulation data obtained from the vehicle connected network simulation platform are taken as sample data, but their impact on traffic accidents, road construction, weather and other special circumstances on the traffic data of the road network was not considered. Therefore, in the next work, the actual traffic data of the road network will be used to verify and analyse the algorithm.

References

[1] AN, K., CHIU, Y. C., HU, X., Chen, X., 2017. *A network partitioning algorithmic approach for macroscopic fundamental diagram-based hierarchical traffic network management.* IEEE Transactions on Intelligent Transportation Systems, 99:1-10.

[2] DAGANZO, C.F., 2007. *Urban gridlock: macroscopic modeling and mitigation approaches.* Transportation Research Part B, 41(1): 49-62.

[3] DAGANZO, C.F., GAYAH, V.V., GONZALES, E.J., 2011. *Macroscopic relations of urban traffic variables: Bifurcations, multivaluedness and instability.* Transportation Research Part B, 45 (1): 278-288.

[4] DAI, B. K., HU, J., 2010. *Research on double-Layer division method of regional traffic district.* Railway Transport and Economy, 32(11), 86-89.

[5] DU, C. J. , CHEN, J. H. , LU, A. Q., 2014. *Transportation zone classification based on cluster analysis with traffic flow contact.* Highway, 59(6), 175-178.

[6] EDIE, L. C., 1963. *Discussion of traffic stream measurements and definitions.* In J.Almond (Ed.), Proceedings of the 2nd International Symposium on International Symposium on Transportation and Traffic Theory, 139-154.

[7] FENG, S. M. , MA, D. C., 2015. *Two-dimensional graphic theory on clustering method of small traffic zones.* Journal of Harbin Institute of Technology, 47(9), 57-62.

[8] GAYAH, V. V., DAGANZO, C.F., 2011. *Clockwise Hysteresis loops in the Macroscopic Fundamental Diagram: An effect of network instability.* Transportation Research Part B, 45 (4): 643-655.

[9] GEROLIMINIS, N. , SUN, J., 2011. *Properties of a well-defined macroscopic fundamental diagram for urban traffic.* Transportation research part b methodological, 45(3), 605-617.

[10] GEROLIMINIS, N., DAGANZO, C.F., 2008. *Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings.* Transportation Research Part B, 42(9): 759-770.

[11] GODFREY, J. W., 1969. *The mechanism of a traffic network.* Traffic Engineering Control, 11(7), 323-327.

[12] GONZALES, E., CHAVIS, C., LI, Y., DAGANZO, C. F., 2009. *Multimodal Transport Modeling for Nairobi, Kenya: Insights and Recommendations with an Evidence based Model.* UC Berkeley, Volvo Working Paper UCB-ITS-VWP-2009-5.

[13] HADDAD, J., SHRAIBER, A., 2014. *Robust perimeter control design for an urban region.* Transportation Research Part B, 68:315-332.

[14] JI, Y., GEROLIMINIS, N., 2012. *On the spatial partitioning of urban transportation networks.* Transportation Research Part B Methodological, 46(10), 1639-1656.

[15] KLOS, M. J., SOBOTA, A., 2019. *Performance evaluation of roundabouts using a microscopic simulation model.* Scientific Journal of Silesian University of Technology. Series Transport,104: 57-67.

[16] LI, X. D., YANG, X. G. , CHEN, H. J., 2009. *Study on traffic zone division based on spatial clustering analysis.* Computer Engineering and

Applications, 45(5), 19-22.

[17] LIN, X. H., XU, J. M., CAO, C. T., 2019. *Simulation and comparison of two fusion methods for macroscopic fundamental diagram estimation.* Archives of Transport,51(3) ,35-48

[18] LIU, B. L., SU, J., 2016. *Improved Canopy-Kmeans algorithm based on double-mapreduce.* Journal of Xi'an Technological University, 36(9), 730-737.

[19] LOPEZ, C., KRISHNAKUMARI, P., LECLERCQ, L., CHIABAUT, N., LINT, J. W. C. H. V., 2017. *Spatiotemporal partitioning of transportation network using travel time data.* Transportation Research Record: Journal of the Transportation Research Board, 2017, 2623:98-107.

[20] MAO, D. H., 2012. *Improved Canopy-Kmeans algorithm based on mapreduce*. Computer Engineering and Applications, 48(27), 22-26+68.

[21] NAGLE A. S., GAYAH V. V., 2014. *Accuracy of networkwide traffic states estimated from mobile probe data*. Transportation Research Record Journal of the Transportation Research Board, 2421(2421):1-11.

[22] PASCALE A., MAVROEIDIS D., LAM H. T., 2015. *Spatiotemporal clustering of urban networks: real case scenario in London.* Transportation Research Record Journal of the Transportation Research Board, 2491(2491), 81-89.

[23] SAEEDMANESH, M., GEROLIMINIS, N., 2015. *Optimization-based clustering of traffic networks using distinct local components*. IEEE, International Conference on Intelligent Transportation Systems. IEEE Computer Society, 2135-2140.

[24] SAEEDMANESH, M., GEROLIMINIS, N., 2016. *Clustering of heterogeneous networks with directional flows based on "Snake" similarities.* Transportation Research Part B Methodological, 91:250-269.

[25] WALINCHUS, R. J., 1971. *Real-time network decomposition and subnetwork interfacing.* Highway Research Record.

[26] WANG, X. X., 2017. *The partition of urban traffic network and classification of traffic status based on clustering.* Beijing Jiaotong University.

[27] XU, J. M., YAN, X. W., JING, B. B., WANG Y. J., 2017. *Dynamic network partitioning method based on intersections with different degree of saturation.* Journal of Transportation Systems Engineering and Information Technology, 17(4), 145-152.

[28] YIN, H. Y. , XU, L. Q. , CAO,Y. R., 2010. *City transportation road network dynamic zoning based on spectral clustering algorithm*. Journal of Transport Information and Safety, 28(1), 16-19+25.

[29] YUN, M. P., YANG, X. G., 2004. *Improvement of road network zoning optimization model for incident management in advanced traffic management systems.* Journal of Highway and Transportation Research and Development, 04:73-76.

[30] ZHAO, H., HE, R., SU, J., 2018. *Multi-objective optimization of traffic signal timing using non-dominated sorting artificial bee colony algorithm for unsaturated intersections*. Archives of Transport, 46(2), 85-96.

[31] ZHOU, Z., LIN S., YUGENG X. I., 2013. A fast network partition method for large-*scale urban traffic networks.* Journal of Control Theory and Applications, 03:37-44.

.