IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received March 2, 2019, accepted March 7, 2019, date of publication March 22, 2019, date of current version April 9, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2905618

A New Combinatorial Characteristic Parameter for Clustering-Based Traffic Network Partitioning

DUANYANG LIU, MENGTING WANG, AND GUOJIANG SHEN
College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China

Corresponding author: Guojiang Shen (gjshen1975@zjut.edu.cn)

This work was supported in part by the Zhejiang Public Welfare Technology Research Program under Grant LGG19F030012, and in part by the Chinese National Natural Science Foundation under Grant 61603339 and Grant 61771430.

ABSTRACT Traffic network partitioning is of great importance in regional coordinated traffic signal control in urban areas. Most partitioning algorithms only use a single traffic parameter to represent dynamic traffic information, which will lead to inaccurate results. Moreover, traditional clustering and heuristic partitioning algorithms are not practical in applications. Thus, in this paper, we first propose a new combinatorial characteristic parameter for clustering-based partitioning algorithm by using the Pearson correlation coefficient and data normalization. Then, we refer to the idea of "snake" algorithm and use a linear programming model to obtain the exact partitioning result, and such algorithm avoids local optimum of heuristic algorithms. Finally, based on the real traffic data of a Chinese city, we conduct the experiments and verify the effectiveness of the new combinatorial parameter.

INDEX TERMS Intelligent transportation systems, partitioning algorithms, clustering methods, correlation, linear programming.

I. INTRODUCTION

For many cities, road congestion in urban traffic has been a large challenge for a long term, especially in the morning and evening rush hours. The regional coordinated traffic signal control method is an effective approach to decrease road congestion level. Such method can improve traffic efficiency of urban road network and reduce parking delay and driving time. Nowadays, using big data technology, human mobility and urban mobility are explored and analyzed in [1] and [2] to alleviate road congestion. But the regional coordinated traffic signal control method is still the most effective and direct way to solve traffic congestion, especially with the help of big traffic data, such method works more effectively.

However, modern urban road network often comprises hundreds of intersections and thousands of road links. As dynamic traffic situation and physical conditions of these intersections and road links are different, the whole traffic network is heterogeneous and some subareas may have different traffic patterns from others. Therefore, before using a regional coordinated traffic signal control scheme, the traffic network needs to be partitioned into homogeneous subareas. Within a traffic subarea, the intersections and road links have similar traffic features and the same signal control strategy can be applied. While among subareas, the traffic situation is different and different control strategies have to be adopted. So, traffic network partitioning has been of great importance in regional coordinated traffic signal control.

In the field of intelligent transportation, clustering is often used to partition traffic network. Its basic idea is to divide data into different sets, data similarity is high within a set and low between sets [3]. Using this idea in partitioning, clusters with similar traffic characteristics can be extracted. When clustering traffic data, spatial adjacency constraints should also be taken into account, and traffic data is aggregated into spatially connected homogeneous subareas. After clustering, we can obtain a certain number of traffic subareas within which traffic situation is similar and inner roads are connected.

According to the using of spatial adjacency constraints, clustering-based traffic network partitioning methods can be divided into two classes. The first class of methods implicitly enforces spatial adjacency conditions and uses an unconstrained clustering algorithm. These methods either hide spatial information into traffic data or consider spatial contiguity when computing data similarity. Based on traffic data with spatial information, such methods build a road network graph with vertices and edges, and partition the graph into a number

The associate editor coordinating the review of this manuscript and approving it for publication was Xiangjie Kong.

of compact regions by using a clustering-based graphic segmentation method. For example, Tan *et al.* [4] utilized graph cut technology to divide the traffic area, Ji and Geroliminis [5] designed a network partition method by using normalized cut, and Anwar *et al.* [6] proposed a new graph cut to partition a well-structured and condensed density peak graph. Different from the first class of partitioning algorithms, the second one explicitly imposes spatial contiguity constraints on the steps of the algorithms. Such algorithms mainly use heuristic methods to obtain traffic subareas. A heuristic algorithm may start by considering each road as one cluster and then merge them iteratively, or start by taking the whole network as one cluster and split it into a number of clusters until a level of homogeneity is reached. For example, the heuristic Newman algorithm was applied in [7] to realize fast traffic network partitioning, the heuristic hierarchical clustering was also adopted in traffic network partition in [8], and a region growing technique that uses heuristics was utilized in [9].

But in practical applications, these two classes of partitioning methods both have their disadvantages. The first one requires traffic network data should be not missing, otherwise, the road network map and its partition will be incomplete. However, data loss or low-quality data is often encountered in the running of intelligent transportation system, and it will lead to incomplete partitioning results by using the first class of partitioning methods. Although the second class of methods can tolerate a certain amount of missing data, the heuristic algorithm can only achieve sub-optimal partitioning results, which have some deviation from actual subareas.

Moreover, most traffic network partitioning methods only use a single characteristic parameter to represent dynamic traffic information. For example, traffic flow is used in [8] and [10]–[12], and vehicle density is applied in [5], [6], [9], and [13]–[16]. Such single parameter cannot accurately represent dynamic traffic situation. Although multiple parameters are adopted in [4] and [7], some of the parameters that they used, such as travel time, queue length and vehicle density, are difficult to collect in actual applications. Therefore, in this paper, considering data accuracy and acquisition easiness, we focus on two related characteristic parameters, i.e., traffic flow and link speed. Using Pearson correlation coefficient of these two parameters, we design a new combinatorial characteristic parameter based on normalized data. Then, referring to the "snake" idea and the experience in [16], we establish a linear programming model based on the result of hierarchical clustering. By solving, we obtain the optimal network partitioning result. Finally, we apply this method to the real-world traffic data, and the experimental results show that the new combinatorial parameter is more effective than other two single parameters, and the partitioning result highly agrees with the practical subareas.

The remainder of the paper is organized as follows: Section 2 reviews previous researches. Section 3 demonstrates the design of the new combinatorial characteristic parameter, and presents the process of the partitioning method. Section 4 conducts experiments on a real urban transportation network, and shows the effectiveness of the new parameter. Section 5 concludes our research and shows some future work.

II. RELATED WORK

Traffic network partition is a contiguity-constrained clustering problem that aggregates spatial data into homogeneous regions connected spatially. These methods can be divided into two major classes, and the first class mainly uses conventional clustering methods. Based on the basic recursive bisection partition, Wei *et al.* [10] proposed an improved road network partition algorithm for parallel microscopic traffic simulation. Ma *et al.* [11] presented a recursive spectral bisection algorithm to partition road network. Using graph cut technology, Tan *et al.* [4] abstracted the road network and divided the traffic region based on congestion factor that is computed from four traffic parameters. Ji and Geroliminis [5] used the normalized cut algorithm and designed a partitioning mechanism on the clustering of transportation network, and the simulation in the real urban network exhibited its effectiveness and robustness. Saeedmanesh and Geroliminis [13] defined a sequence of roads in a connected homogeneous area as "snake" via clustering and computed their similarities, and then utilized Symmetric Non-negative Matrix Factorization (SNMF) to realize the network partitioning. In [14] and [17], a graph partitioning package is adopted to partition urban road network, and some physical parameters are taken into account. Anwar *et al.* [6] presented a framework for traffic congestion-based spatial partitioning of large urban road networks and developed a two-stage algorithm within this framework. The algorithm first generated a well-structured and condensed density peak graph via clustering and link aggregation, and thereafter devised a spectral theory based on graph cut to partition the network. All these methods require that traffic data must be complete, or else, the partitioned subareas will be incomplete.

The second class of partitioning methods is mainly realized by heuristic optimization. Lu *et al.* [12] realized dynamic division of coordinated control subarea based on a clustering algorithm and correlation degree theory, and a genetic algorithm was used to fast optimize the subarea division. This method can search a sub-optimal solution with high probability under a certain scale network. Zhou *et al.* [7] considered both physical characteristics and dynamic traffic information of road links, and applied an agglomerative hierarchical algorithm, i.e., the heuristic Newman fast algorithm, to divide the traffic network. An *et al.* [9] proposed a four-step heuristic partitioning algorithm, and it utilized the lambda-connectedness concept and the region growing technology-based clustering algorithm to find locally homogeneous sub-networks in the network. Chen [15] proposed three principles of road partitioning and a new model to measure the relationship between two intersections. And based on heuristic community detection, it presented a clustering algorithm to partition the traffic network. Comparing with the first class of partitioning algorithms, the second one only obtains suboptimal partitioning results in many cases, which cannot satisfy practical demands.

III. METHODOLOGY

The clustering method of traffic network partitioning comprises two modules: the selection of traffic characteristic parameters and the design of clustering-based partitioning. Most studies focus on the design and improvement of clustering partition methods and ignore the importance of traffic characteristic parameters. In many cases, they usually choose a single parameter, such as traffic flow, vehicle density, etc. Traffic parameters represent dynamic traffic situation, and directly influence the accuracy of partitioning result and the effectiveness of coordinated signal control. Thus, in this paper, we choose two traffic parameters, i.e., traffic flow and link speed, and design a new combinatorial characteristic parameter based on their correlation. Then, using the idea of "snake" in [13] and referring to the snake algorithm in [16], we establish the linear programming model and then obtain the exact partitioning result by optimizing software. Comparing with heuristic methods or graph cut algorithms, such exact solution have more advantages except high computation.

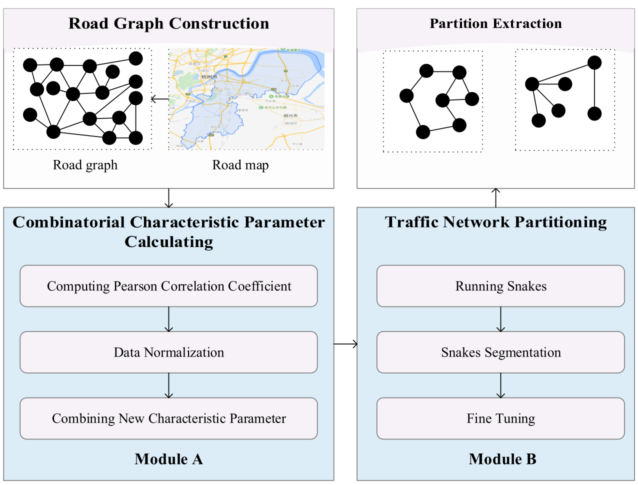

FIGURE 1. Partitioning workflow diagram.

As shown in FIGURE 1, the traffic network can be represented by a graph, and a vertex is a road link while an edge indicates that the two links are connected. Generally, one road has two links, while a one-way road only has one link. The proposed method in this paper includes two modules as follows.

(1) **Module A** - Combinatorial Characteristic Parameter Calculating, which first analyzes the correlation between traffic flow and link speed by Pearson correlation coefficient. Then a new combinatorial characteristic parameter is designed to represent dynamic situation of urban traffic network.

(2) **Module B** - Traffic Network Partitioning, which adopts a three-step clustering-based method to partition a heterogeneously transportation network into a number of homogeneous sub-networks with the guarantee of spatial connectivity.

The details of these two modules are described in the following subsections.

A. COMBINATORIAL CHARACTERISTIC PARAMETER CALCULATING

In the field of urban traffic signal control, there are many characteristic parameters to represent dynamic traffic status, such as travel time, vehicle density, delay, queue length, etc. But in actual urban traffic environment, some parameters are not easy to collect and sometimes the acquisition data has a large deviation. With the development of technology, link speed and traffic flow are relatively easy to acquire at present. Link speed is the average speed of vehicles in a certain road link, and it can be calculated from GPS devices installed on vehicles or mobile phones. Traffic flow is the number of vehicles passing through a link in a period of time. With the widespread use of video detectors (such as gun cameras), traffic flow on urban roads can be obtained by video cameras with the intelligent road vehicle monitoring and recording system. However, these two types of data have their own advantages and disadvantages in practice. The data of link speed can nearly cover all links of the city, but it always has some errors for the calculation only uses a small amount of sample data. With more GPS data, the link speed will be more accurate, while less GPS signals will lead to a large error. Video cameras can accurately record the number of vehicles in a certain direction, but only urban main roads are equipped with such detecting devices. Due to the high cost and the inconvenient installation, cameras cannot cover all roads of the city. Therefore, in this paper, our basic idea is to combine these two characteristic parameters into a new one, so as to make up for their respective disadvantages.

In traffic theory, traffic flow, link speed and vehicle density are three key traffic characteristic parameters. Referring to the fluid theory in physics, their relationship is

$$q_i = v_i k_i \tag{1}$$

where q_i, v_i and k_i are the traffic flow, link speed and vehicle density of road link i, respectively.

Based on (1), Greenshields [18] put forward a linear relationship model of link speed and traffic flow in 1934, which is shown as

$$v_i = v_i^{free}\left(1 - k_i \Big/ k_i^{jam}\right) \tag{2}$$

where v_i^{free} is the link speed of road link i when the number of vehicles is small and every vehicle can travel at a free speed, k_i is the vehicle density of link i, and k_i^{jam} is the vehicle density of link i when the traffic network is blocked seriously. According to (1) and (2), the relationship between link speed and traffic flow is

$$q_i = k_i^{jam}(v_i - v_i^2 \Big/ v_i^{free}) \tag{3}$$

Because the blocking vehicle density k_i^{jam} and the free link speed v_i^{free} are difficult to obtained, we cannot compute the relationship between traffic flow and link speed. So, in this paper, we find a new method to quantify the relationship between link speed and traffic flow directly from the raw collected data.

As shown in FIGURE 1, the new method in **module A** includes three steps, which is described in the following.

1) COMPUTING PEARSON CORRELATION COEFFICIENT

Different from the theoretical model or formulas, relationships between data can be measured by correlation coefficient in the field of big data or data mining. Pearson product-moment Correlation Coefficient (PCC) is a better measurement of linear correlation between two variables, and it was developed by Karl Pearson from a related idea introduced by Francis Galton in the 1880s [19]. PCC is the covariance of the two variables divided by the product of their standard deviations, which can be described as

$$\rho = \frac{cov(Q,V)}{\sigma_Q \sigma_V} = \frac{\sum_{i=1}^{n}(q_i-\bar{q})(v_i-\bar{v})}{\sqrt{\sum_{i=1}^{n}(q_i-\bar{q})^2}\sqrt{\sum_{i=1}^{n}(v_i-\bar{v})^2}} \tag{4}$$

where Q and V denote the average traffic flow and the average link speed of all road links, respectively. $cov(V,Q)$ is the covariance of Q and V, and σ_Q *and* σ_V are the standard deviation of Q and V, respectively. q_i and v_i are the traffic flow and the link speed of road link i, respectively, and $\bar{q}$ and $\bar{v}$ are their average value.

TABLE 1. The relation between values and relevance.

Interdependency	Positive value	Negative value
Uncorrelated	−0.09∼0.0	0.0∼0.09
Low correlation	−0.3∼−0.1	0.1∼0.3
Medium correlation	−0.5∼−0.3	0.3∼0.5
High correlation	−1.0∼−0.5	0.5∼1.0

The value of PCC in (4) is between 1 and −1, where 1 represents total positive linear correlation, 0 is no linear correlation, and −1 is total negative linear correlation. As shown in TABLE 1, the closer the correlation coefficient is to 1 or −1, the more linear the two parameters are. If it tends to be zero, they are uncorrelated.

2) DATA NORMALIZATION

The units of traffic flow and link speed are different and the ranges of their data are also different, so it is necessary to normalize these two kind of data. Data normalization is always needed before clustering, it can scale different kind of data to a specific range and convert the data into dimensionless values. In this way, different parameters can be combined into a dimensionless parameter. The formula of data normalization is

$$\tilde{x}_i = \frac{(x_i-\mu)}{\sigma} \tag{5}$$

where x_i is the value of traffic flow or link speed of link i, and $\tilde{x}_i$ is the normalized value. μ denotes the average value of traffic flow or link speed of all links, and σ denotes the standard deviation. Using data normalization, we can obtain the normalized value of traffic flow and link speed of link i, which is denoted as $\tilde{q}_i$ and $\tilde{v}_i$, respectively.

3) COMBINING NEW CHARACTERISTIC PARAMETER

After PCC calculation and data normalization, we can combine traffic flow and link speed into a new characteristic parameter. We use a linear model as

$$c_i = \rho\tilde{q}_i + (1-\rho)\tilde{v}_i \tag{6}$$

where c_i is the value of the combinatorial characteristic parameter of link i.

As shown in (6), we use the value of PCC as the weight. Considering the theoretical model of traffic flow and link speed in (3), it is reasonable to use such linear combination in (6). Because the value of traffic flow in practical case is more accurate and the value of link speed is more popular, such combination will be better than the single parameter, i.e., traffic flow or link speed.

B. TRAFFIC NETWORK PARTITIONING

This module uses the new combinatorial characteristic parameter computed from the first module and partitions the heterogeneous traffic network into connected homogeneous regions. In the previous research, most partitioning methods are based on traditional clustering or heuristic algorithms. However, these methods have some problems and not suitable in practical traffic network environment. In this paper, we refers to the partitioning method in [16]. Such method uses the concept of ''snake'' in [13] and the model of mixed integer linear programming, and can achieve the exact solution. In addition, the method can guarantee the connectivity of clusters in each step, and is feasible to perimeter control.

As shown in FIGRUE 1, the partitioning method consists of three steps, and the details are described in the following.

1) RUNNING SNAKE

This step is to find adjacent similar links for each road link of the traffic network and use the ''snake'' algorithm which is an agglomerative hierarchical clustering algorithm. For each link, a sequence of links is generated by adding adjacent links iteratively in order to minimize the standard variance of characteristic parameter of all chosen links. This sequence of links is considered as a snake that starts from a link and grows by attracting the most similar adjacent link in each step.

Initially, the algorithm adds a link into the empty snake, and then iteratively add one of the adjacent links to the current snake. When choosing a new link into the snake, its characteristic parameter value is closest to the average parameter value of all links in the snake. This procedure continues until all the links are added. After the running of the snakes, each snake represents a homogeneous connected area around a road that is obtained by absorbing neighboring roads, and the result will be used as a search space in the second step.

2) SNAKE SEGMENTATION

In this step, the main skeleton of each cluster is found based on the results obtained in the first step, and the objective of this step is to choose a few snakes and find homogenous parts of them to partition the traffic network. In this step, a mixing integer linear programming model is used to obtain the exact solution.

The model comprises the optimization objective and some constraints. The optimization objective is to minimize the sum of weighted variance of the selected links. The constraints include the snake segmentation constraint, the number of clusters, the minimum links in the selected snakes, the coverage rate of the selected links, etc. In this model, there are four decision variables. One variable is weighted variance, while other three variables are binary values to decide the link is in a snake, a cluster, multiple clusters or not.

The model can be solved by using the CPLEX optimizer, which is one of the world's leading software packages for solving linear programming, mixer integer programming, and quadratic programming [20]. After solving, the exact partitioning result can be achieved.

3) FINE TUNING

After the second step, some road links may not have been assigned to any one of the clusters or have been assigned to multiple clusters. So these links need to be reassigned to the clusters, and this step is designed to deal with these links.

In this step, a mix integer linear programming model is also built. Such model is similar to the one in the second step except the optimization objective. This step is to minimize the sum of the square of the difference of the parameter value of any two links in all clusters. That is to say, it is to minimize the heterogeneity of all links in each cluster. Likewise, the CPLEX optimizer is also used to solve this linear programming problem. At last, the links that are not assigned to a cluster, as well as those that are allocated to multiple clusters, can find their own clusters.

IV. CASE STUDY AND RESULTS

In this section, the proposed method is implemented on a real transportation road network, and the partitioning results prove the effectiveness.

A. CASE STUDY

The case study is the dataset of Xiaoshan district of Hangzhou. It contains the network structure and daily vehicles data from GPS and video detectors. As shown in FIGURE 2, there are totally 473 intersections connected by 1,312 road links, where a road includes two-direction links and a one-way road has only one link.

When applying regional coordinated traffic signal control, the partitioning of the congested area is vital, and it influences the traffic efficiency of the whole district. So in this section, we choose the main congested area from the dataset, and it involves 55 road links. All these road links form a north-south main road and its branches. This congested area is covered by high-definition cameras and traffic flow can be acquired with high accuracy. And link speed is also accurate because of a number of passing vehicles.

FIGURE 2. Topology map of road network in Xiaoshan District.

The time of the collected data is from 7:45 to 8:00 on the morning of September 19, 2017. This period of time is the morning traffic peak and the traffic is busy and congested. The traffic hotspot in this area is shown in FIGURE 3, the deeper color denotes the roads are more congested. As shown in FIGURE 3, the selected area is highly congested, but the congestion level is different in different subareas. After data processing, the dataset of traffic flow and link speed of the selected area can be obtained.

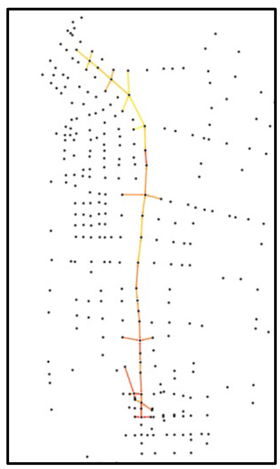

FIGURE 3. Heat map of the selected area.

B. EXPERIMENTS AND RESULTS

In order to verify the effectiveness of the proposed clustering method that uses the new combinatorial characteristic parameter, we design two experiments. In the first experiment, the new combinatorial parameter, traffic flow and link speed are chosen respectively as the characteristic parameter of the clustering-based partitioning method. And in the second experiment, Pearson correlation coefficient (PCC) is compared with other two values as the weight of the combined parameter. In these two experiments, we set number of clusters to two, which matches with the real partition.

Moreover, in order to compare the results of different characteristic parameters and different weights, two metrics, i.e., the normalized total variance TV_N and the similarity rate SR, are defined as follows.

$$TV_N = \frac{\sum_{i=1}^{N_c} N_{A_i} \times var(A_i)}{N \times var(A)} \tag{7}$$

$$SR = \frac{\sum_{i=1}^{N_c} N_{A_i}^s}{N} \tag{8}$$

where A denotes the whole road network, and A_i denotes the i-th cluster, N_C denotes number of clusters, then $A = \bigcup_{i=1}^{N_c} A_i$. N and N_{A_i} denote number of links in the network and in cluster A_i, respectively. And $var(A)$ and $var(A_i)$ denote the variance of characteristic parameter of the links in the corresponding area. $N_{A_i}^s$ denotes number of links in subarea A_i that agree with the real partition.

From (7), TV_N is the ratio of the total variance of the partitioned network to that of the non-partitioned network. The partitioning method is better when TV_N is smaller. From (8), SR is defined as the ratio of the number of links that are in accordance with the actual partitioning scheme to the number of all links. A larger SR indicates that the partitioning result is closer to the actual partition used in practice.

1) DIFFERENT CHARACTERISTIC PARAMETERS

Dynamic traffic information is always represented by characteristic parameters, and different characteristic parameters will lead to different partitioning results. The new combinatorial parameter, traffic flow and link speed are respectively used as the characteristic parameter of the partitioning method. And their partitioning results are shown in FIGURE 4, where FIGURE 4(a) is the result of using traffic flow, FIGRUE 4 (b) is that of using link speed, and FIGURE 4(c) is that of using the new combinatorial parameter. FIGURE 4(d) is the manual partition in running, and its effectiveness has been verified in the practical traffic signal control system for a long time. Because the scheme in FIGURE 4(d) is stable and reliable, it is used as a correct referring to evaluate other partitioning results.

Comparing the four results in FIGURE 4, we can find that only FIGURE 4(c) is almost identical to the actual partition in FIGURE 4(d). And it indicates the combinatorial parameter is more representative of the dynamic traffic information than other two single parameter, i.e., traffic flow and link speed. Observing the results in FIGURE 4, we can find that only the middle road links are not correct in FIGURE 4(a) and (b). When using traffic flow, more links are allocated to the lower cluster (i.e., the blue part), and if using link speed, more links are assigned to the upper cluster (i.e., the red part).

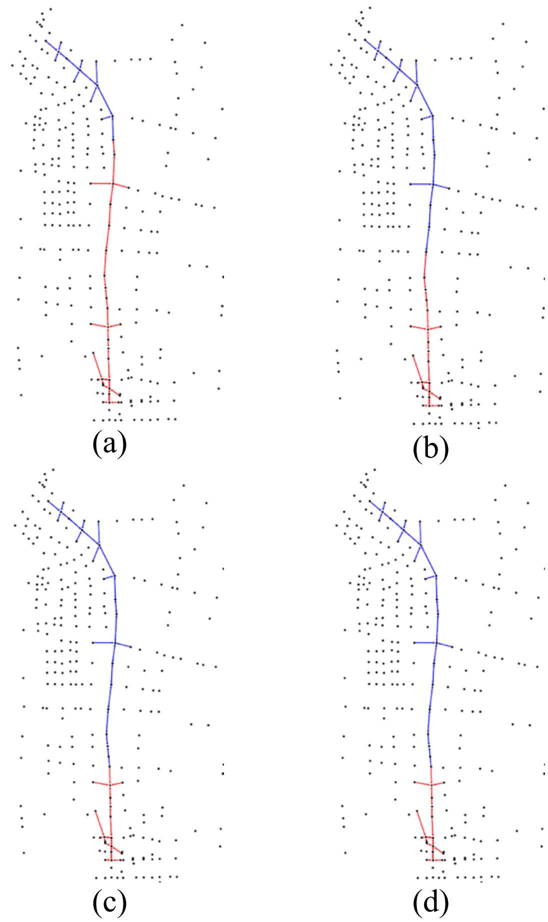

FIGURE 4. Partitioning results (a, b, c) and actual partition (d).

Because neither traffic flow nor link speed is enough to reflect dynamic information of road links, FIGURE 4(a) and (b) are not in accordance with the referring result. But if comparing FIGURE 4(a) and (b), we can find that FIGURE 4(b) is closer to FIGURE 4(d) and only a few of links are not correctly assigned. It implies that link speed is more suitable than traffic flow to represent dynamic traffic situation.

TABLE 2. Metric values using different parameter.

Parameter	TV_N	SR
Combinatorial parameter	0.364	0.96
Link speed	0.416	0.92
Traffic flow	0.635	0.72

TABLE 2 shows the metric values of the partitioning results using different characteristic parameters. As shown in TABLE 2, the TV_N value using the new combinatorial parameter is the least, and it indicates that the variance of the partitioning clusters obtained by using the new parameter is the least. That is to say, the method using the combinatorial parameter can reduce the heterogeneity within the partitions. And other two parameters are not better than this new one.

The SR values in TABLE 2 show that the partitioning result using the combinatorial parameter highly agrees with the actual partition. In fact, we can know that only two links are incorrect from the experimental partitioning data, but comparing FIGURE 4(c) with (d), it is hard to discover this subtle difference. The SR value using link speed is also high, but more links mismatch with the actual partition than using the combinatorial parameter. While the SR value using traffic flow is the lowest among these three values, and its partitioning result is quite different from the actual partition.

2) DIFFERENT WEIGHTS

PCC is the common method to measure the linear correlation between two parameters. In this experiment, we compare PCC with other two values, i.e., 0.5 and 0.8, as the weight of the combined model in equation (6). By computing the traffic data, the PCC value is about 0.2. Using these three weights, the partitioning results are shown in FIGURE 5, and their metric values are shown in TABLE 3.

TABLE 3. Metric values using different weight.

Weight	TV_N	SR
PCC	0.364	0.96
0.5	0.398	0.96
0.8	0.569	0.96

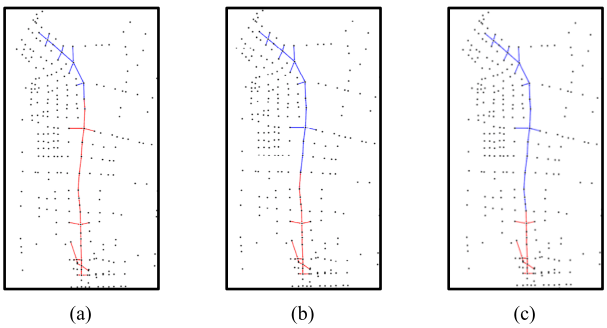

FIGURE 5. Partitioning results using different weights.

In FIGURE 5, PCC, 0.5 and 0.8 are used in (a), (b) and (c), respectively. As shown in FIGRUE 5, these three partitioning results are all the same although their weights are different. And the SR values in TABLE 3 also prove this conclusion. So it indicate that no matter what the combination weight is, the combinatorial parameter will obtain a better partitioning result than single traffic flow or single link speed. That is to say, the combined parameter can make up for the shortcomings of traffic flow and link speed. However, as shown in TABLE 3, the TV_N values of these three weights are different, where PCC's TV_N value is the least, and 0.8's TV_N value is the largest. Therefore, according to the TV_N values and the linear relationship between traffic flow and link speed in (3), it is most reasonable to use PCC than other values as the weight of the combinatorial parameter.

V. CONCLUSION AND FUTURE WORK

In this paper, we analyzed the actual acquisition of two traffic parameters, i.e., traffic flow and link speed. Because a single parameter is not enough to represent dynamic traffic situation, we presented a method to combine traffic flow and link speed into a new combinatorial characteristic parameter by using Pearson correlation coefficient and data normalization. Then, considering disadvantages of traditional clustering and heuristic algorithms, we adopted a snake algorithm to partition the traffic network. The basic idea of such algorithm is to build a linear programming model based on the result of hierarchical clustering and obtain the exact partitioning result by solving. The snake algorithm can avoid falling into local optimization. At last, we conducted experiments on the real traffic data of a Chinese city, and the results proved the effectiveness of our proposed methods.

Although the partitioning method using the new combinatorial parameter achieved a good result, the calculating of the linear programming problem is time-consuming, and it is our next research work in the future. Moreover, we will also study the combinatorial characteristic parameter in a large traffic area to check its performance.

REFERENCES

[1] F. Xia *et al.*, "Exploring human mobility patterns in urban scenarios: A trajectory data perspective," *IEEE Commun. Mag.*, vol. 56, no. 3, pp. 142–149, Mar. 2018. doi: 10.1109/MCOM.2018.1700242.

[2] F. Xia *et al.*, "Modeling and analysis of large-scale urban mobility for green transportation," *IEEE Trans. Ind. Informat.*, vol. 14, no. 4, pp. 1469–1481, Apr. 2018. doi: 10.1109/TII.2017.2785383.

[3] Z. Wu and R. M. Leahy, "An optimal graph theoretic approach to data clustering: theory and its application to image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 11, pp. 1101–1113, Nov. 1993. doi: 10.1109/34.244673.

[4] G. Tan, X. Li, and K. Wang, "The partitioning of traffic signal control network avoiding congestion on boundary," in *Proc. 6th IEEE Joint Int. Inf. Technol. Artif. Intell. Conf. (ITAIC)*, Chongqing, China, Aug. 2011, pp: 361-365.

[5] Y. Ji and N. Geroliminis, "On the spatial partitioning of urban transportation networks," *Transp. Res. B, Methodol.*, vol. 46, no. 10, pp. 1639–1656, Dec. 2012. doi: 10.1016/j.trb.2012.08.005.

[6] T. Anwar, C. Liu, H. L. Vu, and C. Leckie, "Partitioning road networks using density peak graphs: Efficiency vs. accuracy," *Inf. Syst.*, vol. 64, pp. 22–40, Mar. 2017. doi: 10.1016/j.is.2016.09.006.

[7] Z. Zhou, S. Lin, and Y. Xi, "A fast network partition method for large-scale urban traffic networks," *J. Ctrl. Theory Appl.*, vol. 11, no. 3, pp. 359–366, Aug. 2013. doi: 10.1007/s11768-013-2031-0.

[8] G. Shen, C. Chen, Q. Pan, S. Shen, and Z. Liu, "Research on traffic speed prediction by temporal clustering analysis and convolutional neural network with deformable kernels," *IEEE Access*, vol. 6, pp: 51756- 51765, May 2018. doi: 10.1109/ACCESS.2018.2868735.

[9] K. An, Y.-C. Chiu, X. Hu, and X. Chen, "A network partitioning algorithmic approach for macroscopic fundamental diagram-based hierarchical traffic network management," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 4, pp. 1130–1139, Apr. 2018. doi: 10.1109/TITS.2017.2713808.

[10] D. Wei, F. Chen, and X. Sun, "An improved road network partition algorithm for parallel microscopic traffic simulation," in *Proc. IEEE Int. Conf. Mechanic Autom. Ctrl. Eng. (MACE)*, Wuhan, China, Jun. 2010, pp. 2777–2782.

[11] Y.-Y. Ma, Y.-C. Chiu, and X.-G. Yang, "Urban traffic signal control network automatic partitioning using laplacian eigenvectors," in *Proc. 12th IEEE Int. Conf. Intell. Transp. Syst. (ITSC)*, St. Louis, MO, USA, Oct. 2009, pp. 528–532.

[12] K. Lu, J.-M. Xu, S.-J. Zheng, and S.-M. Wang, "Research on fast dynamic division method of coordinated control subarea," *Acta Autom. Sinica*, vol. 38, no. 2, pp. 279–287, Feb. 2012. doi: 10.3724/SP.J.1004.2012.00279.

[13] M. Saeedmanesh and N. Geroliminis, "Clustering of heterogeneous networks with directional flows based on 'Snake, similarities,'" *Transp. Res. B, Methodol.*, vol. 91, pp. 250–269, Sep. 2016. doi: 10.1016/j.trb.2016.05.008.
[14] M. S. Ahmed and M. A. Hoque, "Partitioning of urban transportation networks utilizing real-world traffic parameters for distributed simulation in SUMO," in *Proc. 8th IEEE Veh. Netw. Conf. (VNC)*, Columbus, OH, USA, Dec. 2016, pp. 1–4.
[15] C. Chen, "Dynamic network zoning method based on community detection," in *Proc. 17th IEEE Int. Conf. Intell. Transp. Syst. (ITSC)*, Qingdao, China, Oct. 2014, pp. 2749–2755.
[16] M. Saeedmanesh and N. Geroliminis, "Dynamic clustering and propagation of congestion in heterogeneously congested urban traffic networks," *Transp. Res. B, Methodol.*, vol. 105, pp. 193–211, Nov. 2017. doi: 10.1016/j.trb.2017.08.021.
[17] L. Dimitriou and P. Nikolaou, "Dynamic partitioning of urban road networks based on their topological and operational characteristics," in *Proc. 5th IEEE Int. Conf. Models Technol. Intell. Transp. Syst. (MT-ITS)*, Naples, Italy, Jun. 2017, pp. 457–462.
[18] B. D. Greenshields, "A study of traffic capacity," in *Proc. Highway Res. Board*, Washington, DC, USA, May 1935, pp. 448–477.
[19] M. S. Stigler, "Francis Galton's account of the invention of correlation," *Stat. Sci.*, vol. 4, no. 2, pp. 73–79, Apr. 1989. doi: 10.1214/ss/1177012580.
[20] (2018). *CPLEX Optimizer*. [Online]. Available: https://www.ibm.com/analytics/cplex-optimizer.

DUANYANG LIU was born in Hunan, China, in 1975. He received the B.Sc. degree in mechanical design and manufacture from Xiangtan University, Xiangtan, China, in 1997, and the M.S. degree in mechanical manufacturing and automation and the Ph.D. degree in computer science and technology from Zhejiang University, Hangzhou, China, in 2000 and 2003, respectively.

He is currently an Associate Professor with College of Computer Science and Technology, Zhejiang University of Technology. His current research interests include data mining, machine learning, and intelligent transportation systems.

MENGTING WANG was born in Zhejiang, China, in 1994. She received the bachelor's degree in computer science technology from Zhejiang University of Traditional Chinese Medicine, Hangzhou, China, in 2016. She is currently pursuing the master's degree in computer technology with Zhejiang University of Technology. Her current research interests include data mining, machine learning, and intelligent transportation systems.

GUOJIANG SHEN received the B.Sc. degree in control theory and control engineering and the Ph.D. degree in control science and engineering from Zhejiang University, Hangzhou, China, in 1999 and 2004, respectively. He is currently a Professor with College of Computer Science and Technology, Zhejiang University of Technology. His current research interests include artificial intelligence, theory, big data analytics, and intelligent transportation systems.