IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received April 23, 2018, accepted May 27, 2018, date of publication June 4, 2018, date of current version June 20, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2843564

An Enhanced Clustering-Based Method for Determining Time-of-Day Breakpoints Through Process Optimization

XIANG SONG[1], WENJING LI[2], DONGFANG MA[2], YEZHOU WU[2], AND DAXIONG JI[3]
[1]Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[2]Institute of Marine Sensing and Networking, Zhejiang University, Zhoushan 316021, China
[3]Institute of Marine Robot, Zhejiang University, Zhoushan 316021, China

Corresponding author: Dongfang Ma (mdf2004@zju.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61773337, Grant 61773338, and Grant 61304191, in part by the Zhejiang Provincial Natural Science Foundation under Grant LY17F030009, in part by the Key Research and Development Program of Zhejiang Provincial under Grant 2018C01007, in part by the Fundamental Research Funds for the Central Universities under Grant 2018QNA4050, and in part by the Hangzhou Science and Technology Development Plan Program under Grant 20160533B97.

ABSTRACT The fixed-time strategy is crucial in traffic signal control which applies signal plans to different time periods of the day. One critical step is to determine the optimal breakpoints to divide one day into periods with homogeneous traffic flow. Most existing methods are based on k-means clustering algorithm and have to select the optimal number of clusters. Since direct k-means and time-incorporated k-means clustering will result to noncontiguous time periods, several adjustments are needed including further partitioning and re-clustering via empirically adjustment which merges short time periods into adjacent longer ones to finalize the time-of-day (TOD) partition plan. Such adjustments can make the previous optimal number of clusters selection suboptimal. This paper proposes an enhanced method to determine optimal TOD breakpoints through optimizing the process. Instead of choosing the optimal partition plan before adjustments, we propose to determine the optimum after all the adjustments. A case study based on Qingdao City in China is presented to evaluate the added value of the enhanced method. It is shown through simulation experiments that the enhanced method can avoid over-partitioning and substantially improve the traffic operational efficiency especially during high demand periods.

INDEX TERMS Time-of-day, TOD breakpoint, k-means clustering, empirical adjustment.

I. INTRODUCTION

Traffic signal control is one of the most effective methods to ensure the safety and improve the mobility efficiency in urban transportation systems. A large number of traffic signal control systems have been designed, which can be divided into online and offline systems. Online systems use traffic information collected in real-time from loop detectors to develop responsive signal control strategies [1]–[3]. They can conduct short term traffic flow forecasting based on real-time and historical data [4]–[6]. With the forecasting results, an optimal signal timing plan can be presented with specific control strategy and target. Online systems can be quite effective when facing high fluctuation of the traffic flow, and can improve the operation efficiency of one isolated intersection or even an area especially for peak hours.

However, the loop detectors usually suffer from a fairly high level of failure rate, and the failure rates were reported to be between 24-29% in different states in the United States (U.S.) [7], [8]. Additionally, for the 12,225 traffic signals in New York, more than 95% of them are offline pre-timed, with no detectors deployed at these intersections [9]. Due to many practical issues, a lot of current signal control systems take the fixed time strategy as alternative option, which will be operated when the loop detectors or the fibers to transfer the data from loop detectors become invalid. In addition, the fixed time strategy is also recommended for low traffic demand and is operated between adaptive control periods [10].

In recent years, many emerging sensors such as wireless magnetic sensors, video-imaging sensors, mobile sensors have been used to obtain traffic flow data [11]–[14]. Although these sensors can provide more stable and reliable traffic flow data, they are not connected with existing traffic signal control systems therefore cannot be used for responsive traffic signal control strategies.

The offline systems use historical traffic data as inputs to generate signal plans, and are implemented with fixed time strategies [15], [16], which consist of two steps: (1) determine the optimal breakpoints to divide one day into periods with homogeneous traffic flow, i.e., to determine time-of-day (TOD) plan; and (2) generate the signal timing plans off-line based on traffic conditions within each time period.

In summary, fixed time strategy is still a crucial component in traffic signal control practice. A significant amount of literature has been devoted to signal timing plan optimization with various control objectives including delay time minimum, queue length management, and output maximum for both online and offline systems [17]–[20].

This paper focuses on the other major task for fixed time strategy, i.e., the TOD breakpoint optimization. Although the TOD breakpoint optimization can be crucial for operating efficiency of traffic signal under certain conditions, most existing literature on signal control has been focusing on responsive traffic signal control especially in recent years and limited attention has been paid to this practically important problem. Since there would be many feasible TOD plans, the optimization problem is not easy to solve. A good solution to the time-of-day breakpoints optimization problem can significantly improve the operating efficiency of traffic signal control systems.

II. LITERATURE REVIEW AND MOTIVATION

In the current practice, the experience of traffic engineers and an imprecise analysis of traffic volume data usually determine the current day plan schedules [21], [22]. Specifically, the time of one day is usually spilt into morning peak-hour interval, flat-hump interval, evening peak-hour interval, and off-peak hour interval, which is the general fluctuation pattern within a day [23], [24]. The methods have been proved to be feasible and usually perform well in actual applications. However, they are hard to automate this method as it entirely relies on engineers' subjective judgment.

There are also some algorithm-based approaches, which can be divided into two categories in terms of methodology: artificial intelligence and clustering. The former kind of approaches, such as genetic algorithm (GA) [25], [26], can have good operational performance and computational efficiency, but the results often have to be modified manually due to premature convergence issue [27].

The majority of existing methods for TOD breakpoints optimization are built upon clustering methods. Smith et al. (2001) first used direct k-means (DKM) clustering method, which is the most common clustering analysis method, to divide intervals of one day into similar groups; the method takes the traffic flow and time occupancy in each time interval as input [28]. The DKM cluster-based methods are semi-intelligent as the total number of groups or clusters (denoted by k) has to be set in advance. Later, many other clustering methods were introduced to improve the method proposed by Smith, including improved k-means clustering with subtraction, and spectral clustering [29]–[32].

These methods can generate the optimal breakpoints plan for given number of clusters. Some researchers have recently used two ways to determine the final optimal plan across different values of ks. One is to select the final optimal number of clusters by comparing performance of plans given different values of ks through simulation experiments [33]. The other is to identify optimal number of clusters through statistical methods such as "elbow" method, silhouette measure and gap statistic [34]–[37]. The latter two methods require the use of a "reference distribution" which has to be chosen by the researchers and they may not work very well in practice [38]. As a result, "elbow" method is often used. The "elbow" method plots the sum of intra-cluster variances versus number of clusters and pick the optimal number of cluster as the "elbow" point which uses smallest number of clusters to explain most of the variances in the data.

In the conventional methods above, the optimal number of clusters is selected by "elbow" method after applying k-means clustering across different numbers of clusters. However, as most of the existing methods above have not explicitly treated the traffic data as time series data, the samples of traffic data belonging to the same cluster may need to be further divided into different time periods, since they are not contiguous in the time dimension. In consequence, the number of resulting time periods is usually greater than the corresponding number of clusters, and it often leads to short time periods. These short time periods should be merged into their neighbors, as frequent transitions between different signal timing plans may harm both the safety and operational performance. In actual applications, the merging procedure can be based on engineers' subjective decisions or some rules and is called empirical adjustment. For convenience, the TOD plans before and after empirical adjustment are regarded as original and final plans. Due to the noncontiguous issue and the need of empirical adjustment, the final plan can deviate a lot from optimal number of time periods or clusters determined from the "elbow" method. Therefore, the final plan would likely be suboptimal.

Guo et al. (2014) considers the time of traffic occurring as the additional dimension of input in clustering analysis, called as time-incorporated k-means (TKM) [27]. TKM might mitigate but can not eliminate the issue of having noncontiguous time intervals within one cluster from the clustering result. The mitigation comes from the increased sum of intra-cluster variances if noncontiguous time intervals are selected into one cluster as the differences in time of traffic occurring will be large. It is obvious that it cannot eliminate the issue as TKM does not add constraints or reformulate the problem to directly partition the time-of-day into contiguous time periods. Since the noncontiguous time intervals still exist, additional steps are still needed to adjust the clustering results.

The enhancement for DKM can also be applied with TKM to improve its performance.

The primary purpose of this paper is to enhance the existing approach of determining TOD breakpoints through optimizing the whole process of the method. The main contribution includes (1) illustrate the issue of traditional approach to determine TOD breakpoints; (2) clearly illustrate the details of necessary steps in k-means clustering based TOD breakpoints method including further partitioning, re-clustering via empirical adjustment, optimal cluster number determination through modified "elbow" method; (3) propose a new approach through process optimization to enhance the traditional approach. The remainder of this paper is organized as follows: section three illustrates the main methodology and the detailed process of the proposed method. In section four, we will evaluate the added value of the proposed method against the benchmark through a simulation-based case study based on real data. In the end, we summarize the finding of this paper.

III. TOD BREAKPOINTS OPTIMIZATION

Traffic volumes (or rate of traffic flow, measured in vehicles/hour) at intersection approaches are the most commonly-used clustering elements in existing TOD breakpoints optimization studies [25]–[32], [39], [40], and we also follow the element in this paper. Since k-means clustering algorithms, including DKM and TKM could not deal with contiguous time series data partitioning, several steps need to be taken to adjust the results from k-means clustering including further partitioning and empirical adjustment. As the whole processes based on DKM and TKM are similar and also exist the same limitations, they can be enhanced with the same idea. In this section, we just propose an enhanced method through optimizing the whole process based on DKM clustering, and the TKM-based method can also be improved in the same way. For each possible number of clusters k (e.g., from 2 to 15), first apply DKM clustering algorithm, then further partition each of these clustering results into contiguous partitions, and then conduct empirical adjustment to merge short time periods into neighboring time periods, finally use a modified elbow method to select the optimal partition results from all these adjusted partitions which can be led back to those many k-means clustering results with different values of ks at the beginning.

The enhanced method starts with DKM clustering results with various values of ks, then conduct further partitioning and empirical adjustment for each of clustering results respectively, finally select the optimal partition results (TOD plans). The traditional method also starts with DKM clustering results with various values of ks, then it selects the "optimal" k or "optimal" clustering results. Further partitioning and empirical adjustment is applied on this "optimal" clustering results therefore the final partition result is likely suboptimal. The framework of the comparison between traditional and enhanced method is shown in Fig.1. Both the traditional and enhanced methods contain four steps,

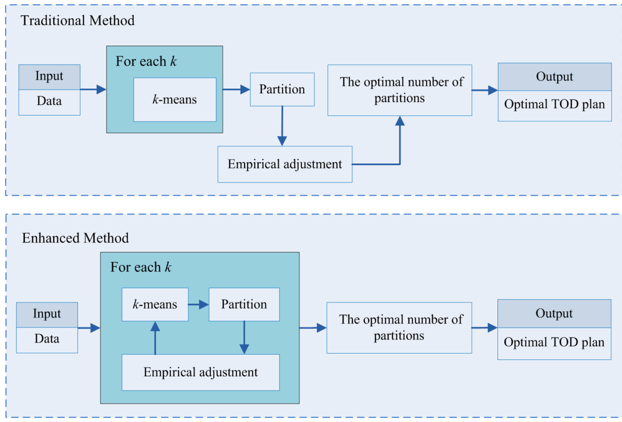


FIGURE 1. The framework of the comparison between traditional and enhanced method.

but some differences still exist in the order of the steps. In the end of this section, we also present overall summary of the traditional and enhanced methods.

In this paper, we follow the common practice that uses data from a historical day as input for TOD breakpoints optimization. More specifically, we divide one-day data into Δ-minute intervals and there are T intervals in total ($T = 1440/\Delta$) for one day. For a specific intersection, there would be multiple lanes from different directions. So the data would be a matrix with rows representing the lanes and columns representing the (Δ-minute) time intervals.

A. CLUSTERING

Let X be the matrix of the entire data and $X = (x_1, x_2, \cdots, x_T)$ where x_t is traffic flow for t-th time interval. The k-means clustering method is to find an optimal solution for the following problem [40], [41].

$$\min_{G_1, G_2, \cdots, G_k} \sum_{i=1}^{k} \sum_{x_t \in G_i} \| x_t - E(x_t) \|_2 . \tag{1}$$

where G_k is the k^{th} time period; k is the number of clusters, which should be fixed before the clustering, $k \in N^+$; $G_1 \bigcup G_2 \bigcup \cdots \bigcup G_k = X = (x_1, x_2, \cdots, x_T)$ and $G_p \bigcap G_q = \varnothing$; $\| \; \|_2$ is l_2 norm of a vector and E is expectation over T time intervals. Any k-means clustering algorithm can be used to solve this problem [43]–[45]. We don't elaborate the solution algorithm here.

To determine optimal TOD plan, we apply k-means algorithms on the data with different inputs of k, and use statistical methods such as elbow or silhouette to determine which inputting number of clusters is optimal. It often requires to explain most variances among data (minimizing sum of intra-cluster variances) with few number of clusters. More details about selecting optimal number of clusters can be found in the section D. However, due to issues of using k-means algorithm for partitioning time series data, additional steps are needed to finalize the TOD plan. Therefore, the evaluation of optimal number of clusters at this stage would be sub-optimal.

B. FURTHER PARTITIONING

Due to noncontiguous issue of resulting clusters from applying k-means clustering algorithms on time series data, both traditional method and enhanced method need to apply an additional step to split or partition those time intervals within the same cluster but not contiguous in time into separate partitions. For example, k-means clustering could cluster time intervals in morning peak and evening peak into one cluster as the values of traffic flow during these intervals may be close to each other, which obviously requires further partitioning.

The algorithm or procedure for further partitioning is very simple. For example, In $G_1, G_2, \cdots, G_k$, if all the clusters contain contiguous time intervals except that G_1 contains $x_1, x_2, x_3, x_7, x_8, x_9$ and G_2 contains x_4, x_5, x_6. Since G_1 contains noncontiguous time intervals, we have to further partition G_1 into $G_1'=(x_1, x_2, x_3)$ and $G_1''=(x_7, x_8, x_9)$. Then we order those new partitions in terms of time. We would have a sequence of $k+1$ partitions as $G_1', G_2, G_1'', G_3, \cdots, G_k$. Fig.2 describes the pseudo code for further partitioning.

Algorithm 1: Pseudocode for further partitioning
For each G_i, i=1,…, k
check whether all the time intervals in G_i are contiguous
If not contiguous, split them into subgroups or partitions with contiguous time intervals.
End for
Re-order all the partitions in terms of time.

FIGURE 2. Pseudo code for further partitioning.

In the enhanced method, we would further partition for each of many clustering results respectively. The resulting number of partitions may be the same even if the starting number of clusters are different. For example, both 3-means clustering and 4-means clustering can result to 7 number of partitions but their breakpoints can be quite different. Note that in traditional method, there is only one clustering result that is "optimal" considering clustering results before further partitioning.

C. RE-CLUSTERING VIA EMPIRICAL ADJUSTMENT

After further partitioning, an issue of over-partitioning rises as there might exist too many short time periods as shown in Fig.3a. Due to safety and operational efficiency concerns, frequent switching signal plans are not desired [27]. Therefore, an empirical adjustment is needed to merge short periods into long periods and eliminate the short time periods. Fig.3b shows the resulting partitions after empirical adjustment.

Although it is common to adjust short periods after further partitioning, no explicit algorithm has been provided in existing literature. In actual applications, those short periods are merged to their adjacent partitions based on the experience of the traffic engineers. This paper first present a detailed empirical adjustment algorithm which can be used for both traditional and enhanced method.

First we specify the desirable length of shortest time period denoted by p (e.g., 12, which means 60 minutes). It's an iterated procedure and we use j to index iteration. Assume that there are in total $N(j)$ partitions for the whole day at iteration j. A TOD partition plan at iteration j can be denoted as $\mathbb{G}^j = (G_1^j, G_2^j, \cdots, G_i^j, \cdots, G_{N(j)}^j)$. G_i^j is the i-th partition at iteration j. At beginning, j=1, and G^1 can be written as follows.

$$\begin{aligned}\mathbb{G}^1 &= \underbrace{[x_1, x_2, \cdots, x_{t_1^1-1}]}_{G_1^1}\underbrace{[x_{t_1^1}, x_{t_1^1+1}, \cdots, x_{t_2^1-1}]}_{G_2^1}, \cdots, \\ &\quad \underbrace{[x_{t_{i-1}^1}, x_{t_{i-1}^1+1}, \cdots, x_{t_i^1-1}]}_{G_i^1}, \cdots \underbrace{[x_{t_{N-1}^1}, x_{t_{N-1}^1+1}, \cdots, x_T]}_{G_N^1}.\end{aligned} \tag{2}$$

Let l_i^1 present the length of partition G_i^1, then

$$l_i^1 = t_i^1 - t_{i-1}^1. \tag{3}$$

As discussed above, G_i^1 should be merged to its former or latter partitions if $l_i^1 < p$.

At the first iteration, we can have two alternative TOD plans by merging G_i^1 to G_{i-1}^1 or G_{i+1}^1 for each short partition or time period. The plan with a lower sum of inter-partition variances should be selected, which is to solve the following problem.

$$\min_{G \in \{G_i^+; G_i^-\}} \sum_G \sum_{x_t \in G} \| x_t - E(x_t) \|_2 . \tag{4}$$

where $G_i^+ = (G_{i-1}^1 \bigcup G_i^1, G_{i+1}^1)$ and $G_i^- = (G_{i-1}^1, G_i^1 \bigcup G_{i+1}^1)$, which denote partition plans after merging to former partition or latter partition respectively.

We keep iterating the procedure until the length of all partitions are greater than p. We use j^* to denote the index of last iteration. The procedure is described as in Fig.4.

D. DETERMINE OPTIMAL TOD PLAN

In this section, we focus on how to determine optimal number of partitions or optimal TOD plan for the enhanced method by introducing the modified elbow method. There have been many attempts to formulate a measure of clustering performance in the past. The proposed measures include Calinski-Harabasz (CH), Davies-Bouldin (DB), Weighted inter-intra (Wint), Akaike information criterion (AIC), Bayesian information criterion (BIC), Deviance information criterion (DIC), sum of intra-cluster variances (SIV). By utilizing those measures, several methods have been proposed to determine the optimal cluster number, such as the elbow method, information criterion approach, and silhouette method. The elbow method using SIV as the measure is

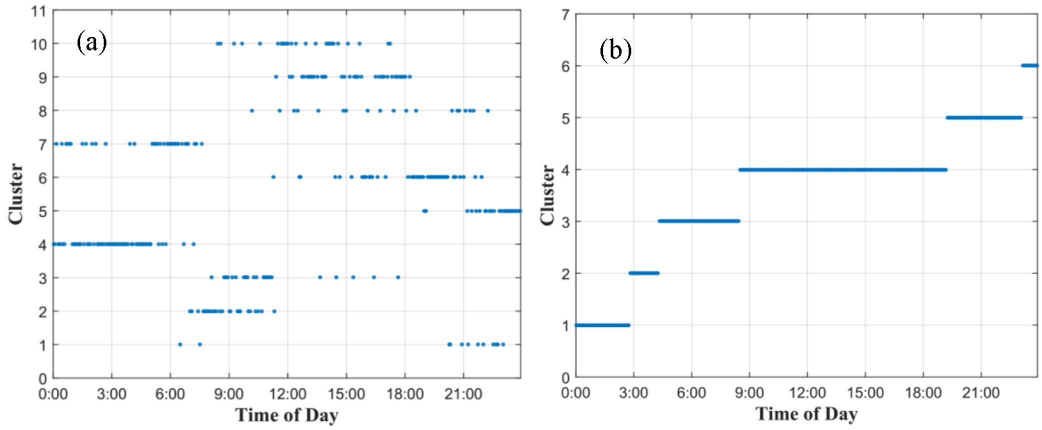


FIGURE 3. (a) TOD plan with clustering; (b) TOD plan after adjustment.

Algorithm 2: Pseudocode for empirical adjustment

Input X and initial partition result G_1

Initialize iteration index j=1

While there exists i-th partition whose length (denoted by l_i^j, for all i=1,…,N(j)) is smaller than p (prefixed threshold), **do**

1. Compute both sums of inter-partition variances for i-th partition G_i^j merging to former or latter adjacent partition
2. Merge the i-th partition G_i^j into the adjacent partition with lowest sum of inter-partition variances

Output partition result G^{j*}

FIGURE 4. Pseudo code for empirical adjustment.

used frequently in real world applications [46]. In this paper, elbow method is selected to determine the optimal number of clusters.

Assume that the whole day is finally divided into k^* partitions after empirical adjustment with k as the initial number of clusters. Then, each k produces a final TOD plan after further partitioning and empirical adjustment, and some different ks will result to the final TOD plans with same number of partitions, which means that each value of k^* is corresponding to multiple values of k. On the other hand, some values of k^* will do not relate to any final TOD plan. From Fig.5, it can be observed that there are no partition results at 2, 6, 9, and 11 partitions but there are more than one partition results at 4, 5, and 10 partitions.

The SIV will decrease with the increasing of number of partitions, and we always want to explain most of the variances by the smallest number of partitions. Thus, we can still use the ideas from existing techniques such as elbow method or silhouette to determine optimal number of partitions. But we need to deal with two issues here. First there

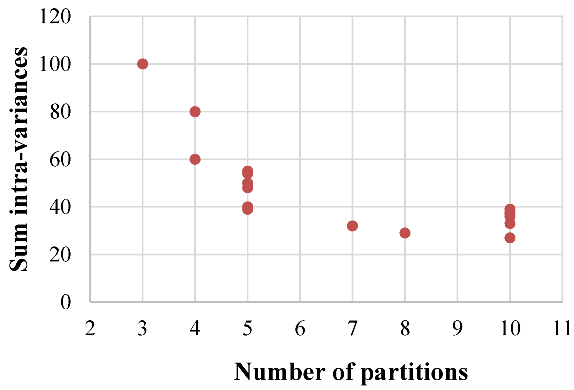


FIGURE 5. Sum of intra-partition variances versus number of partitions.

might be several partition plans given the same number of partitions from different initial inputs of k. Second, there might be no partition plan at certain numbers of partitions.

For the first issue, we select the local optimal partition plan, i.e. the one with lowest sum of intra-partition variances among all partition plans with the same number of partitions. For the second issue, we create quasi-value of sum of intra-partition variances for those numbers of partitions where partition plans are not available. It is created by extrapolating linearly with the known values.

In the elbow method, the reduction in terms of intra-partition variances from having an additional partition is regarded as first-order marginal gain, which is diminishing as the increasing of the number of partitions. We can determine the optimal number of partitions by identifying the highest second-order marginal gains. In other words, we want to find the "elbow" point. We also call the second-order marginal gain as acceleration in this paper and it is denoted by α_k. For $k_{i-1}^*, k_i^*, k_{i+1}^*$, we denote that their local optimal sum of intra-partition variances are $v_{k_{i-1}^*}$, $v_{k_i^*}$ and $v_{k_{i+1}^*}$ respectively, then $\alpha_{k_i^*}$ equals to zero if $v_{k_i^*} - v_{k_{i-1}^*} = v_{k_{i+1}^*} - v_{k_i^*}$.

If there is no partition plan at k^*, we will add a quasi-value of sum of intra-partition variances assuming the reduction rate is linear therefore its acceleration is zero. Let that (k^*-m) and (k^*+n) are the closest number of partitions with at least one partition plans before and after the k^*, then the quasi-value of sum of intra-partition variances at k^* can be computed as

$$v_{k^*} = v_{k^*-m} + \frac{m \cdot (v_{k^*+n} - v_{k^*-m})}{m+n}. \tag{5}$$

Given the set of sum of intra-partition variances with different k^* values denote as $\{v_1, v_2, \cdots, v_{K^*}\}$ where K^* is max number of k^*, the acceleration of k^* partitions can be written as

$$\alpha_k = v_{k^*+1} - 2v_{k^*} + v_{k^*-1}; \quad 2 \le k^* \le N^*. \tag{6}$$

Note that K^* is different from $k_{\max}$ after two steps of adjustments.

We choose the optimal number of partitions when the acceleration reaches its maximum, which can be shown as follows

$$\widehat{K}_E = \arg\max_k(\alpha_k). \tag{7}$$

where $\widehat{K}_E$ is the optimal number of partitions. Note that it's not just a number of partitions but it corresponds to a specific partition plan.

E. SUMMARY

In summary, for the enhanced method based on DKM clustering, we start from many initial values of ks which are imputed into DKM clustering and adjusting each of those many k-means clustering results respectively to have adjusted partitioning results due to noncontiguous issue of clustering algorithm. In the final step, we select which partition results are optimal according to the sum of intra-cluster variances. We use sum of intra-cluster or intra-partition variances to evaluate the partitioning results in all these steps since it is the objective of traffic signal control to have homogenous traffic flow within each time period.

However, for the traditional method, we also start from many initial values of ks which are imputed into DKM clustering. The clustered results of DKM cannot be used directly since a morning time period and an evening time period can be put into the same cluster. However, the traditional method would first choose which k is optimal based on resulting clusters. Then further partition and empirical adjustment would be applied on clustering result of this single "optimal" ks clustering results.

The following Fig.6 and Fig.7 show the high level process of two methods. For detailed implementation of each part, one can refer to previous subsections. The process and summary based on TKM clustering are similar with DKM-clustering-based method.

Algorithm 3: Pseudocode for traditional method

Input data X=(x_1,…,x_T)

1. **For** k=2,…,$k_{\max}$, **do**
   a. k-means algorithm on X
   End for
2. Determine the optimal number of clusters by the elbow method or other statistical methods
3. Further partition noncontiguous clusters into contiguous partitions
4. Empirical adjustment to merge short time periods (partitions)

FIGURE 6. Pseudo code for traditional method.

Algorithm 4: Pseudocode for enhanced method

Input data X=(x_1,…,x_T)

1. **For** k=2,…,$k_{\max}$, **do**
   a. k-means algorithm on X
   b. Further partition noncontiguous clusters into contiguous partitions
   c. Empirical adjustment to merge short time periods (partitions)
   End for
2. Determine the optimal number of partitions by the modified elbow method

FIGURE 7. Pseudo code for enhanced method.

IV. QINGDAO-BASED CASE STUDY

In this section, we present a Qingdao-based case study to show the added value of the enhanced method compared to the existing method. Simulation experiments are conducted based on actual infrastructure characteristics and traffic flow data of an intersection in Qingdao China.

The frequent-used time intervals for TOD breakpoints optimization are 5 minutes and 15 minutes [47], which are determined by the traffic flow data counting interval of the traffic control system. The traffic flow data may changed severely in a short time period, and a short counting interval can result in more appropriate TOD breakpoints, ensuring that the breakpoints are close to the severely changed time points. Thus, a short time interval to counting traffic flow is a better choice for TOD optimization than any longer one. As the signal controller belonging to the studied intersection count the traffic flow data with 5-minute interval, thus we presented the case study with 5-minute traffic flow data.

A. BACKGROUND AND DATA

We collected the data at an intersection of Jiangxi Road and Fuzhou South Road at Qingdao City in China. Detailed information about this intersection is as follows. (1) There are two lanes at the east and west approaches, and there are six and five lanes at the north and south approaches respectively. The channelization and order phase sequence were shown in Fig.8. (2) The length of four links, i.e., the north, south, west and east approaches are 562m, 426m, 413m and 443m respectively. (3) The intersection now implements a fixed-time strategy, and a whole day is divided into four time periods including morning peak hours, afternoon peak hours, off-peak period and mid-night period. (4) There are no road-side parking spaces.

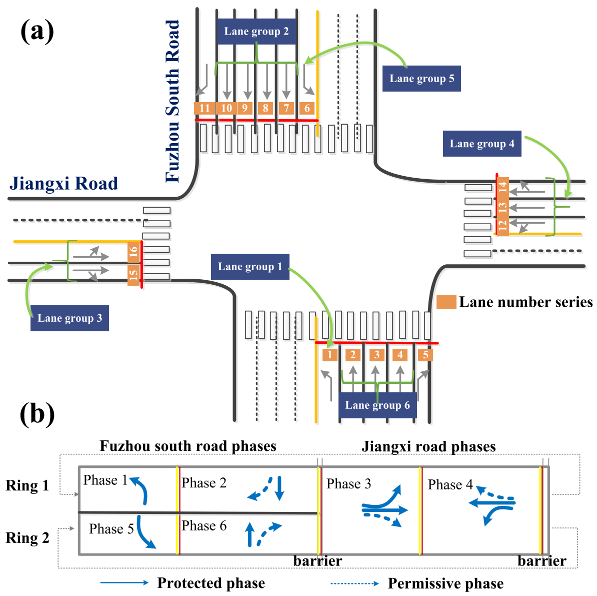


FIGURE 8. Illustration of the intersection: (a) channelization; (b) phases order.

The currently-used signal timing plan at the intersection contained four time periods, namely, 0:00-7:00, 7:00-17:00, 17:00-20:00 and 20:00-24:00. Since there is no ''no-turn-on-red'' restriction at this intersection, we ignore the right turning traffic. We use the traffic flow data from six lane groups including through lanes of north and south approaches, left turning lanes of north and south approaches, lanes of east approach and lanes of west approach.

The Automatic Violation Detecting and Recording System (AVDRS) was installed at this intersection, and at least one video-imaging detectors were assigned to each approach. This new kind of detectors collects the lane-based traffic flow, speed, and other parameters [48], and we use the traffic flow data in July 2016 for the case study. Fig.9a and Fig.9c show the traffic flow patterns of six lane groups on July 4 and 5; Fig.9b and Fig.9d show the aggregate traffic flow pattern of the same day. In the latter part of the case study, our analysis focus on these two days which have quite different traffic flow patterns. On July 4, there are two peak hours, namely 7:30-8:30 and 17:30-18:00. On July 5 however, there is no palpable peak hours.

B. TOD PLAN

We use the traffic flow data from six lane groups with five minutes as the intervals, and calculate the TOD partition plan using the two methods described in section 3. The optimal number of clusters after k-means algorithms in traditional method and the optimal number of partitions after all the adjustments in enhanced methods for two test days are shown in Fig.10. The elbow points can be identified by the maximum acceleration as defined in section 3. We can observe that 3 is chosen to be optimal number of clusters which is used for traditional DKM-based method but note that it is not the final number of partitions after adjustments. The enhanced DKM-based method selects 6 to be optimal number of partitions. Results for the traditional and the enhanced TKM-based methods can also be found in Fig.10. Fig.11 shows sixteen graphs of TOD partition results including intermediate results. The first row of four pictures are for July 4 and the second is for July 5 based on DKM algorithm. The other eight figures are for the traditional and enhanced TKM-based method.

Meanwhile, we calculate the sums of the intra-partition variances for the final TOD partition plans with the traditional and enhanced methods for the two days, which are shown in Table 1.

TABLE 1. Sum of intra-variance of the final plans on different days.

Date and method	Number of partitions	Intra-partition variances
T-DKM on July.4	10	6871.7
E-DKM on July.4	5	7722.6
T-TKM on July.4	13	6475.8
E-TKM on July.4	6	7236.5
T-DKM on July.5	5	9547.6
E-DKM on July.5	6	8717.6
T-TKM on July.5	5	9159.4
E-TKM on July.5	5	8521.8

Notes: *T-DKM=Traditional direct k-means; E-DKM=Enhanced direct k-means;T-TKM=Traditional time-incorporated k-means; E-TKM=Enhanced time-incorporated k-means.*

From Table 1, we can see that the enhanced methods for DKM or TKM result to smaller sum of intra-partition variances for July 5, which implies that it is a better way to determine the TOD breakpoints using the enhanced methods. The number of partitions of the final TOD plan with the traditional method on July 4, including DKM and TKM, is twice as many as the enhanced methods; however, it just brings a slight decrease in terms of the sum of intra-partition variances, which implies that the traditional methods may over-partition the time of day. We use simulation experiments to illustrate the operational performance under two different methods.

C. SIMULATION EXPERIMENTS

The most effective evaluation experiment is to collect and analyze the actual traffic flow data under different plans at one or more intersections during the same time periods. However, such evaluation experiment, is not feasible in real life for various reasons. For example, the traffic flow patterns

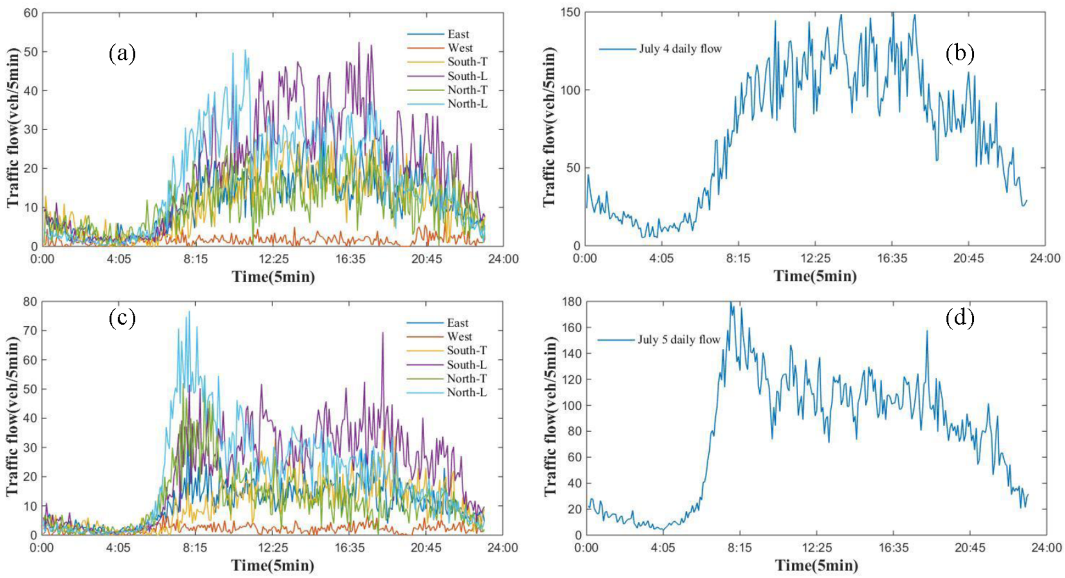


FIGURE 9. (a) (c) traffic flow of six lane groups on July 4 and 5; (b) (d) aggregate traffic flow on July 4 and 5.

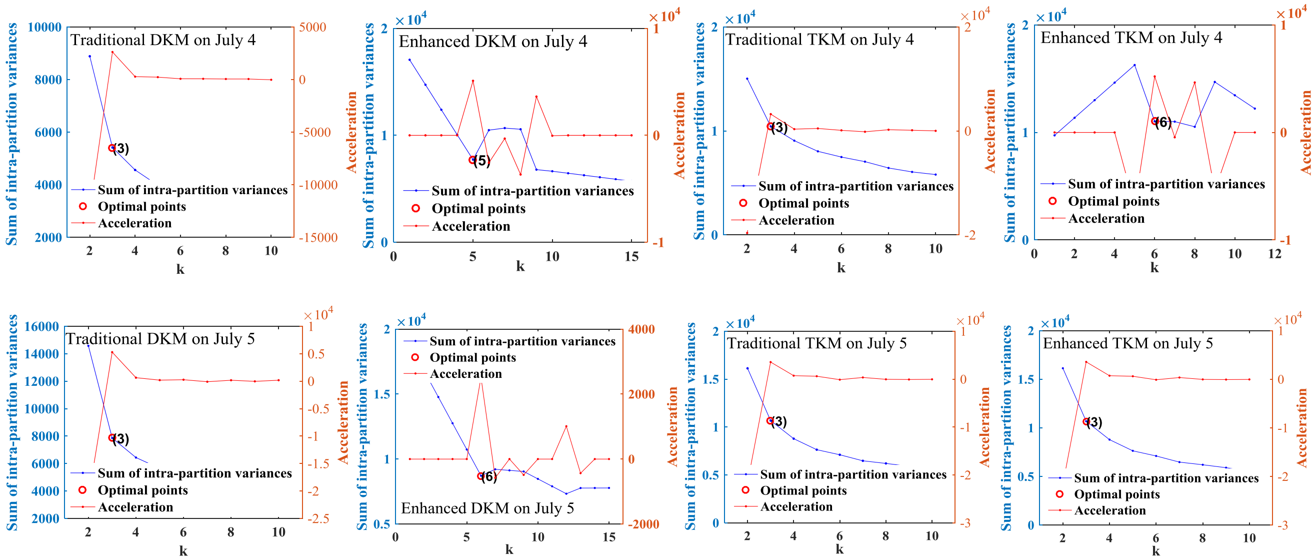


FIGURE 10. Optimal number of clusters for traditional and enhanced methods on July 4 and 5.

are different every day, it is impossible to select two days with the exactly same traffic pattern. Therefore, in this case study, we evaluate the enhanced method against the traditional one through simulation experiments, where TOD plans optimized by different methods are applied under the same traffic condition.

The simulation model is built by VISSIM, a microscopic simulation software developed by PTV Company [49]. Traffic flow data are imputed with 5 min as the interval, shown as in Fig.9. For each time period, we optimize the signal timing plan using Webster algorithm [50], [51], and assign the split for each phase with the principle to ensure saturation degree among different phases uniform.

$$C_0 = \frac{1.5L+5}{1-\sum_{p=1}^{Z}\frac{q_p}{s_p}}$$
$$g_p = \max\{g_{p,\min}, \frac{q_p/s_p}{\sum_{p=1}^{Z}\frac{q_p}{s_p}}\}. \quad (8)$$

where C_o is the optimum cycle length (s); L is sum of the lost time for all phases, usually taken as the sum of the intergreen periods (s), in this paper, and this value is set to be 20; q_p/s_p is the ratio of the flow rate to the saturation flow rate for the critical approach p or lane p in each phase, and we

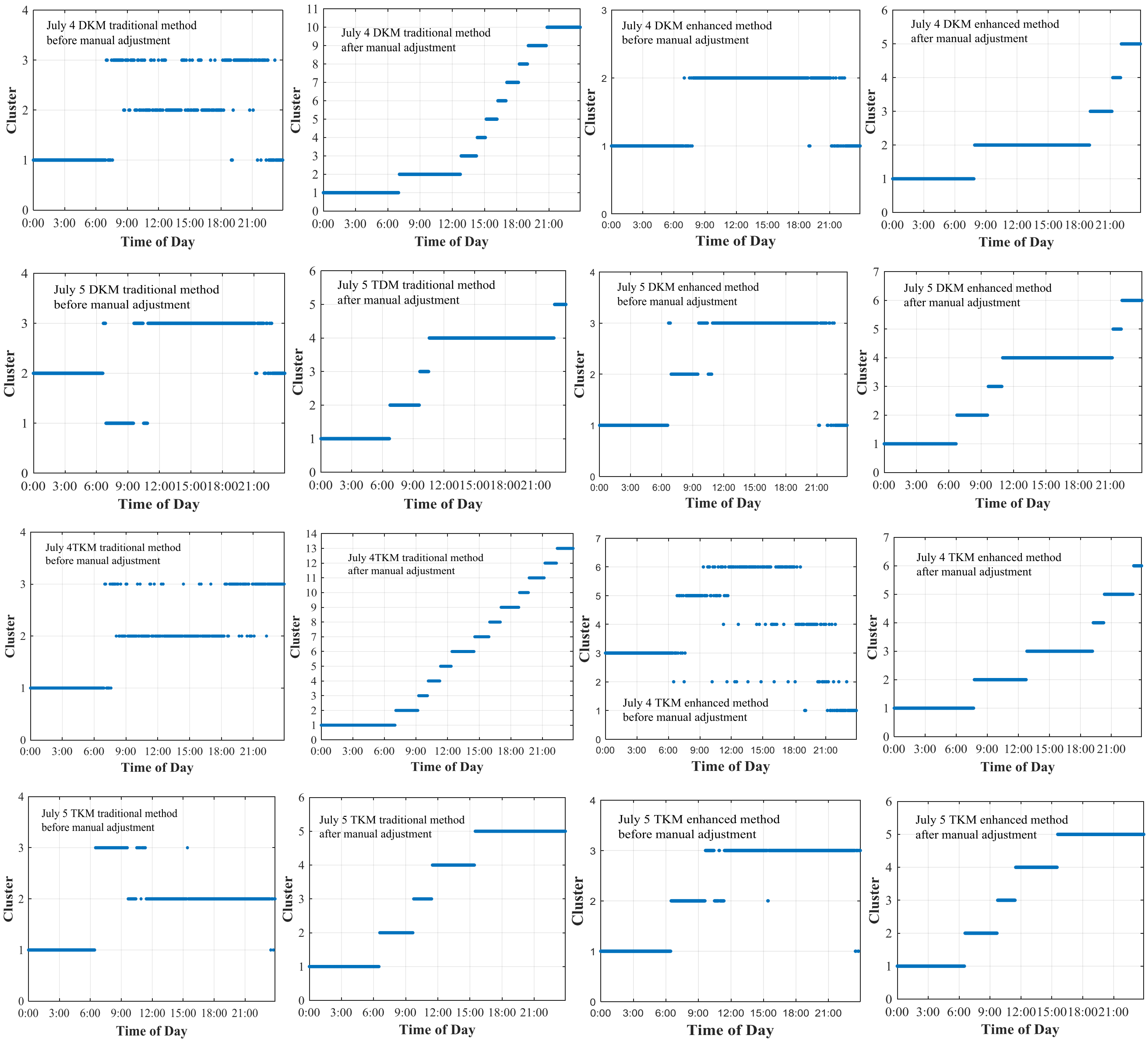


FIGURE 11. Partition results with different methods.

set the saturation flow rate for each lane is $0.5pcu/s$; Z is the number of phases; g_p is the green time for phase p (s); $g_{p,\min}$ is the minimum green time to ensure the safety of cross-street pedestrians, and it set as $15s$ in the simulations. The final TOD plans, the green time for each phase and the cycle length within different time periods (partitions) determined by DKM and TKM clustering method are shown in Table 2.

Under given traffic flow, the operation performance is also influenced by the distribution of the vehicle arrivals. In VISSIM simulations, the distribution of the vehicle arrivals vary with the change of the parameter of random seed. The simulation was run 10 times for each day, with the random seed varying from 42 to 51. and the average results of the ten simulations are used to compare different TOD plans.

D. RESULT ANALYSIS

There are many performance measures to assess the operation efficiency of traffic signal control including maximum queue length (MQL), delay time (DT), average throughput(AT), stop time (ST), and fuel consumption (FC). According to High Capacity Manual (HCM) and Traffic Signal Timing Manual (TSTM), the first two is the primary performance measures at individual intersections [52].

VISSIM can output the two performances with a given time intervals, and we use fifteen minutes as the time period in this paper. Fig.12 illustrates the two performance measures under TOD plans generated by the traditional and the enhanced methods across the whole day for July 4 and 5. We also optimized the signal timing plans using Webster

TABLE 2. Signal timing parameters under different time periods.

Date and Clustering method	Enhanced method						Traditional method					
	breakpoint	Phase green time				Cycle length	breakpoint	Phase green time				Cycle length
		E	W	NS-T	NS-L			E	W	NS-T	NS-L	
7.4 DKM	00:00-07:55	15	15	15	15	60	00:00-07:05	15	15	15	15	60
	07:55-19:05	24	15	44	24	107	07:05-12:50	17	15	33	15	80
	19:05-21:15	21	15	33	19	89	12:50-14:20	23	15	49	26	114
	21:15-22:05	15	15	30	15	75	14:20-15:10	20	15	33	21	88
	22:05-24:00	15	15	15	15	60	15:10-16:15	19	15	37	21	91
	--	--	--	--	--	--	16:15-17:05	18	15	47	21	101
	--	--	--	--	--	--	17:05-18:15	22	15	44	21	102
	--	--	--	--	--	--	18:15-19:05	15	15	27	17	75
	--	--	--	--	--	--	19:05-20:50	18	15	28	16	77
	--	--	--	--	--	--	20:50-24:00	15	15	23	15	68
7.4 TKM	00:00-07:45	15	15	15	15	60	00:00-08:00	15	15	18	15	63
	07:45-12:50	25	15	53	27	120	08:00-10:15	19	15	46	22	102
	12:50-19:15	22	15	38	21	96	10:15-11:25	18	15	34	19	86
	19:15-20:20	21	15	34	20	80	11:25-12:25	18	15	36	20	89
	20:20-23:10	17	15	23	16	71	12:25-13:45	20	15	38	17	90
	23:10-24:00	15	15	15	15	60	13:45-16:10	16	15	30	18	78
	--	--	--	--	--	--	16:10-18:10	24	15	48	25	112
	--	--	--	--	--	--	18:10-19:15	22	15	43	22	102
	--	--	--	--	--	--	19:15-19:55	17	15	33	18	83
	--	--	--	--	--	--	19:55-20:20	17	15	30	17	79
	--	--	--	--	--	--	20:20-21:10	18	15	28	16	77
	--	--	--	--	--	--	21:10-22:55	15	15	23	15	68
	--	--	--	--	--	--	22:55-24:00	15	15	17	15	62
7.5 DKM	00:00-06:45	15	15	15	15	60	00:00-06:45	15	15	15	15	60
	06:45-09:40	25	15	82	50	173	06:45-09:40	25	15	82	50	173
	09:40-11:00	17	15	38	20	90	09:40-10:35	17	15	34	19	85
	11:00-21:15	17	15	37	25	94	10:35-22:50	16	15	36	24	90
	21:15-22:05	15	15	30	19	79	22:50-24:00	15	15	19	15	64
	22:05-24:00	15	15	22	15	67	--	--	--	--	--	--
7.5 TKM	00:00-06:35	15	15	15	15	60	00:00-06:35	15	15	15	15	60
	06:35-09:45	25	15	82	50	173	06:35-09:45	25	15	82	50	173
	09:45-11:30	17	15	32	17	81	09:45-11:30	17	15	32	17	81
	11:30-15:35	18	15	38	22	93	11:30-15:30	18	15	39	23	95
	15:35-24:00	22	15	31	16	84	15:30-24:00	22	15	31	16	84

Notes: *E and W are the phase for east and west approach respectively; NS-T and NS-L are the through and left-turning phases respectively both for the south and north approaches.*

algorithm together with the real-used TOD plan, and then presented the performance results in Fig.12. In Fig.12, each point represents the average performance within the time interval of 15 minutes; plans 1, 2, 3, 4 and 5 are currently-used, traditional DKM-based [28], enhanced DKM-based, traditional TKM-based [27] and enhanced TKM-based ones. We can observe that the the performance currently-used plan is the worst compared with the four optimization ones on both the two days, which indicates that both the two clustering methods, traditional or enhanced ones, are useful for traffic signal control. Besides, on July 5, the enhanced method is, during most time of the day, better than traditional one. However, on July 4, it is hard to tell which one is better. The traditional method divided the day into ten partitions,

TABLE 3. Comparison of the performance measures with the enhanced and traditional methods.

Date and clustering method	Whole day						06:30-19:00					
	MAQ1/m	MAQ2/m	Rc/%	AD1/s	AD2/s	Rc/%	MAQ1/m	MAQ2/m	Rc/%	AD1/s	AD2/s	Rc/%
July 4 DKM	86.51	85.19	1.55	55.29	54.97	0.58	132.11	132.71	-0.45	81.73	79.23	3.16
July 5 DKM	66.67	77.42	-13.89	53.63	58.10	-7.69	107.42	126.82	-15.30	79.34	87.13	-10.98
July 4 TKM	84.13	83.53	0.72	53.51	52.55	1.83	136.01	135.67	0.25	77.58	76.24	1.76
July 5 TKM	74.65	75.08	-0.57	55.82	55.94	-0.21	120.69	121.50	-0.67	81.11	81.34	-0.28

Notes: *MAQ1 and MAQ2 are the maximum queue length with the new and traditional methods respectively; AD1 and AD2 are the average delay time with the new and traditional methods respectively; Rc1 and Rc2 are relative change rates for maximum queue length and average delay respectively.*

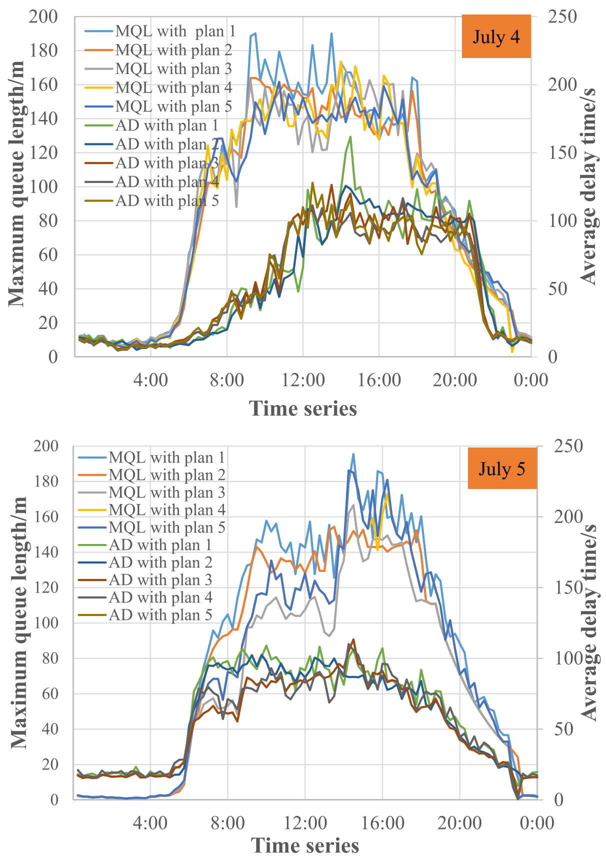


FIGURE 12. Performances under different methods on July 4 and 5.

about twice comparing with the enhanced method; however, the performance measures did not improved significantly, implying that the day was over-partitioning by the traditional methods. The plans with the enhanced methods are better choices for traffic signal control.

During low traffic flow period, for example 00:00-06:30, the signal timing plans are similar across different days, during which the minimum cycle length are used to ensure safety of street-crossing pedestrians. As a result, the performance measures during these periods fluctuate slightly. For better illustration, we select the relatively high traffic flow hours (06:30-20:00) as the observational periods to evaluate the advantage of the proposed method.

In order to quantity the average change of performance measures of enhanced method against traditional method, the relative change rate for each measure is introduced here, which can be calculated by

$$R_c = \frac{y_e - y_t}{y_t}. \tag{9}$$

where R_c is the relative change rate for performance measures during one certain time period; y_e is the measure value with the enhanced method; y_t is the measure value with the traditional method.

The comparisons across the whole day and the high traffic flow time period on both July 4 and 5 are summarized in Table 3. On July 5, the new and traditional DKM-based methods partition the whole day into five and six partitions, and the enhanced method reduces MAQ and AD by 13.89% and 7.69% respectively on average for whole day; meanwhile, for the time periods from 06:30 to 20:00, the improvements are respectively 15.30% and 10.98%, which indicate that the enhanced method can improve the operation efficiency of traffic flow for the traffic signals with fixed-time strategy, especially during high traffic flow period.

On July 4, the average MAQ and AD with the enhanced DKM-cluster method are increased by 1.55% and 0.58% respectively compared to the traditional method for the whole day, which means that the new method performs slightly worse than the traditional method. However, we should note that the number of partitions with the enhanced and traditional methods are respectively 5 and 10, thus, the transition times of signal timing plans with the traditional method are twice as much as the enhanced. One important principle for TOD is to minimize the transition times if the operation efficiency can be guaranteed. However, the improvements from having extra partitions are very minor or there is even no improvement for MAQ during high traffic periods. Thus, we can make a conclusion that the traditional methods sometimes produce over-partitioning plan, which can be avoided by the enhanced methods.

For the the traditional and enhanced methods based on TKM clustering algorithms, a similar conclusion can be deduced by analyzing the results in Table 3. Thus, the enhancement of process optimization is helpful to improve the performance of TOD for both DKM and TKM.

V. CONCLUSION

In this paper, due to issues of traditional clustering-based algorithms for TOD breakpoint optimization, we propose an enhanced method with optimized process. The main difference between the two methods are when to select the optimal number of partitions/clusters given the need of empirical adjustment after k-means clustering algorithm. We point out that the traditional method would lead to suboptimal partition plans by selecting optimal number of clusters before adjustments. Additionally, we also present procedure of empirical adjustment and how to determine optimal number of partitions given complicated partition results, which have not been shown in previous studies.

We present a case study which compares the traditional and enhanced methods through simulation experiments based on the data collected by AVDRS. Then, we presented a experiment to evaluate the benefits of the enhanced method with VISSIM simulation, after determining the signal timing parameters with a common algorithm. The results show the new method can avoid the issue of over-partitioning, and it can perform better than the traditional method in general when traditional method does not generate too many partitions.

REFERENCES

[1] E. Christofa, I. Papamichail, and A. Skabardonis, "Person-based traffic responsive signal control optimization," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 3, pp. 1278–1289, Sep. 2013, doi: 10.1109/TITS.2013.2259623.

[2] R. Evers and S. Proost, "The myth of traffic-responsive signal control: Why common sense does not always make sense," *Transp. Res. A, Policy Pract.*, vol. 77, pp. 350–357, Jul. 2015, doi: 10.1016/j.tra.2015.05.004.

[3] D. Ma *et al.*, "Gating control for a single bottleneck link based on traffic load equilibrium," *Int. J. Civil Eng.*, vol. 14, no. 5, pp. 281–293, Jun. 2016, doi: 10.1007/s40999-016-0043-0.

[4] Z. Zhao, W. Ding, J. Wang, and Y. Han, "A hybrid processing system for large-scale traffic sensor data," *IEEE Access*, vol. 3, pp. 2341–2351, Nov. 2015, doi: 10.1109/access.2015.2500258.

[5] D. Xia, H. Li, B. Wang, Y. Li, and Z. Zhang, "A map reduce-based nearest neighbor approach for big-data-driven traffic flow prediction," *IEEE access*, vol. 4, pp. 2920–2934, Jun. 2016, doi: 10.1109/access.2016.2570021.

[6] O. Tayan, Y. M. Alginahi, M. N. Kabir, and A. M. Al Binali, "Analysis of a transportation system with correlated network intersections: A case study for a central urban city with high seasonal fluctuation trends," *IEEE Access*, vol. 5, pp. 7619–7635, Apr. 2017, doi: 10.1109/access.2017.2695159.

[7] *Traffic Detector Handbook*, 3rd ed., Federal Highway Admin., Washington, DC, USA, 2006.

[8] M. M. Ahmed and M. A. Abdel-Aty, "The viability of using automatic vehicle identification data for real-time crash prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 2, pp. 459–468, Oct. 2013, doi: 10.1109/TITS.2011.2171052.

[9] X. J. Ban, P. Hao, and Z. Sun, "Real time queue length estimation for signalized intersections using travel times from mobile sensors," *Transp. Res. C, Emerg. Technol.*, vol. 19, no. 6, pp. 1133–1156, 2011, doi: 10.1016/j.trc.2011.01.002.

[10] M. Papageorgiou, C. Diakaki, V. Dinopoulou, A. Kotsialos, and Y. Wang, "Review of road traffic control strategies," *Proc. IEEE*, vol. 91, no. 12, pp. 2043–2067, Dec. 2003, doi: 0.1109/JPROC.2003.819610.

[11] D. Ma *et al.*, "Traffic demand estimation for lane groups at signal-controlled intersections using travel times from video-imaging detectors," *IET Intell. Transp. Syst.*, vol. 11, no. 4, pp. 222–229, May 2017, doi: 10.1049/iet-its.2016.0233.

[12] P. Zheng and M. McDonald, "Estimation of travel time using fuzzy clustering method," *IET Intell. Transp. Syst.*, vol. 3, no. 1, pp. 77–86, Mar. 2009, doi: 10.1049/iet-its:20080021.

[13] L. Li, D. Wen, and D. Yao, "A survey of traffic control with vehicular communications," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 1, pp. 425–432, Feb. 2014, doi: 10.1109/TITS.2013.2277737.

[14] T. Tettamanti, T. Luspay, B. Kulcsár, T. Péni, and I. Varga, "Robust control for urban road traffic networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 1, pp. 385–398, Feb. 2014, doi: 10.1109/TITS.2013.2281666.

[15] T.-H. Chang and J.-T. Lin, "Optimal signal timing for an oversaturated intersection," *Transp. Res. B, Methodol.*, vol. 34, no. 6, pp. 471–491, Aug. 2000, doi: 10.1016/S0191-2615(99)00034-X.

[16] L. Zhao, X. Peng, L. Li, and Z. Li, "A fast signal timing algorithm for individual oversaturated intersections," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 1, pp. 280–283, Mar. 2011, doi: 10.1109/TITS.2010.2076808.

[17] A. Stevanovic, J. Stevanovic, K. Zhang, and S. Batterman, "Optimizing traffic control to reduce fuel consumption and vehicular emissions: Integrated approach with VISSIM, CMEM, and VISGAOST," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2128, pp. 105–113, Dec. 2009, doi: 10.3141/2128-11.

[18] Z. He, L. Zheng, P. Chen, and W. Guan, "Mapping to cells: A simple method to extract traffic dynamics from probe vehicle data," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 32, no. 3, pp. 252–267, Mar. 2017, doi: 10.1111/mice.12251.

[19] Z. He, L. Zheng, and W. Guan, "A simple nonparametric car-following model driven by field data," *Transp. Res. B, Methodol.*, vol. 80, no. 10, pp. 185–201, Oct. 2015, doi: doi.org/10.1016/j.trb.2015.07.010.

[20] G. Zhang and Y. Wang, "Optimizing minimum and maximum green time settings for traffic actuated control at isolated intersections," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 1, pp. 164–173, Mar. 2011, doi: 10.1109/TITS.2010.2070795.

[21] J. Zheng, H. Liu, and S. Misgen, "Fine-tuning time-of-day transitions for arterial traffic signals," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2488, pp. 32–40, Oct. 2015, doi: 10.3141/2488-04.

[22] A. Muralidharan, S. Coogan, C. Flores, and P. Varaiya, "Management of intersections with multi-modal high-resolution data," *Transp. Res. C, Emerg. Technol.*, vol. 68, pp. 101–112, Jul. 2016, doi: 10.1016/j.trc.2016.02.017.

[23] S. Lavrenz, J. Sturdevant, and D. Bullock, "Strategic methods for modernizing traffic signal maintenance management and quantifying the impact of maintenance activities," *J. Infrastruct. Syst.*, vol. 23, no. 4, p. 05017004, Dec. 2017, doi: 10.1061/(ASCE)IS.1943-555X.0000361.

[24] A. D. Lidbe, E. G. Tedla, A. M. Hainen, and S. L. Jones, "Analytical techniques for evaluating the implementation of adaptive traffic signal control systems," *J. Transp. Eng., A, Syst.*, vol. 143, no. 5, p. 04017011, May 2017, doi: 10.1061/JTEPBS.0000034.

[25] B. Park, D.-H. Lee, and I. Yun, "Enhancement of time of day based traffic signal control," in *Proc. IEEE Int. Conf. Syst., Man Cybern.*, Washington, DC, USA, vol. 4, Oct. 2003, pp. 3619–3624.

[26] M. M. Abbas and A. Sharma, "Optimization of time of day plan scheduling using a multi-objective evolutionary algorithm," in *Proc. 84th Annu. Meeting Transp. Res. Board*, Washington, DC, USA, Jan. 2005.

[27] R. Guo and Y. Zhang, "Identifying time-of-day breakpoints based on nonintrusive data collection platforms," *J. Intell. Transp. Syst.*, vol. 18, no. 2, pp. 164–174, Jun. 2014, doi: 10.1080/15472450.2013.802151.

[28] B. L. Smith, W. T. Scherer, T. A. Hauser, and B. B. Park, "Data–driven methodology for signal timing plan development: A computational approach," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 17, no. 6, pp. 387–395, Nov. 2002, doi: 10.1111/1467-8667.00285.

[29] N. T. Ratrout, "Subtractive clustering-based K-means technique for determining optimum time-of-day breakpoints," *J. Comput. Civil Eng.*, vol. 25, no. 5, pp. 380–387, Sep. 2011, doi: 10.1061/(ASCE)CP.1943-5487.0000099.

[30] Y. K. Wong and W. L. Woon, "An iterative approach to enhanced traffic signal optimization," *Expert Syst. Appl.*, vol. 34, no. 4, pp. 2885–2890, May 2008, doi: 10.1016/j.eswa.2007.05.007.

[31] Y. Jun and Y. Yang, "Using Kohonen cluster to identify time-of-day break points of intersection," in *Emerging Technologies for Information Systems, Computing, and Management*. New York, NY, USA: Springer, 2013, pp. 889–896.

[32] C. Dong, Y. Su, and X. Liu, "Research on TOD based on Isomap and K-means clustering algorithm," in *Proc. 6th Int. Conf. Fuzzy Syst. Knowl. Discovery*, Tianjin, China, vol. 1, Aug. 2009, pp. 515–519.

[33] B. B. Park, P. Santra, I. Yun, and D.-H. Lee, "Optimization of time-of-day breakpoints for better traffic signal control," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1867, pp. 217–223, Aug. 2004, doi: 10.3141/1867-25.

[34] T. M. Kodinariya and P. R. Makwana, "Review on determining number of cluster in K-means clustering," *Int. J. Adv. Res. Comput. Sci. Manage. Stud.*, vol. 1, no. 6, pp. 90–95, 2013.

[35] Q. Ni, Q. Pan, H. Du, C. Cao, and Y. Zhai, "A novel cluster head selection algorithm based on fuzzy clustering and particle swarm optimization," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 14, no. 1, pp. 76–84, Jan./Feb. 2017, doi: 10.1109/TCBB.2015.2446475.

[36] D. Tian, J. Zhou, Z. Sheng, Y. Wang, and J. Ma, "From cellular attractor selection to adaptive signal control for traffic networks," *Sci. Rep.*, vol. 6, Feb. 2016, Art. no. 23048, doi: 10.1038/srep23048.

[37] R. Zhang, Y. Ma, F. You, T. Peng, Z. He, and K. Li, "Exploring to direct the reaction pathway for hydrogenation of levulinic acid into γ-valerolactone for future clean-energy vehicles over a magnetic Cu-Ni catalyst," *Int. J. Hydrogen Energy*, vol. 42, no. 40, pp. 25185–25194, Oct. 2017, doi: 10.1016/j.ijhydene.2017.08.121.

[38] A. Zambelli. (2016). "A data-driven approach to estimating the number of clusters in hierarchical clustering." [Online]. Available: https://arxiv.org/abs/1608.04700

[39] S. Coogan, C. Flores, and P. Varaiya, "Traffic predictive control from low-rank structure," *Transp. Res. B, Methodol.*, vol. 97, pp. 1–22, Mar. 2017, doi: 10.1016/j.trb.2016.11.013.

[40] J. Lee, J. Kim, and B. B. Park, "A genetic algorithm-based procedure for determining optimal time-of-day break points for coordinated actuated traffic signal systems," *KSCE J. Civil Eng.*, vol. 15, no. 1, pp. 197–203, Jan. 2011, doi: 10.1007/s12205-011-1093-0.

[41] L. Liang, W. Wang, Y. Jia, and S. Fu, "A cluster-based energy-efficient resource management scheme for ultra-dense networks," *IEEE Access*, vol. 4, pp. 6823–6832, Sep. 2016, doi: 10.1109/ACCESS.2016.2614517.

[42] Z. Wang, "Determining the clustering centers by slope difference distribution," *IEEE Access*, vol. 5, pp. 10995–11002, Jun. 2017, doi: 10.1109/ACCESS.2017.2715861.

[43] J. Lu and Q. Zhu, "An effective algorithm based on density clustering framework," *IEEE Access*, vol. 5, pp. 4991–5000, Apr. 2017, doi: 10.1109/ACCESS.2017.2688477.

[44] W. Yang, K. Hou, B. Liu, F. Yu, and L. Lin, "Two-stage clustering technique based on the neighboring union histogram for hyperspectral remote sensing images," *IEEE Access*, vol. 5, pp. 5640–5647, Apr. 2017, doi: 10.1109/ACCESS.2017.2695616.

[45] R.-H. Zhang, Z.-C. He, H.-W. Wang, F. You, and K.-N. Li, "Study on self-tuning tyre friction control for developing main-servo loop integrated chassis control system," *IEEE Access*, vol. 5, pp. 6649–6660, Feb. 2017, doi: 10.1109/ACCESS.2017.2669263.

[46] Y. Zhao and G. Karypis, "Criterion functions for clustering on high-dimensional data," in *Grouping Multidimensional Data*, J. Kogan, C. Nicholas, and M. Teboulle, Eds. Berlin, Germany: Springer, 2006, doi: 10.1007/3-540-28349-8.

[47] A. Gholami and Z. Tian, "Using stop bar detector information to determine turning movement proportions in shared lanes," *J. Adv. Transp.*, vol. 50, no. 5, pp. 802–817, Aug. 2016, doi: 10.1002/atr.1376.

[48] D. Ma, X. Luo, S. Jin, D. Wang, W. Guo, and F. Wang, "Lane-based saturation degree estimation for signalized intersections using travel time data," *IEEE Intell. Transp. Syst. Mag.*, vol. 9, no. 3, pp. 136–148, Jul. 2017, doi: 10.1109/MITS.2017.2709881.

[49] *VISSIM 5.30-05 User Manual*, 1st ed., PTV, Karlsruhe, Germany, 2011.

[50] Y. Zhao, H. Gao, S. Wang, and F.-Y. Wang, "A novel approach for traffic signal control: A recommendation perspective," *IEEE Intell. Transp. Syst. Mag.*, vol. 9, no. 3, pp. 127–135, Jul. 2017, doi: 10.1109/MITS.2017.2709779.

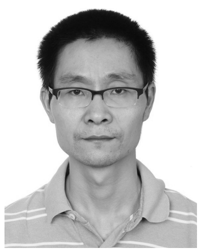
[51] G. Sharon *et al.*, "Network-wide adaptive tolling for connected and automated vehicles," *Transp. Res. C, Emerg. Technol.*, vol. 84, no. 11, pp. 142–157, Nov. 2017, doi: 10.1016/j.trc.2017.08.019.

[52] G. R. Iordanidou, I. Papamichail, C. Roncoli, and M. Papageorgiou, "Feedback-based integrated motorway traffic flow control with delay balancing," *IEEE Trans. Intell. Transp. Syst.*, vol. 8, no. 9, pp. 2319–2329, Sep. 2017, doi: 10.1109/TITS.2016.2636302.

XIANG SONG received the M.S. degree in transportation and the M.S. degree in management research from the Massachusetts Institute of Technology, USA, in 2013 and 2016, respectively, where he is currently pursuing the Ph.D. degree in transportation. His current research interests include personalized transportation through choice-based optimization, Bayesian econometrics, and optimal learning theory and application.

WENJING LI received the B.E. degree in Internet of Things from the Jiangxi University of Finance and Economics, Nanchang, China, in 2016. She is currently pursuing the master's degree with Zhejiang University. Her current research is data analysis in traffic and marine.

DONGFANG MA received the M.S. and Ph.D. degrees in traffic information engineering and control from Jilin University, Changchun, China, in 2009 and 2012, respectively. From 2012 to 2014, he was a Post-Doctoral Fellow with the Department of Civil Engineering and Architecture, Zhejiang University. Since 2015, he has been with the Institute of Marine Sensing and Networking, Zhejiang University. His current research interests include intelligent transportation systems, traffic planning and management, and big data mining.



YEZHOU WU received the M.S. degree in mathematics from the University of Science and Technology in China, Hefei, China, in 2006, and the Ph.D. degree in mathematics from West Virginia University, USA, in 2012. From 2012 to 2014, he was a Post-Doctoral Fellow with the Department of Mathematics, The University of Hong Kong. Since 2014, he has been with the Institute of Marine Information Science and Technology, Zhejiang University. His current research interests include combinatorics and graph theory, network topology structure, and computer algorithms.

DAXIONG JI received the Ph.D. degree in pattern recognition and intelligent system from the Graduate University of Chinese Academy of Sciences, Beijing, China, in 2008. He was an Associate Researcher with the State Key Laboratory of Robotics, Shenyang Institute of Automation, Chinese Academy of Sciences, China. He has been an Associate Professor with the Institute of Marine Robot, Ocean College, Zhejiang University, China, since 2015. His research interests include robot control theory, marine robotics, and artificial intelligence.