ISSN 2616-5767 Vol.7, Issue 4: 122-128, DOI: 10.25236/AJETS.2024.070418

A study on the problem of predicting traffic signal cycles by travelling trajectories

Zhengen Lv[1,a], Qi Liu[2,b], Yicheng Shen[3,c], Pengyang Wei[4,d], Zhicheng Pan[5,e,*]

[1]*College of Mechanical and Control Engineering, Guilin University of Technology, Guilin, China*
[2]*School of Energy and Electricity Engineering, Qinghai University, Sining, China*
[3]*College of Mechanical and Control Engineering, Guilin University of Technology, Guilin, China*
[4]*School of Biomedical Engineering and Technology, Tianjin Medical University, Tianjin, China*
[5]*School of Petroleum and Natural Gas Engineering, Changzhou University, Changzhou, China*
[a]*2542558270@qq.com,* [b]*2675668147@qq.com,* [c]*3523242189@qq.com,* [d]*3118963479@qq.com,*
[e]*pzzzc@qq.com*
[*]*Corresponding author*

Abstract: *With the acceleration of urbanisation, the problem of traffic congestion is becoming more and more prominent, and effective traffic management and optimisation is increasingly becoming an important issue in urban operation. In this paper, we will focus on the relevant data and discuss in depth how to use vehicle trajectories to estimate the operation cycle of traffic signals. In this paper, the K-means model is first used to analyse the periodicity of five specified intersections in a particular direction based on the vehicle trajectories within one hour and accurately calculate the periodicity of traffic signals at these intersections in that direction. The clustering model is continuously adjusted through iterative optimisation to minimise intra-class distances and maximise inter-class distances. Then, the peak detection method is used to solve the red light phase duration and calculate the continuous stop state interval to estimate the green light duration. In this paper, for dynamic cycle change detection, peak analysis is used to identify the major stop duration peaks and the CUSUM method is further applied to detect cycle change points. This helps in mapping the vehicle trajectory using the provided data and confirming the approximate orientation of the vehicle as a means of categorising the data and constructing an accurate cycle model.*

Keywords: *Cluster analysis, Dynamic cycle detection, Data visualisation, CUSUM, Peak analysis*

1. Introduction

As urban traffic congestion becomes more and more prominent, effective management and optimisation of signal cycles have become the key to improving traffic flow and easing congestion. Intelligent traffic management systems, which have attracted much attention in the construction of smart cities, use big data and cloud computing technology to process driving trajectory data, improve road network utilisation efficiency, shorten commuting time and reduce tailpipe emissions. By optimising the signal cycle, it can reduce drivers' waiting time and the frequency of starting and stopping, thus improving driving safety and efficiency.

Some scholars have already studied related models. Wu et al [1] propose a two-level fuzzy control method to design the signal cycle of road sections according to the actual traffic conditions for the case of multi-intersection road sections and the existence of multiple main roads. The experimental simulation results show that the method proposed in this paper performs well. Luo et al [2] provide a visual analytics system containing multiple linked views that allow users to observe and compare vehicle trajectories and trajectory entropy, combined with cluster analysis and correlated interactions, to help users discover meaningful vehicle behaviours. Lv [3] combined various anomaly recognition models to denoise the trajectory data of transport vehicles in order to obtain more accurate trajectory routes of transport vehicles, and secondly, used various clustering models to screen the driving behaviour of drivers. Wu et al [4] studied the cumulative sum of change points (CUSUM, for short) estimator. Simulation studies and two real data examples are also provided to support the theoretical results. Ivanovs et al [5] introduce a new Lévy fluctuation theoretic method to analyze the cumulative sum (CUSUM) procedure in sequential change-point detection. Ahad et al [6] propose a modified version of CUSUM that we call data-adaptive symmetric CUSUM (DAS-CUSUM). Experiments on simulated and real-world data show the utility of

ISSN 2616-5767 Vol.7, Issue 4: 122-128, DOI: 10.25236/AJETS.2024.070418

DAS-CUSUM.

In this paper, we first implement the classification of roads based on K-means cluster analysis and use peak detection to calculate the corresponding durations of traffic lights. Then CUSUM is used to identify the critical points of cyclic changes. After detecting the dynamic cyclic changes in traffic signals, this paper uses the given vehicle related data to plot the trajectory and performs the cyclic modelling work to reveal the dynamic change pattern under traffic signal control. In order to solve the model established in this paper, the http://hzbmmc.com:9000/hzb/jeditor/ B%E9%A2%98%EF%BC%9A%E4%BD%BF%E7%94%A8%E8%A1%8C%E8%BD%A6%E8%BD%A8%E8%BF%B9%E4%BC%B0% E8%AE%A1%E4%BA%A4%E9%80% 9A%E4%BF%A1%E5%8F%B7%E7%81%AF%E5%91%A8%E6%9C%9F%E9%97%AE%E9%A2%98_1713424222679.zip the data provided.

2. Signal Fixation Period Estimation based on K-means Cluster Analysis

2.1. Data Preprocessing

In this paper, we first preprocess the data provided by the problem. First, we plot scatter plots to roughly analyse the distribution of the data. Next, we write code to draw Quantile-Quantile plots to test whether the data are normally distributed. In this process, we use the 3σ principle to judge the outliers and analyse the results in detail in context.

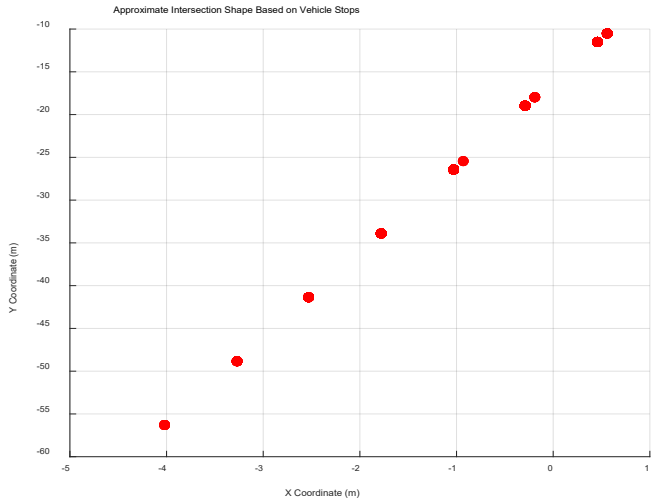

Figure 1: Scatter Plot of data set

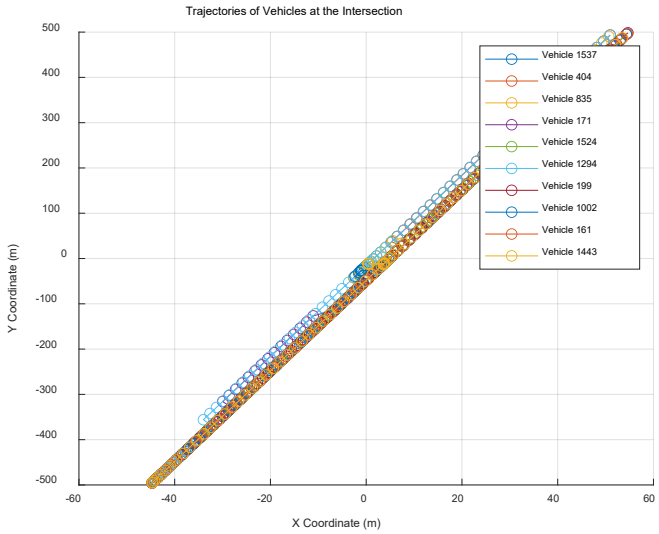

Figure 2: Quantile-Quantile Plot

ISSN 2616-5767 Vol.7, Issue 4: 122-128, DOI: 10.25236/AJETS.2024.070418

The 3σ principle is used here to identify outliers. The 3σ principle, also known as the 3σ rule, can be succinctly summarised as follows: for data that conforms to a normal distribution, any value that deviates from the standard deviation of the mean (σ) by more than a factor of three is considered an extreme outlier.

As shown in Figure 1 and Figure 2, we did not find outliers in this dataset, obeying a normal distribution.

2.2. K-means Cluster Analysis

In this paper, we choose the hierarchical clustering algorithm which is highly similar to it. Hierarchical clustering algorithms are those that begin by treating each data point as a single cluster, and then merge (or aggregate) the classes sequentially until all the classes are merged into a single cluster containing all the data points.

2.2.1. Cluster modelling

K-means is one of the most commonly used methods in clustering and calculates the best category attribution based on the similarity of the distance between points.

The K-means algorithm clusters the data by trying to separate the samples into several groups of equal variance to minimise the objective function. The algorithm requires the number of clusters to be specified. It scales better to large numbers of samples and is currently used in a wide range of applications in a large number of fields.

The idea of this algorithm is roughly as follows: first randomly select k samples from the sample set as cluster centres and calculate the distances of all samples to these k "cluster centres", for each sample, divide it into clusters with the closest "cluster centre", for the For each sample, it is divided into the cluster with the closest "cluster centre", and for each new cluster, a new "cluster centre" is calculated for each cluster.

Based on the above description, the K-means algorithm proceeds as follows.

Step 1: Selection of the number of clusters k and the initial cluster centre for each cluster;

Step 2: Distance from each sample point to the "cluster centre";

Step 3: Assign the samples to the nearest neighbour clusters according to the minimum distance principle and update the cluster centres according to the newly divided clusters;

Step 4: Repeat the previous step until the centre of clustering no longer changes;

Step 5: Output the final clustering centre and k cluster divisions.

2.2.2. K-means Model solving

The problem also involves trajectory distances and this paper therefore also introduces the Euclidean distance method.

Euclidean distance (ECD) is one of the most commonly used distance measures for calculating the straight line distance between two points in a multidimensional space. In two or three dimensions, it can be considered as the actual distance between two points, similar to the distance we measure using a straightedge.

That is, the difference between the corresponding coordinate values on each dimension is squared and then summed, and then the Euclidean distance is obtained by performing a square root operation on that sum. Using the shortest distance method, clustering is performed.

We identify five possible lane locations by clustering. The derived road conditions are as follows.

Table 1: Lane Road Condition Table

A1	A2	A3	A4	A5
0 Increasing	0 Increasing	0 Increasing	0 Increasing	0 Increasing
1 Increasing	1 Decreasing	1 Increasing	1 Increasing	1 Increasing
	2 Decreasing	2 Increasing	2 Decreasing	2 Increasing
	3 Increasing	3 Increasing	3 Decreasing	3 Increasing
	4 Increasing			4 Increasing
				5 Increasing

ISSN 2616-5767 Vol.7, Issue 4: 122-128, DOI: 10.25236/AJETS.2024.070418

2.3. Estimation of Fixed-cycle Signal Cycles

In this paper, peak analysis is used to estimate the period of fixed-period signals.

2.3.1. Event Analysis

In order to track vehicle stop (start) and start (stop) state transition events, it is necessary to monitor successive changes in the stop state (represented by the 'stopped' column in the dataset) data column for each vehicle. When the vehicle state changes from non-stop to stopped, i.e., the first record in the timing data is a travelling state and the immediately following record is a stationary state, the event can be marked as a stop event (stop_change = 1). Conversely, when the vehicle state changes from stopped to non-stop, i.e. from stationary to travelling, we mark this as a start event (stop_change = -1).

In order to ensure that the processed data is in the correct order, and thus to accurately identify and track these events, the dataset must first be sorted according to the vehicle identification number (vehicle_id) and timestamp (time). This step is crucial as it ensures accuracy in the subsequent analysis regarding temporality and individual vehicle behaviour. Sorting the data according to the aforementioned metrics ensures that vehicle-by-vehicle analyses of the variability of their driving status on the timeline are performed.

2.3.2. Parking Duration Measurement and Statistics

In order to accurately quantify and analyse the stopping and starting characteristics of vehicles, the following standardised process has been adopted:

①Extraction of event timestamps:

From the dataset sorted by vehicle_id and timestamp (time) sequences, all time points indicating that a vehicle has been stopped (stop_change = 1) and started (stop_change = -1) are independently extracted.

②Stop event duration determination: operate according to a reliable procedure and perform the following steps:

Iteratively traverse the set of identified stop events;

For each particular stop event, find the closest start event with a timestamp greater than the stop event in the same vehicle ID;

Determine the duration of a single stop event by calculating and storing the time interval between the stop and the detected start event.

③Stop duration data generalisation:

The stop event duration data obtained above was compiled and converted into Series objects in the pandas framework, an operation that facilitates the execution of subsequent statistical processing and data analysis.

④Descriptive Statistics Extraction:

Descriptive statistical metrics for parking duration data, including but not limited to totals, means, standard deviations, minimums, and maximums, were calculated and presented to provide basic data characterisation perceptions for the study.

⑤Peak analysis to estimate red light durations:

Peak detection of duration data using the find_peaks method with a reasonable minimum height threshold (e.g. 31 seconds in this case) was aimed at identifying significant peaks that could be interpreted as traffic signal red light durations.

Similarly, descriptive statistics of these peak data as red light durations were extracted.

2.3.3. Peak Analysis Model solving

This paper plots the trajectories of 15 vehicles travelling at an intersection. Figure 3 shows that the vehicles slow down near the intersection and then accelerate through it. Some vehicles stop before the intersection and then continue to drive.

ISSN 2616-5767 Vol.7, Issue 4: 122-128, DOI: 10.25236/AJETS.2024.070418

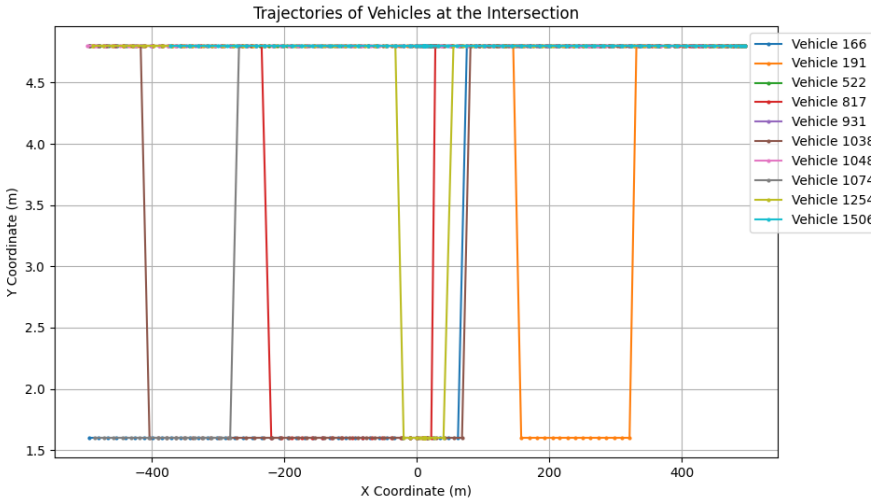

Figure 3: Vehicle Trajectory Peak Analysis

For example, the results of the analysis using the A1 vehicle as a reference showed that a total of 87 valid stopping duration events were recorded. Descriptive statistics indicate a mean parking duration of approximately 31.5 with a standard deviation of 22.3 seconds, showing a high degree of dispersion in the data. The shortest recorded stop duration was 1 second, while 25 per cent of stop events lasted 10.5 seconds or less, 50 per cent (i.e., the median) had a stop duration of 28 seconds, 75 per cent of the data was less than or equal to 50 seconds, and the longest duration was 69 seconds.

The distribution of vehicle stopping durations exhibits significant non-uniformity, showing several significant peaks in stopping frequency. In particular, the frequency of parking events is higher in the time intervals of 0 to 10 seconds, 20 to 30 seconds and 50 to 60 seconds. According to the relevant literature, the minimum waiting time for traffic signals varies between cities, e.g., the average signal waiting time is 12.5 seconds in Weihai, Shandong, 14.5 seconds in Daqing, Heilongjiang, and 15.6 seconds in Tongchuan, Shaanxi. In order to accurately analyse the parking duration data, the shorter parking duration data were screened and deleted. Peak analysis was used to set a threshold higher than the average stopping duration, and the average stopping duration was used as a height threshold so as to exclude very small value data. After performing these data cleaning steps, more accurate and representative results of the parking duration distribution were obtained.

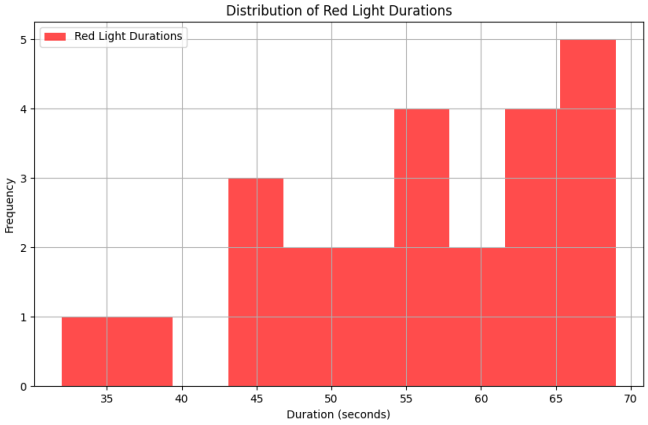

Figure 4: Duration of Red Light after Treatment

As is shown in Figure 4, the average signal red duration obtained was estimated to be 56.0 seconds. In the absence of direct signal state change observations, an extrapolation method based on indirect observations was used to estimate the duration of the green and red phases of the signal control system. The logical assumption of this extrapolation method is based on the premise that the vehicle will be forced to stop during the red phase and resume travelling when the signal switches to the green phase.

ISSN 2616-5767 Vol.7, Issue 4: 122-128, DOI: 10.25236/AJETS.2024.070418

Solve for the green light time in the same way. The specific results are shown in Figure 5.

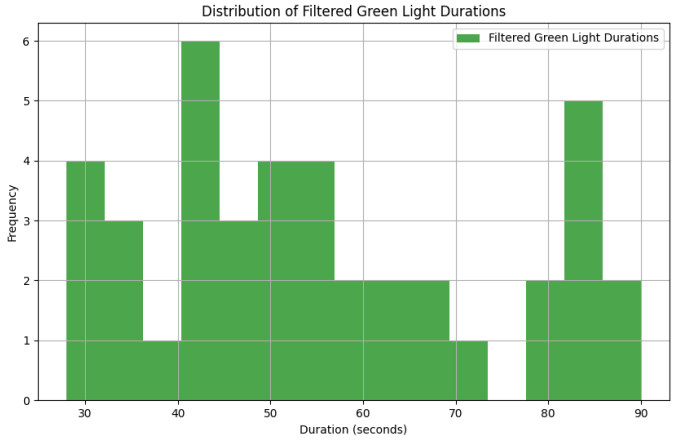

Figure 5: Distribution of Green Light Hours

Figure 5 shows the distribution of green light durations. The green light durations are normally distributed between 30 and 90 seconds. The longest green light durations are between 40 and 50 seconds and the shortest are between 70 and 80 seconds.

3. CUSUM-based identification of alternating old and new cycles

3.1. Prediction of Effective Continuous Stopping Time

The majority of vehicles had stopping durations between 0 and 10 seconds, followed by between 10 and 20 seconds. Only a small number of vehicles had parking durations of more than 20 seconds. This suggests that the car park has a high level of utilisation of parking spaces, although there are also some spaces that are unused for longer periods of time.

3.2. CUSUM Statistical Analysis

The CUSUM (cumulative and control chart) method is an effective technique widely used in the field of statistical quality control to monitor and identify change points in a process or system. The method is particularly suitable for analysing traffic data flows, especially when there is a need to identify changes in traffic signal cycles or other changes in the nature of the system.

3.2.1. CUSUM solving

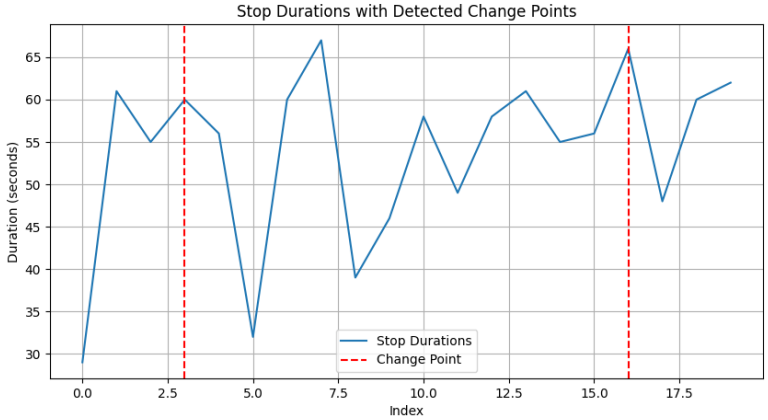

Figure 6: Plot of the results of the cycle analysis based on CUSUM

The graph shows the variation of parking duration over time. Parking duration is the amount of time a vehicle spends in a car park. As can be seen in Figure 6, parking duration is constantly changing

ISSN 2616-5767 Vol.7, Issue 4: 122-128, DOI: 10.25236/AJETS.2024.070418

throughout the day. During peak hours, parking durations are longer, while during low peak hours, parking durations are shorter. In addition, the graph shows whether there are other vehicles in the car park. When there are other vehicles in the car park, the parking duration is longer. This may be due to the fact that when there are other vehicles in the car park, drivers need to wait for the other vehicles to leave before they can park.

4. Conclusions

In this paper, the given data are first pre-processed to determine whether there are outliers and missing values. K-means clustering analysis is performed on the "Y-axis data" to achieve the classification of roads, and the direction is determined by the displacement change of the "X-axis data". On this basis, the speed of each data point is calculated to determine the vehicle status. Subsequently, the stop and start time points of all vehicles at each intersection are collected, and the corresponding duration of the traffic light is calculated. Summarising the red and green light durations of the signals at each intersection, the red and green light durations (in seconds) at intersections A1, A2, A3, A4, and A5 are 56.0, 44.7, 57.1, 46.6, 51.2; 55.8, 81.8, 55.0, 45.4, and 61.0 respectively. In order to detect the dynamic cycle changes of the traffic signals, this paper calculates the effective vehicle stopping durations. Using peak analysis technique, this paper identifies the major peaks in the stopping duration data. Subsequently, CUSUM is used to identify the critical points of cyclic variations. The switching moments (in seconds) of the traffic light cycles at intersections C1~C6 are obtained as 226,839; 246,1770; 164,546,593; 87,118; 1661,2073; and 96,417, respectively.

In the future, considering the synergistic effect of multi-directional traffic flows on each other, the data will be classified and processed, and relying on the relevant models in this paper, we will carry out periodic modelling work to reveal the dynamic change law under traffic signal control.

References

[1] Q. Z. Wu, J. Zhou. Research on the signal cycle of traffic signals in multi-intersection road sections based on fuzzy control[J]. Industrial Control Computer, 2019, Vol. 32(6): 81-82, 85

[2] Y. T. Luo, T. Wang, M.N. Yang, Y.K. Zhang. A visual analysis method for vehicle behaviour based on historical driving trajectory set [J]. Computer Science, 2021, Vol. 48 (9): 86-94

[3] R. Lv. Research on trajectory and driving behaviour of transport vehicles based on clustering [D]. Chang'an University, 2022

[4] Wu, Yi; Wang, Wei; Wang, Xuejun. Convergence of the CUSUM estimation for a mean shift in linear processes with random coefficients [J]. Computational Statistics, 2024: 1-26

[5] Jevgenijs Ivanovs; Kazutoshi Yamazaki. A series expansion formula of the scale matrix with applications in CUSUM analysis [J]. Stochastic Processes and their Applications, 2024, Vol.170: 104300

[6] Ahad, Nauman; Davenport, Mark A.; Xie, Yao. Data-adaptive symmetric CUSUM for sequential change detection [J]. Sequential Analysis-Design Methods and Applications, 2024, Vol. 43(1): 1-27