

# The Promotion Effect of LOESS Smoothing Technique in Short-term Traffic Volume Clustering

Jia-juan CHEN<sup>1</sup>, Yu-bang LIU<sup>1</sup>, Zhi-yuan WANG<sup>1,\*</sup> and Chuan-tao WANG<sup>2</sup>

<sup>1</sup>School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China

<sup>2</sup>School of Mechanical-Electronic and Vehicle Engineering, Beijing University of Civil Engineering and Architecture, Beijing 100044, China

\*Corresponding author

**Keywords:** Short-term traffic volume, Clustering, LOESS smoothing technique, Intelligent transport system.

**Abstract.** In short-term traffic volume clustering, one important issue is the representation of traffic profiles. This article focuses on how the LOESS smoothing technique enhances the clustering effect and what the best value of parameter span of LOESS is. This article used K-Means clustering algorithm and compared the clustering effect using raw data and smoothed data. The experiment result shows when the traffic profiles are slight smoothed, the clustering effect enhances from 39.15% to 66.48%. And the best range of parameter span is 0.2~0.4 to keep the balance of clustering effect and profiles details. This article verifies the promotion effect of LOESS smoothing technique in short-term traffic volume clustering and gives advice on the best value of LOESS parameter.

## Introduction

With the rapid development of information techniques such as Internet of things, big data, cloud computing, Intelligent Transport System (ITS) is making great progress not only in laboratories but also in real traffic management systems in China. Short term traffic analysis is the essential component of ITS to identify specific traffic patterns and make forecasting of short-term traffic flow.

Short term traffic clustering, which can identify similar traffic flow of different sections of road on different day, is enlightening to find similar traffic patterns and support traffic management. Some researchers used clustering to support data driven forecasting methods such as K Nearest Neighbor regression and used locally estimated scatterplot smoothing (LOESS) to smooth the noise in traffic flow profiles. However, the promotion effect of LOESS to identify similar traffic profiles was rarely been focused on. This research aims to fill up this gap to find whether LOESS can enhance the clustering effect of short term traffic profiles and what the best parameter of LOESS is to keep the balance of maintaining the profiles details and clustering effect.

Short term traffic profile is one type of time series. Aghabozorgi (2015) [1] has reviewed the four main elements of time series clustering, including representation method, similarity degree, prototype definition, clustering algorithm. This article uses Euclidean distance as similarity measurement, average value as prototype definition and K-Means as clustering algorithm. Raw profiles and LOESS smoothing profiles as two kinds of representation methods will be conducted in clustering experiments and the clustering effects will be compared.

## Methodology

### K-Means Clustering Methods

K-Means is one of the most popular clustering algorithms based on partition. K-Means produces k clusters from n unlabeled objects to make sure that there is at least one object in each cluster [1]. The clustering prototype of K-Means is the average of the objects. The principle of K-Means is to minimize the sum of the distances (generally expressed in Euclidean distance) between all objects in the cluster and their cluster center (i.e. prototype). So this article chooses the result of between-cluster sum of squares by total squared distance as an index of clustering index. For the data in Euclidean

space, total sum of the squared error (SSE) is calculated as Eq. 1, which  $x$  is each sample point and  $c_i$  is the center of cluster  $i$ .

$$SSE = \sum_{i=1}^k \sum_{x \in C_i} dist(x, c_i). \tag{1}$$

K-Means clustering algorithms steps are as follows. First, select K centers randomly. Secondly, going through all profiles, dividing each profiles into the nearest centers. Third, calculate the mean value of each cluster and act as a cluster centroid. Forth, repeat steps 2 and 3 until the K midline points no longer changes, or the specified number of iterations is reached.

**LOESS Smoothing Technique**

Locally estimated scatterplot smoothing or locally weighted regression (LOESS) is to estimate the regression surface by a multivariate smoothing process, fitting the function of the independent variables locally and moving in a similar way to how to calculate the moving average of a time series [2,3]. It's a nonparametric regression technique with highly flexible which can capture the pattern of the data without making any assumption on the nature of the raw data. Many previous related researches used LOESS to smooth the noise in traffic flow analysis [4].

The fitting process is completed locally. In other words, at point  $x$ , the points used for fitting are points near  $x$ , and the weights are determined by their distance from  $x$ . The range of the neighborhood is adjusted by the parameter  $\alpha$  (or span). If  $\alpha < 1$ , then the neighborhood will include proportion  $\alpha$  of the points, which have tricubic weighting. All points will be included, if  $\alpha > 1$ . The weight of profile  $x$  when  $\alpha < 1$  is calculated as Eq. 2.

$$W(x) = \left(1 - \left(\frac{dist_x}{\max dist}\right)^3\right)^3. \tag{2}$$

**Experiments**

This article uses the traffic volume data of one traffic survey station in Guizhou province of China from September 19<sup>th</sup> to October 12<sup>th</sup> 2016. The traffic volume aggregates every 5 minutes so there are 288 records every day if no loss in detecting and aggregating.

First, the raw traffic volume profiles are introduced in K-Means clustering. The clustering index with different parameter K are as Fig. 1 (a) and the clustering effect of traffic profiles when K is 2 as Fig. 1 (b). From Fig. 1 (a), when K is 2, the clustering index is only 39.15%. From Fig. 1 (b), the centers of clusters also contains noise.

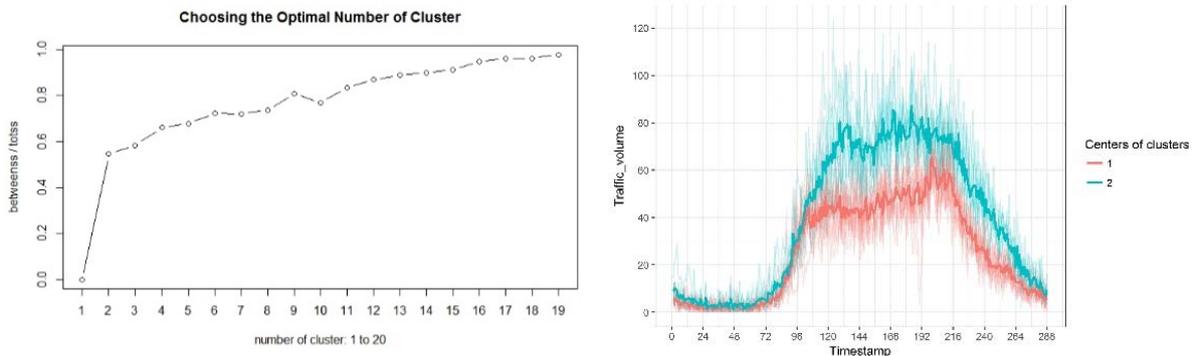


Figure 1. (a) Clustering index with different value of K. (b) The clustering effect when K is 2.

Second, the LOESS smoothed profiles are introduced in K-Means clustering. The clustering indexes with different values of parameter span of LOESS are as Table 1. When span is 0.1, which means the traffic profiles are slightly smoothed, the clustering index is 66.48%. The clustering index is larger when parameter span is larger, which means the clustering effect is better.

Table 1. Clustering indexes with different values of parameter span of LOESS

Span	0.1	0.2	0.3	0.4	0.5
Clustering index	66.48%	71.19%	73.71%	75.70%	76.92%
span	0.6	0.7	0.8	0.9	1.0
Clustering index	78.48%	79.90%	81.01%	81.48%	81.71%

However, when span is larger than 0.4, many details in traffic profiles has been smoothed as Fig.2 and some interesting traffic patterns may be ignored. So the best range of parameter span of LOESS is 0.2~0.4 in the context of traffic volume profiles clustering.

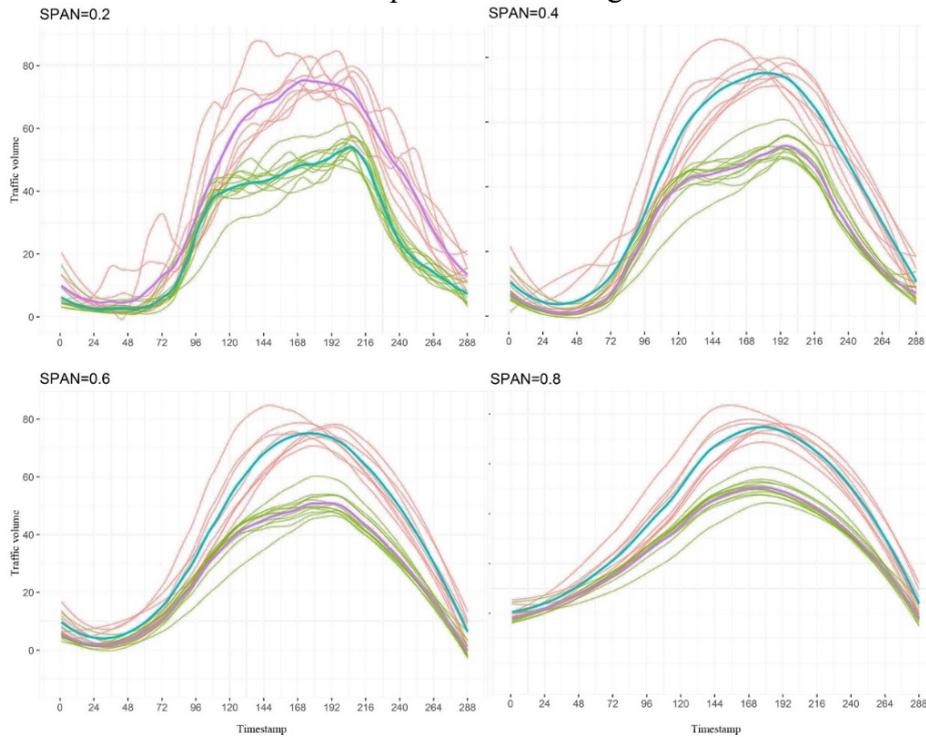


Figure 2. Smoothing effect with different value of span

## Conclusions

In short term traffic clustering, the representation method of traffic profiles is one of the most important issues. When raw traffic profiles are used in clustering, the clustering index (the result of between-cluster sum of squares by total squared distance) is only 39.15%. Smoothing profiles is a better way of representation. This article uses LOESS smoothing technique to verify this conclusion. When span is 0.1, which means traffic profiles are slightly smoothed, the clustering index enhances to 66.48%. The larger of parameter span, the better of clustering effect. However, when span is larger than 0.4, many details of traffic profiles are smoothed. It may lose the insight of interesting traffic patterns. So the best range of parameter span is 0.2~0.4 in traffic volume profiles clustering.

## Acknowledgement

This research was financially supported by the National Key R&D Program of China No. 2018YFB1601603.

## References

- [1] Aghabozorgi S, Shirkhorshidi A S, Wah T Y, Time Series Clustering - A Decade Review, J, Clustering, 2015, 53(May): 16–38.
- [2] Cleveland W S, Devlin S J, Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting, J, Journal of the American Statistical Association, 1988, 83(403):596-610.

- [3] Cleveland W S. Robust Locally Weighted Regression and Smoothing Scatterplots, J, Journal of the American Statistical Association, 1979, 74(368):829-836.
- [4] Habtemichael F G, Cetin M, Short-term traffic flow rate forecasting based on identifying similar traffic patterns, J, Transportation Research Part C, 2016, 66:61-78.