Cluster Analysis for a Series of Microscopic Traffic Simulation Results

Mikoto Yanai¹, Kazuki Abe¹, Tomonori Yamada⁰*, Hideki Fujii¹, Shinobu Yoshimura¹

¹Department of Systems Innovation, School of Engineering, The University of Tokyo
⁰tyamada@sys.t.u-tokyo.ac.jp

Received: March 14, 2017; Accepted: September 27, 2017; Published: March 31, 2018

Abstract. The use of traffic simulators is getting increasingly popular for the assessment of policies to reduce traffic jams. However, simulators based on multi-agent models show some variability in results even if the input data and parameters are identical, because they use probabilistic phenomena, such as lane change of vehicles, which is determined by random numbers. Results of such simulations have been evaluated and analyzed by taking the mean of several trials, but such an approach fails to account for phenomena that have a low probability of occurring, but are still possible nonetheless, and therefore appropriate decisions may not be made. This paper verifies that possible phenomena can be taken into account by the cluster analysis combing a self-organizing map (SOM) and hierarchical clustering. This study clustered traffic volume data obtained from 600 traffic simulations near Okayama Station, grouped the traffic patterns, and analyzed the results.

Keywords: Traffic simulation, Cluster analysis, Self-organizing map

1. Introduction

1.1. Theoretical approach and limits of its application

Traffic problems have a serious impact on society, and a wide range of policies have been proposed and implemented over the years to solve this problem. When traffic infrastructure did not meet the traffic demand once, the impact of a new road could be measured by static traffic volume allocation and policies can be assessed using static indicators.

However, a more diverse assessment of urban traffic policies has been needed in recent years, as the conventional theoretical approach has been getting insufficient. For example, existing indicators, such as the degree of congestion and the degree of saturation, are already at their applicable limits. Degree of congestion is an indicator that measures the traffic
conditions in road sections, defined as the ratio of traffic capacity and traffic demand [1]. Degree of saturation is an indicator that measures the traffic conditions at a signalized intersection, and is defined as the ratio of traffic capacity to traffic demand at an intersection [2]. Both indicators depend on traffic demand which is given as a mean of the range segmented by a given time period. They are unable to appropriately assess traffic conditions when the time variation in traffic demand is large. Degree of saturation of an intersection is an indicator that considers a single intersection, and it carries another problem in that it is unable to take the spatial impact of multiple intersections into account.

1.2. Simulation-based approach

Traffic simulation is used to deal with the limitations of the static traffic demand allocation method. In particular, microscopic simulation has three following advantages [3].

The first advantage is its ability to simulate and predict traffic jams. Traffic jams, in which multiple intersections affect each other, and ones which are caused by exceeded demand in a short period of time can be simulated, while they cannot be evaluated by conventional methods.

The second advantage is the ability to reproduce the behavior of the vehicles. Behaviors of individual vehicles such as acceleration, turns and lane changes can be considered in detail. In addition, special behaviors can also be simulated, such as the behavior of a right-turn vehicle at an intersection and the behaviors of buses.

The third advantage is the ease of visualization. Because traffic phenomena can be visualized as a model on a computer, the result is an intuitive understanding of dynamic changes in traffic conditions.

While the introduction of simulations can overcome some limits of conventional indicators, traffic phenomena are fundamentally probabilistic. It is impossible to represent every detail of vehicle behavior in a deterministic manner, and thus, lane change of vehicles, for example, is expressed as probabilistic phenomena using random numbers. Therefore, the outputs of simulators vary in each trial even if the input data and parameters are identical. Because simulations aim to recreate reality as an average, accurate interpretation of simulation results requires an interpretation that takes the range of the variance of random numbers into account rather than merely interpreting the aggregate values.

1.3. Prior research

Conventional evaluation and analysis of traffic simulations were based on mean values from several trials. Simulation trials are typically run five times [3], but this number of trials is empirical. Therefore, a study was conducted to quantify the uncertainty of the number of trials as an indicator of reliability [4].

Several studies have applied the clustering method as an analytical approach to the results
of traffic simulations that do not depend on a mean value. Celikoglu (2014) clustered the results from macroscopic traffic simulations based on a cell transmission model by applying a neural network. As a result, Celikoglu successfully grouped congestions by clustering data on varying speed and density values obtained from the simulations. Silgu and Celikoglu (2015) clustered changes in dynamic traffic behavior from simulations by the fuzzy c-means method and grouped patterns in changes in the state of traffic flow [6]. Taking changes in dynamic traffic behaviors into account, clustering performances were evaluated by comparing dynamic and static treatment of the center point of the clusters, and it was found that the performance improved under a dynamic treatment.

Other studies have clustered observations of actual traffic rather than the results of simulations to help the understanding of the phenomena. For example, Wijermars and van Berkum (2005) performed a hierarchical clustering using Ward's method for chronological data of the traffic volume, and showed that the period of peak traffic volume varied by cluster [7]. Chen et al. (2014) extracted scenarios for simulations by clustering actual observed data and providing them as the input for a traffic flow simulator. These studies have shown the usefulness in clustering actual traffic phenomena. By assuming that the mechanism of traffic phenomena exists in the world of simulation as it does in the real world, clustering the results from simulations is expected to be as effective in determining traffic conditions as the results from the real world.

1.4. Objective of the study

Evaluating the results based on the mean of a few trials has the risk of missing unlikely but possible and important phenomena. On the other hand, possible phenomena can be taken into account by grouping traffic phenomena based on multiple indicators and accounting for the occurrence of multiple dependencies on various factors. As a result, the results of the simulations can be determined appropriately. There are various clustering methods and one of them is a method combing a self-organizing map (SOM) and hierarchical clustering. This method has many advantages and has been applied in various fields, but not in the traffic field. Therefore, this study verifies that the method can effectively cluster simulation results based on the presence of traffic jams. Traffic volume data from 600 traffic simulations in Okayama Station area were clustered and grouped into several traffic patterns. Grouped results were visualized and analyzed, and characteristics of the grouped results were discussed.

The city of Okayama, chosen as the model area for this study, is a typical Japanese regional city, which has been considering an extension of its tramline in recent years. The urban environment of the city is also changing significantly through developments such as the opening of a new large-scale shopping center in front of the train station in 2014. The city has also been making progress on traffic policies such as the assessment of traffic jams and improvement of public transportation service [9,10,11].
2. Simulation settings

Traffic volume data from a past study [4] was used for this study. The past study used a multi-agent traffic and environment simulator ADVENTURE_Mates, in short MATES, to simulate traffic flow [12, 13], and the following section explains the simulation conditions for the development of data.

2.1. Multi-agent model

The multi-agent traffic and environment simulator, MATES, is a kind of microscopic simulators described in section 1.2. In the multi-agent model, individual vehicles that behave according to specified rules are modeled. As a result of interaction among agents and environment, various traffic phenomena such as traffic jams are reproduced.

2.2. Measurement of the traffic volume

The number of passing vehicles was measured in the simulation at 14 intersections on a road network in a 2.5 km grid near Okayama Station at which actual traffic volume surveys were conducted, which is discussed in the next section. Specifically, all 57 road cross sections that approach the 14 intersections were examined. Fig. 1 shows the network surrounding the Okayama Station and the 14 intersections at which traffic surveys were conducted.

Figure 1: A map of the area around Okayama Station and intersections surveyed for traffic volume
Given that vehicles are able to approach the intersection from four directions, the number of passing vehicles was measured for each direction over road cross sections a–d, as shown in Fig. 2. Intersection H was a junction of five roads, and measurements were taken in five directions.

An ID was assigned to each point of measurement in a counter-clockwise manner starting at the road cross section on the north side for each intersection. The ID was assigned incrementally from the road cross section a in intersection A to the road cross section d in intersection N in order from 1 to 57.

2.3. Occurrence of vehicles

An origin-destination (OD) matrix was used to describe the occurrence of vehicles. An OD matrix defines a pair with a starting point and a destination, and the traffic volume that travels through the pair (OD traffic volume). In MATES, vehicles are generated based on the OD traffic volume by treating the matrix as the input data. An OD matrix was developed based on the results of traffic census conducted between 15:00 and 16:00 on Sunday, April 22, 2012 in Okayama City [14]. This census was conducted for the introduction of a large-scale shopping center. The census was conducted at this time of day as traffic volume was at its peak on Sundays during this time, and traffic jams occurred as a result. The OD matrix was manually adjusted to match the census results with the traffic volume for the individual roads.

Two cases were configured to take traffic jams (and lack thereof) into account. Case 1 assumed the number of generated vehicles as defined by OD traffic volume, and a traffic jam
was observed in the simulation. On the other hand, case 2 was configured as a case in which traffic jams did not occur, and the OD traffic volume was assumed to be half of the original volume. This case limited the number of vehicles in the simulation to half of the original OD matrix to prevent traffic jams from occurring during the simulation. The number of passing vehicles in simulation results were doubled in case 2, which was treated as data comparable with that of case 1. When the traffic jam does not occur, MATES has linearity between the number of generated vehicles and the number of passing vehicles. The number of generated vehicles followed a Poisson distribution, and the OD traffic volume was the mean value of the distribution.

2.4. Route search

All vehicles that ran across the map decided its route to be the shortest distance for a given OD pair, regardless of the congestion on the road in our simulations. However, vehicles rerouted and found a new shortest path to the specified destination if they could not make a turn or change lanes due to congestion.

2.5. Data extraction of simulation results

A 90-minutes simulation trial was conducted 300 times per case using the inflow from case 1 and case 2. Since vehicles did not exist in the simulated area at the beginning, the first 30 minutes after the start was discarded to avoid the transient state in which vehicles were flowing in. The simulation was evaluated using the number of vehicles that passed through the measurement points during the last 60 minutes. The results were treated as vectors by ordering the traffic volume at each point in ascending order of the ID. Therefore 600 vectors with 57 dimensions were analyzed.

A total of 300 sample data from case 1 and case 2 were combined and 600 sample data were clustered. More detailed clustering and analysis were then performed on 300 sample data from case 1.

3. Proposed method

Workflow for the proposed method consisted of the organization of simulation data with a SOM, clustering of the results, and an analysis of the clusters. At first, the simulation data is grouped into each node of a SOM. Then reference vectors corresponding to nodes are clustered. Finally, the clusters are analyzed. The advantages of this method are to maintain the nonlinear relationship, to be reproducible (not affected by random numbers), and to reduce calculation cost in clustering, compared with other clustering methods.

3.1. Self-organizing map
A SOM is a type of neural network capable of unsupervised learning. It is used to visualize data structures [15]. The network consisted of an input layer formed by input data and a mapped competitive layer, which mapped multi-dimensional data to arbitrary data. Nodes were placed throughout the competitive layer like a lattice, and each node was assigned a reference vector that had the same dimension as the input vector. Input vectors were grouped to the node with the reference vector in greatest agreement, and reference vectors for that node and the surrounding nodes were updated to better match the input vector. By repeating this learning process, reference vectors became similar to each other as they were closer to each other on the node map, and diverged as they became further away from each other. Since input vectors were grouped to each node, they were mapped to a low dimension while retaining the nonlinear relationship and the data structure could be visualized.

The algorithm follows the following general procedures:
I. Pre-processing of input data
II. Configuration of the SOM node map
III. Initialization of the reference vectors
IV. Learning of reference vectors
V. Grouping of input vectors

The following section discusses the details of each process and the parameter settings for this study.
I. Pre-processing of input data

The objective of this study was to group patterns in traffic jams, and data were used without pre-processing because actual traffic volume is important in evaluating traffic jams, and the components were all unit-less.

II. Configuration of the SOM node map

Node placement was determined on a 2-dimensional plane. The shape of the node lattice and the total number of nodes were determined and the XY ratio of the map was defined. Node lattice shape is typically either a rectangle or a hexagon, and this study employed a hexagonal lattice due to its ability to represent complex data with fewer nodes. The total number of nodes, M, was calculated by approximation using the number of total input data points n in (1) [16]:

\[ M = 5\sqrt{n} \]  

(1)

The XY ratio of the map was calculated by (2) [16]. \( n_1 \) represents the number of nodes in the vertical direction, \( n_2 \) represents the number of nodes in the horizontal direction, and \( e_1 \) and \( e_2 \) represent the first and second components of the covariance matrix of all the input data, respectively.
\[
\frac{n_1}{n_2} = \frac{e_1}{e_2}
\]

III. Initialization of the reference vectors

The reference vector \( w_i \) was initialized based on the major component analysis of input data, as shown in (3) [17]:

\[
w_i = x_{ave} + 5\sigma_2 b_2 l + 5\sigma_1 b_1 j
\]

where \( x_{ave} \) is the mean value of all input vectors, \( b_1 \) and \( b_2 \) are the first component vector and second component vector of the major component analysis, respectively, and \( \sigma_1 \) and \( \sigma_2 \) are the standard deviations of the input vectors against the axis. \( I \) and \( J \) represent the relative positions on the map, and their ranges are \(-1/2 \leq I \land J \leq 1/2\), respectively. Assuming the position of each node on the two-dimensional map to be \((p_i, q_i)\), and their mean as \((p_{ave}, q_{ave})\), \( I \) and \( J \) are given by the following (4) and (5):

\[
I = \frac{p_t - p_{ave}}{max p_i - min p_i}
\]

\[
J = \frac{q_t - q_{ave}}{max q_i - min q_i}
\]

IV. Learning of reference vectors

Learning style of reference vectors in SOM is grouped into sequential learning and batch learning. Sequential learning updates every time an input vector and a reference vector is compared, and is used in cases where there is a correlation in the order of the data and the order of data input affects the outcome. On the other hand, batch learning updates at once after all input data are compared with the reference vectors. Batch learning was applied for this study given that there were no correlations in the order of input data, and that outcomes would be more favorable without the impact of the order of the input. The algorithm for the batch algorithm was as follows [15]:

For each input vector \( x_j \), a node \( c \) with a reference vector, whose Euclidean distance is the shortest, is assigned as shown in (6):

\[
\| x_j(t) - w_c(t) \| = \frac{e_1}{e_2}
\]

where \( t \) is the number of learning steps. The reference vector was updated based on the node assigned using (7):
where \( c \) is the node assigned to an input vector \( x_j \), \( n \) is the number of input data values, and \( h_{ci} \) is an approximation function. \( h_{ci} \) represents the approximate relationship between node \( c \) and node \( i \), given by (8):

\[
h_{ci} = \exp \left( -\frac{||r_c - r_i||^2}{2\sigma^2} \right)
\]

where \( r_c \) and \( r_i \) are positional vectors of node \( c \) and node \( i \) on the map, respectively, and \( \sigma \) is the approximate radius. The approximate radius was given by (9) using the number of learning steps \( t \), number of total learning steps \( T \), and initial value of \( \sigma \), \( \sigma_0 \):

\[
\sigma = \sigma_0 - (\sigma_0 - 1) \frac{t}{T}
\]

V. Grouping of input vectors
After learning reference vectors, each input vector was grouped to a node with a reference vector whose Euclidean distance was the shortest as given by (6).

3.2. Clustering
Hierarchical clustering by Ward's method was performed to classify data into several clusters [18]. The DBI value was calculated for each number of clusters [19] to determine the number of clusters. The DBI value is an indicator that represents the cohesion within clusters and discreteness between clusters, as shown in (10), where clustering is optimized as the DBI value gets smaller.

\[
DBI = \frac{1}{N} \sum_{i=1}^{N} \max_{(i \neq j)} \frac{S_i + S_j}{M_{ij}}
\]

In the equation, \( N \) is the number of clusters, \( S_i \) is the center of a cluster \( i \) and the mean Euclidean distance between the elements in the cluster \( i \), and \( M_{ij} \) is the Euclidean distance between the centers of clusters \( i \) and cluster \( j \). The latter two parameters are given by (11) and (12). \( T_i \) is the number of elements in a cluster \( i \), and \( A_i \) is the center of cluster \( i \). \( X_{i,c} \) is an element in the cluster \( i \). In this study, a set of \( X_{i,c} \) is the same set of learned reference vectors.
3.3. Analysis method

In a clustering of 300 samples from case 1 alone, components were compared to further understand the characteristics of the clusters. Relationships in mean size were tested with data from both case 1 and case 2 using Steel's multiple comparison method. Next, in order to identify the difference between clusters, clusters with significant differences were extracted and tested by ANOM, and the results were grouped into three levels of traffic volume size. These three levels were relative to each cluster, and were not absolute levels that were compared with case 2. Traffic volume at intersections was visualized on a map for analysis, comparisons, and discussions. The following section explains each method.

1 Steel's multiple comparison method

When clustering, the size of the mean of the results from each cluster needs to be compared. Generally, a multiple comparison method is used to compare mean values from three or more groups, and Steel's method is one such method. Specifically, Steel's method is a non-parametric method that does not assume that the ensemble of each group follows a normal distribution, and simultaneously tests only the control group and other treatment groups [20].

2 Analysis of Means

Analysis of Means (ANOM) [21] was used to determine whether there were differences in mean values between groups. While variance analysis (ANOVA) is a common approach to testing significant differences in mean values, ANOVA tests to see if a significant difference exists by comparing mean values of the groups against each other while ANOM tests to see if there are significant differences between the mean of the whole and mean of a group [22]. If the mean of all groups is outside the confidence interval defined by the line of the limit of determination with respect to the individual group mean, then said group is deemed to have a significant difference with the mean of all the groups.

4 Characteristics analysis

A Kruskal-Wallis test was performed for each component of the data, i.e. traffic flow data at 57 observation points, to identify the major characteristics of each cluster. The Kruskal-Wallis test is a non-parametric testing method that tests the significance of the difference in medians among groups whose population does not follow a normal distribution [23]. Assuming that the
significant probability is \( p \), then the confidence interval is \( 1 - p \) and the significance of the difference is strengthened as the value of \( p \) decreases.

4. Results of the numerical tests

A total of 600 samples from a case with traffic jams and a case without traffic jams were clustered, to validate that the clustering method used by this study could effectively group simulation results based on the presence of traffic jams. Another clustering was performed on 300 samples with traffic jams, to demonstrate that the method could obtain more detailed results that focus on the correlation relationships among traffic volume.

4.1. Validation of the method combining SOM and hierarchal clustering

First, the traffic volume data from each intersection in case 1 and case 2 were combined, and SOM was applied to the 600 sample data points for clustering. Since there was a difference between the cases in terms of the occurrence of traffic jams, the validity of clustering could be verified in terms of traffic jams if they could be grouped.

SOM nodes were defined as 27 (vertical) by 5 (horizontal) using (1) and (2) in section 3.1, while the initial value of the approximate radius \( \sigma_0 \) was set at 10. Cluster analysis was performed based on the developed SOM by the method discussed earlier. Fig. 3 shows the relationship between the number of clusters and DBI values. 2 was determined to be the appropriate number of clusters that minimizes DBI value.

![Figure 3: Relationship between the number of clusters and DBI values for clustering of traffic jams](image)

The nodes were divided into two clusters, and Fig. 4 shows the cluster boundary as a bold line on a SOM node map with the plotted data. Blue and red represent the data from case 1 and
case 2, respectively. To ensure the number of individual datum included in the nodes were visually recognizable, variance was given to positions of the data displayed within the boundary of the affiliated node. Variation of each position in the vertical and horizontal directions was given as random numbers that follows the normal distribution with average 0 and standard deviation 0.12.

As shown in Fig. 4, data from case 1 and case 2 were clearly divided on the SOM node map. Data were clearly divided also by Ward’s method and clustering using the DBI value. The results show that case 1 and case 2 had a different relationship in traffic volume at intersections, and therefore different traffic conditions throughout the map.

This result was reasonable since traffic jams occurred in case 1 while they rarely occurred in case 2. Therefore, the clustering method used by this study effectively classified simulation results based on the presence of traffic jams.

4.2. Clustering of traffic jam patterns

Next, 300 sample data from case 1 were clustered and analyzed by applying SOM with 10 (vertical) x 9 (horizontal) nodes and an initial value of approximate radius $\sigma_0$ set at 4 to group the pattern for the case in which traffic jams occurred.

SOM was developed first, and clustering analysis was performed following the procedure discussed in section 3.3. Fig. 5 shows the SOM node map, and Fig. 6 shows the relationship between the number of clusters and DBI values. As with Section 4.1, variance was given to the data shown in Fig. 5.
As a result, 12 was the number of clusters at which DBI value was minimized. However, DBI values did not necessarily drop in an extreme manner compared to the DBI values from the clustered 600 samples shown in Fig. 3. This was because the characteristics of the data were not divided as clearly as when there were 600 data samples.

Considering the difficulty in comparisons after visualization if the number of clusters is 12, the number of clusters was set to 5, at which the DBI value was in a local minimum. The fact that DBI was in a local minimum suggests that the number of clusters with relatively more
significance was 5. Considering that the analysis was based on a hierarchical clustering, 5 clusters encapsulated 12 clusters, and making comparisons between 5 clusters was sufficiently useful.

Fig. 7 shows the node map colored for 12 clusters with boundaries drawn to represent 5 clusters. The clusters were also assigned numbers as shown in the figure. The average and variance of traffic volume in each cluster are displayed in Table 2 in Appendix. A.

Figure 7: Affiliated cluster of each node

Five clusters were tested for mean values using Steel's multiple comparison method and the ANOM method to discretize the traffic conditions. A MATES map was visualized based on the size of the traffic volumes obtained from the tests. As an example, Fig. 8 shows the results of clusters 1 and 4. Traffic volume that was greater than case 2 is shown in blue, traffic volume that was less than case 2 is shown in red, and equal traffic volume is shown in black. The size was expressed by the size of the circle, where a greater size represented a greater difference. "Low-traffic volume case" means that traffic volume was decreasing due to traffic jams, while "high-traffic volume case" means vehicles that take a detour as a result of surrounding traffic jams were observed.
Fig. 8 shows that inflow traffic in the north-south direction at intersections I, J, and K (which are aligned in a north-south direction) was in cluster 1, while the inflow traffic was greater in cluster 4. Compared to cluster 1, cluster 4 was characterized by an increase in detoured vehicles on its roads in a north-south direction.

### 4.3. Characteristics analysis of each cluster

Finally, a Kruskal-Wallis test was performed on the variables to find the significance probability $p$ to summarize discussions for the entire clusters, in a manner similar to that of the previous section. Relationship in traffic volume was discussed for each cluster for the top 15 points whose probability $p$ was low, in other words, significance in difference was strong. Fig. 9 shows the 15 roads. Intersections J, K, and M were continuous intersections, and were combined into a single road given that they exhibited similar changes when visualized. Each road was also assigned a number, as shown in the figure.
Table 1 shows a summary of the correlations between roads in terms of the difference in data obtained by conventional methods and the proposed method. Signs +, −, and ± in the table show whether the value was larger, smaller, or at an equal level compared to the overall tendencies, respectively. A value of 0 in a traffic jam meant that traffic volume was low to begin with and traffic jams never occurred. Conventional methods were all marked with ± since they resulted only in mean values.

| Method                        | Proposed Method | Conventional Method | Total |
|-------------------------------|-----------------|---------------------|-------|
| Cluster                       | 1   | 2   | 3   | 4   | 5   |       |      |
| Traffic jams in roads 1, 9, and 11 | +   | -   | +   | -   | ±   | ±     | ±    |
| Detoured vehicle in road 6    | +   | -   | +   | ±   | ±   | ±     | ±    |
| Traffic jam in road 3         | 0   | 0   | +   | -   | ±   | ±     | ±    |
| Detoured vehicle in road 8    | -   | -   | ±   | +   | +   | ±     | ±    |
| Traffic jam in road 7         | -   | -   | ±   | -   | ±   | ±     | ±    |
| Traffic jam in road 10        | +   | +   | ±   | -   | ±   | ±     | ±    |
| Detoured vehicle in road 5    | +   | +   | -   | ±   | -   | ±     | ±    |

As shown in Table 1, correlations between points of observation were clearer. More detailed data were obtained by clustering analysis of characteristics compared to the analysis based on the mean value of the whole. In particular, traffic jams and levels of detoured vehicles in specific clusters could be understood, and clusters could be treated as scenarios that might occur under specific conditions in cases where simulation results were treated in a qualitative
Traffic phenomena appear when multiple phenomena affect each other, and obtaining more detailed information that takes the correlations into account is very useful, and it can be said that it is appropriate information. In light of the above, the clustering method used by this study was effective in the analysis of the traffic simulations.

5. Conclusions

This study proposed a cluster analysis method on a series of traffic simulation results, and applied the method to an actual traffic volume data from the Okayama Station area. A total of 600 samples from a case with traffic jams and a case without traffic jams were clustered, and it was verified that data can be grouped with the presence of traffic jams. Another clustering was performed on 300 samples with traffic jams, and it was found that more detailed results that focus on the correlation relationships among traffic volume could be obtained. Clusters can be used as scenarios that occur under specific conditions because the results can be grouped into an appropriate number of clusters that provide intuitive feedback for the assessment of traffic policies and for new traffic planning and design.

Data used for this study was traffic volume data under specific parameters. However, traffic simulation results may change significantly or cause errors in practice. This is due to a large number of unknowns and parameters, such as road environment and traffic demand, which are often set as arbitrary values or estimations. Therefore, there is room for further evaluations in order for the proposed method in this study to be considered a practical method. The first point of consideration is whether the same method can perform the analysis as intended for simulation data in which parameters vary. Such an assessment is crucial for comparisons of policy assessment with multiple conditions. Another point of consideration is the evaluation of the sensitivity to changes in parameters. Errors are bound to be mixed into simulation results in such a case, and the robustness of the method can be evaluated in detail by understanding how properties of the clusters change by different degree of errors and variance. More specific tests and discussions will be conducted in future studies to address these concerns.

Acknowledgment

The Okayama Prefectural Police Department provided data and knowledge for this study. A part of this study was also funded by JSPS Kakenhi 15H01785. We would like to express our sincere gratitude.

Appendix A. Average and variance of traffic volume in each cluster

See Table 2. ID in the table denotes the ID for the point of measurement.
Table 2: Average ($\mu$) and variance ($\sigma^2$) of traffic volume in each cluster

| ID | Cluster 1 $\mu$ | Cluster 1 $\sigma^2$ | Cluster 2 $\mu$ | Cluster 2 $\sigma^2$ | Cluster 3 $\mu$ | Cluster 3 $\sigma^2$ | Cluster 4 $\mu$ | Cluster 4 $\sigma^2$ | Cluster 5 $\mu$ | Cluster 5 $\sigma^2$ |
|----|-----------------|----------------------|-----------------|----------------------|-----------------|----------------------|-----------------|----------------------|-----------------|----------------------|
| 1  | 731             | 489                  | 734             | 656                  | 731             | 847                  | 731             | 795                  | 726             | 564                  |
| 2  | 535             | 379                  | 529             | 492                  | 526             | 481                  | 529             | 430                  | 516             | 374                  |
| 3  | 419             | 655                  | 402             | 718                  | 410             | 842                  | 401             | 870                  | 415             | 789                  |
| 4  | 346             | 436                  | 367             | 498                  | 362             | 470                  | 364             | 610                  | 382             | 414                  |
| 5  | 993             | 821                  | 981             | 804                  | 972             | 611                  | 980             | 790                  | 967             | 1048                 |
| 6  | 676             | 685                  | 660             | 1259                 | 656             | 720                  | 660             | 845                  | 637             | 1058                 |
| 7  | 444             | 416                  | 456             | 527                  | 455             | 517                  | 462             | 622                  | 471             | 327                  |
| 8  | 644             | 973                  | 659             | 662                  | 642             | 642                  | 646             | 698                  | 653             | 1016                 |
| 9  | 704             | 685                  | 752             | 822                  | 700             | 1215                 | 729             | 999                  | 735             | 1128                 |
| 10 | 894             | 476                  | 893             | 587                  | 888             | 459                  | 885             | 553                  | 884             | 473                  |
| 11 | 136             | 185                  | 138             | 156                  | 140             | 171                  | 139             | 129                  | 143             | 121                  |
| 12 | 460             | 1672                 | 464             | 1281                 | 468             | 1360                 | 442             | 1290                 | 433             | 1123                 |
| 13 | 214             | 285                  | 230             | 271                  | 222             | 482                  | 219             | 299                  | 231             | 298                  |
| 14 | 372             | 573                  | 374             | 459                  | 379             | 416                  | 363             | 383                  | 382             | 551                  |
| 15 | 962             | 933                  | 956             | 979                  | 966             | 978                  | 982             | 740                  | 964             | 1025                 |
| 16 | 589             | 913                  | 581             | 866                  | 601             | 860                  | 576             | 1164                 | 579             | 1051                 |
| 17 | 889             | 1078                 | 891             | 843                  | 862             | 872                  | 882             | 865                  | 877             | 671                  |
| 18 | 1070            | 210                  | 1070            | 336                  | 1066            | 219                  | 1072            | 246                  | 1074            | 179                  |
| 19 | 882             | 763                  | 873             | 570                  | 883             | 1086                 | 899             | 473                  | 883             | 751                  |
| 20 | 834             | 372                  | 835             | 367                  | 832             | 366                  | 845             | 503                  | 836             | 507                  |
| 21 | 410             | 370                  | 412             | 298                  | 397             | 329                  | 407             | 368                  | 400             | 451                  |
| 22 | 1225            | 1007                 | 1223            | 986                  | 1215            | 901                  | 1211            | 1216                 | 1240            | 459                  |
| 23 | 457             | 457                  | 455             | 291                  | 456             | 418                  | 452             | 381                  | 451             | 343                  |
| 24 | 425             | 247                  | 429             | 360                  | 427             | 224                  | 434             | 288                  | 427             | 225                  |
| 25 | 465             | 775                  | 452             | 462                  | 441             | 442                  | 450             | 645                  | 433             | 505                  |
| 26 | 0               | 0                    | 0               | 0                    | 0               | 0                    | 0               | 0                    | 0               | 0                    |
| 27 | 539             | 419                  | 542             | 592                  | 546             | 709                  | 534             | 735                  | 546             | 605                  |
| 28 | 369             | 898                  | 373             | 1073                 | 345             | 1253                 | 354             | 837                  | 368             | 887                  |
| 29 | 76              | 75                   | 77              | 77                   | 76              | 98                   | 78              | 65                   | 75              | 77                   |
| 30 | 407             | 1048                 | 429             | 902                  | 450             | 1435                 | 434             | 1096                 | 488             | 834                  |
### Journal of Advanced Simulation in Science and Engineering

**continued**

| ID | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 | Cluster 5 |
|----|-----------|-----------|-----------|-----------|-----------|
|    | $\mu$    | $\sigma^2$| $\mu$    | $\sigma^2$| $\mu$    | $\sigma^2$| $\mu$    | $\sigma^2$| $\mu$    | $\sigma^2$|
| 31 | 59       | 62        | 60       | 70        | 57       | 90        | 60       | 53        | 55       | 63        |
| 32 | 307      | 299       | 304      | 312       | 307      | 332       | 321      | 333       | 313      | 285       |
| 33 | 689      | 406       | 676      | 712       | 659      | 524       | 675      | 695       | 640      | 791       |
| 34 | 778      | 322       | 824      | 556       | 776      | 475       | 809      | 634       | 803      | 827       |
| 35 | 98       | 89        | 99       | 103       | 99       | 96        | 99       | 73        | 97       | 84        |
| 36 | 890      | 536       | 902      | 611       | 915      | 605       | 937      | 441       | 922      | 430       |
| 37 | 107      | 142       | 95       | 152       | 105      | 180       | 103      | 206       | 97       | 183       |
| 38 | 654      | 348       | 693      | 386       | 655      | 380       | 682      | 491       | 677      | 431       |
| 39 | 135      | 125       | 131      | 59        | 131      | 76        | 133      | 114       | 125      | 43        |
| 40 | 783      | 630       | 799      | 542       | 811      | 639       | 833      | 484       | 822      | 502       |
| 41 | 6        | 6         | 5        | 5         | 6        | 7         | 5        | 5         | 6        | 5         |
| 42 | 510      | 724       | 551      | 583       | 499      | 754       | 533      | 808       | 520      | 564       |
| 43 | 720      | 240       | 721      | 228       | 719      | 235       | 722      | 312       | 718      | 240       |
| 44 | 873      | 562       | 881      | 589       | 894      | 581       | 932      | 492       | 906      | 482       |
| 45 | 2        | 2         | 2        | 3         | 1        | 2         | 2        | 1         | 3        |           |
| 46 | 85       | 274       | 84       | 272       | 90       | 289       | 87       | 226       | 88       | 198       |
| 47 | 141      | 63        | 141      | 51        | 143      | 57        | 143      | 52        | 143      | 56        |
| 48 | 766      | 233       | 761      | 189       | 753      | 204       | 759      | 208       | 745      | 190       |
| 49 | 224      | 389       | 223      | 373       | 221      | 320       | 240      | 308       | 225      | 303       |
| 50 | 960      | 452       | 992      | 436       | 946      | 514       | 977      | 632       | 964      | 419       |
| 51 | 44       | 56        | 44       | 50        | 44       | 36        | 45       | 34        | 47       | 57        |
| 52 | 1094     | 801       | 1099     | 648       | 1118     | 746       | 1157     | 623       | 1130     | 565       |
| 53 | 44       | 40        | 46       | 23        | 45       | 23        | 46       | 26        | 42       | 27        |
| 54 | 144      | 217       | 145      | 402       | 150      | 324       | 151      | 406       | 150      | 297       |
| 55 | 108      | 101       | 109      | 112       | 111      | 130       | 113      | 116       | 113      | 100       |
| 56 | 160      | 140       | 158      | 252       | 154      | 271       | 152      | 224       | 156      | 177       |
| 57 | 292      | 206       | 295      | 530       | 288      | 364       | 296      | 411       | 285      | 466       |

### References

[1] Y. Iida: *Koutsu Kougaku* [Traffic Engineering], Kokumin Kagakusha, Tokyo, 1992. (in Japanese)

[2] Japan Society of Traffic Engineering: *Kaitei Heimen Kousa no Keikaku to Sekkei Kiso hen Dai 3 han* [Planning and Design of Grade Crossing: The Basis (Rev. 3rd ed.)], Maruzen, Tokyo, 2007. (in Japanese)

[3] Japan Society of Traffic Engineering: *Koutsuu Simulation Katsuyou no Susume* [Recommendation of Exploiting Traffic Simulation], Maruzen, Tokyo, 2012. (in Japanese)

[4] T. Yamada, K. Ishikawa, K. Abe, H. Fujii, et al: Uncertainty Quantification in Multi-Agent Traffic Simulation, *Transaction of the Japan Society for Simulation Technology*, 9:1 (2017), 1-9. (in Japanese).

[5] H. B. Celikoglu: Dynamic classification of traffic flow patterns simulated by a switching multimode discrete cell transmission model, *IEEE Transactions on Intelligent Transportation Systems*, 15:6 (2014), 2539-2550.

[6] M. A. Silgu, H. B. Celikoglu: Clustering Traffic Flow Patterns by Fuzzy C-Means
Method: Some Preliminary Findings, in *International Conference on Computer Aided Systems Theory*, Springer International Publishing, 756-764, 2015.

[7] W. Weijermars, E. van Berkum: Analyzing highway flow patterns using cluster analysis, in *2005 IEEE Intelligent Transportation Systems Conference*, 308-313, Vienna, Austria, 2005.

[8] Y. Chen, J. Kim, H. S. Mahmassani: Pattern recognition using clustering algorithm for scenario definition in traffic simulation-based decision support systems, in *17th Int. IEEE Conf. Intelligent Transportation Systems*, 798-803, Qingdao, China, 2014.

[9] H. Fujii, T. Sakurai, S. Yoshimura: Virtual Social Experiment of Tram Railway Extension Using Multi-Agent-Based Traffic Simulator, *Journal of Advanced Computational Intelligence and Intelligent Informatics*, 15:2 (2011), 226-232.

[10] H. Uchida, H. Fujii, S. Yoshimura, S. Arai: Learning Routing Policy for Changes in Road Network, *IPSJ Journal*, 53:11 (2012), 2409-2418. (in Japanese)

[11] H. Fujii, S. Yoshimura, T. Kanou, S. Ninomiya: Microscopic Simulation-based Assessment of Impact of Large-scale Retail Store onto Regional Traffic, in *Proc. 14th International Conference on Computers in Urban Planning and Urban Management*, Cambridge, USA, 2015.

[12] S. Yoshimura: MATES: Multi-Agent Based Traffic and Environment Simulator-Theory, Implementation and Practical Application. *Computer Modeling in Engineering and Sciences*, 11:1 (2006), 17-25.

[13] H. Fujii, S. Yoshimura, K. Seki: Multi-agent-based Traffic Simulation at Merging Section Using Coordinative Behavior Model. *Computer Modeling in Engineering and Sciences*, 63:3 (2010), 265-282.

[14] H. Fujii, T. Kanou, H. Uchida, S. Yoshimura: Simulation-based Assessment of Impact of Large-scale Retail Store onto Regional Traffic and Its Management, in *Proc. of 14th World Conference on Transport Research*, Shanghai, China, 2016.

[15] T. Kohonen: *Self-Organizing Maps* (translated by H. Tokutaka, M. Oyabu, K. Horio, K. Fujimura, M. Ohkita) (Rev. ed.), Springer Japan, Tokyo, 2005. (in Japanese)

[16] L. G. Hilario, M. G. Ivan: Self-organizing Map and Clustering for Wastewater Treatment Monitoring, *Engineering Applications of Artificial Intelligence*, 17:3 (2004), 215-225.

[17] T. Abe, S. Kanaya, M. Kinouchi, T. Ikemura: Genome DNA Hairetsu ni Hisondeiru Deibutushu ni Kosei wo Akiraka ni Suru Shinki na Toukei Suuriteki Shuhou [New Statistic-Mathematical Method to Reveal Characteristics of Species Concealed in Genome DNA Sequence], *Proceedings of the Institute of Statistical Mathematics*, 52:1 (2004), 207-215. (in Japanese)

[18] M. R. Anderberg: *Cluster Analysis and Its Applications* (translated by H. Nishida, T. Sato, K. Eto, Y. Terao, M. Miyai), Uchida Roukakuho, Tokyo, 1988. (in Japanese)

[19] D. L. Davies, D. W. Bouldin: A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1:2 (1979), 224–227.
[20] R. G. D. Steel: A multiple comparison rank sum test: Treatments versus control, *Biometrics*, 15 (1959), 560-572.

[21] E. R. Ott: Analysis of Means -- A Graphical Procedure, *Industrial Quality Control*, 24 (1967), 101-109. Reprinted in *Journal of Quality Technology*, 15 (1983), 10-18.

[22] P. Pallmann, L. A. Hothorn: Analysis of means (ANOM): A generalized approach using R, *Journal of Applied Statistics*, 43:8 (2016), 1541-1560.

[23] W. H. Kruskal, W. A. Wallis: Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association*, 47:260 (1952), 583-621.