

# Cluster Analysis of Car Parking Data, and Development of their Web Applications

Takafumi Kubota<sup>1,a</sup>, Takayuki Hayashi<sup>b</sup>, Tomoyuki Tarumi<sup>c</sup>

<sup>a</sup>Risk Analysis Research Center, The Institute of Statistical Mathematics

<sup>b</sup>Graduate School of Environmental Science, Okayama University

<sup>c</sup>Admission Centre, Okayama University

---

## Abstract

In this paper, we apply cluster analysis to “Okayama parking data” that is one of the spatial point patterns data that includes locations and the fare structure of car parking space in Okayama central area. This study classifies the characteristics of small areas through Okayama parking data as well as visualizes the results of the cluster analysis. We develop web applications that connect the results of a cluster analysis and overlay objects including points of balloons and rectangles of small areas over a map of Okayama central area.

Keywords: Cluster analysis, parking-data, web-application.

---

## 1. Introduction

Recently the number of cars has been increasing due to lifestyle changes. This situation increases the necessity of car parking spaces in a central area. A typical example of this increased need for parking space exists in the medium-sized city of Okayama in Japan.

For previous studies about car parking space, Blackburn town centre renaissance (2007) showed parking strategy and transport strategy at Blackburn, MVA for South East of England Regional Assembly (2005) showed regional assembly parking standards.

Little attention has been given to an analysis for data including locations and fare structure, even though users who want to park their cars need information where car parking spaces are and how much money they should pay for parking. Therefore, the first purpose of our study is to perform a cluster analysis for the fare structure data of car parking space in Okayama central area. A result of this analysis shows the clustering of fare structure data of car parking space as well as the clustering of characteristics in the Okayama central area. Our objective is neither the prediction of the price of car parking nor the creation general measurement of criteria for the price of car parking. The objective is for the grouping of the similar price structures of car parking to characterize industry or commerce of Okayama central area through the use of cluster analysis.

Two kinds of difficult problems sometimes arise in some cluster analysis results. One is a linkage between a particular datum and its corresponding cluster from the result. Another is a number of clusters that suits for the result of cluster analysis. In order to solve these two problems we develop web application for the second purpose of our study.

We describe cluster analysis and applying mesh data for Okayama parking data in Section 2, we show development of web application and its example in Section 3 and, finally present our concluding remarks and future studies in Section 4.

---

<sup>1</sup> Corresponding author: Assistant Professor, Risk Analysis Research Center, The Institute of Statistical Mathematics, 10-3, Midori-chou, Tachikawa-shi, Tokyo 190-8562, Japan. E-mail: tkubota@ism.ac.jp

Table 1: The fare structure of  $p_{i,j,k}$ 

| Type of date<br>O'clock | weekday       |               |     |                | weekend       |               |     |                |
|-------------------------|---------------|---------------|-----|----------------|---------------|---------------|-----|----------------|
|                         | 0-1           | 1-2           | ... | 23-24          | 0-1           | 1-2           | ... | 23-24          |
| 1                       | $p_{0,1,1}$   | $p_{1,1,1}$   | ... | $p_{23,1,1}$   | $p_{0,2,1}$   | $p_{1,2,1}$   | ... | $p_{23,2,1}$   |
| 2                       | $p_{0,1,2}$   | $p_{1,1,2}$   | ... | $p_{23,1,2}$   | $p_{0,2,2}$   | $p_{1,2,2}$   | ... | $p_{23,2,2}$   |
| ⋮                       | ⋮             | ⋮             | ⋮   | ⋮              | ⋮             | ⋮             | ⋮   | ⋮              |
| 300                     | $p_{0,1,300}$ | $p_{1,1,300}$ | ... | $p_{23,1,300}$ | $p_{0,2,300}$ | $p_{1,2,300}$ | ... | $p_{23,2,300}$ |

Table 2: The example of fare structure

| Type of date<br>O'clock   | weekday |     |     |       | weekend |     |
|---------------------------|---------|-----|-----|-------|---------|-----|
|                           | 0-1     | 1-2 | ... | 12-13 | ...     | ... |
| ⋮                         | ⋮       | ⋮   | ⋮   | ⋮     | ⋮       | ⋮   |
| Park'n park Nakasange 7th | 33      | 33  | ... | 200   | ...     | ... |
| ⋮                         | ⋮       | ⋮   | ⋮   | ⋮     | ⋮       | ⋮   |
| C.K parking               | 400     | 400 | ... | 200   | ...     | ... |
| ⋮                         | ⋮       | ⋮   | ⋮   | ⋮     | ⋮       | ⋮   |

## 2. Analysis for Okayama Parking Data

### 2.1. Details of the data

In this study we target the eastern area from Okayama station (a nearly square area that covers 2 km times 2 km both north-south and east-west directions) as the central area of Okayama city. In this area there are 300 car parking spaces. For field work, we gathered the data of location (longitude and latitude) and fare structure of them. In addition to explanations about fare structure, it is different from each parking space in a point of time (daytime or nighttime) and date (weekday or weekend). Therefore in this study we restructure from the original fare structure data to price per one hour, both weekday and weekend. There are many kinds of fare structure such as 100 yen per 15 minutes and 200 yen per 6 hours. Comparing these several kinds of price data is difficult and we calculate by hour to define 48 kinds of price variable as  $p_{i,j,k}$ ; the hour price (yen) from  $i$  o'clock to  $(i + 1)$  o'clock ( $i = 0, 1, \dots, 23$ ) in type of  $j$  ( $j = 1, 2$ ; 1 means weekday, 2 means weekend) at the index of parking ( $k = 1, 2, \dots, 300$ ). Table 1 shows the fare structure of  $p_{i,j,k}$ , and Table 2 shows an example of it.

### 2.2. Cluster analysis

The theoretic of cluster analysis is specified in Tanaka and Tarumi (1995). Cluster analysis is the assignment of a set of observation into subsets (called clusters) in order to combine observations that have similar senses from the mixture senses of observations. There are the hierarchical cluster analysis and non-hierarchical ones. We use the hierarchical cluster analysis because our objective is to develop hierarchical structures of Okayama car parking data in order to characterize the Okayama central area. The order of cluster analysis is as follows.

Step 1. Each element is supposed as each independent cluster.

Step 2. Dissimilarities  $d_{ij}$  are referred to combine clusters that have minimum dissimilarity.

Step 3. If there is only one cluster, then the cluster analysis is completed. If there are more than two clusters, then move to Step 4.

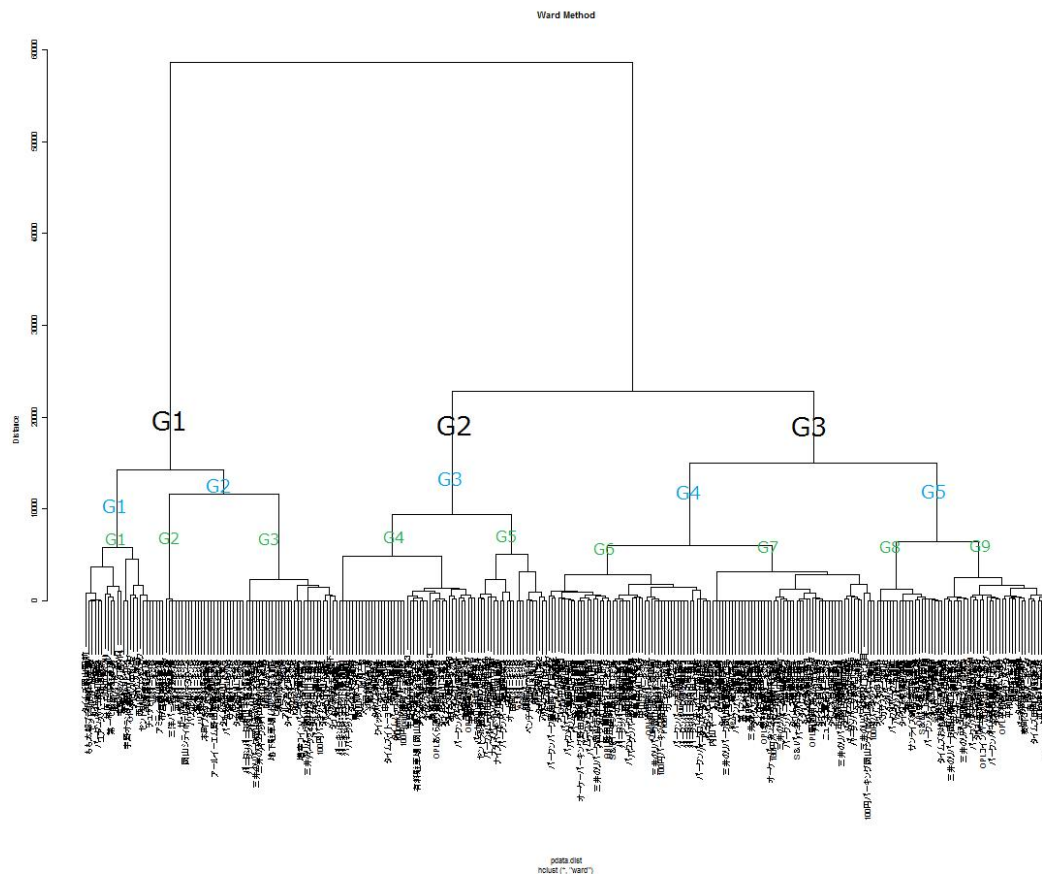


Figure 1: The result of cluster analysis(dendrogram)

Step 4. Dissimilarities are updated by new cluster that is combined at Step 2 and the other clusters, then back to Step 2.

In the Step 2 there are four kinds of distance definitions for measuring the observation dissimilarities.

- Squared Euclidean distance
- Standardized squared Euclidian distance
- Mahalanobis' generalized distance
- Minkowsky distance

We avoid to complicate the calculation for clustering to use normal squared Euclidean distance, because we use only fare price variables.

In the Step 4 there are seven kinds of methods to combine clusters as follows.

- Nearest neighbor method
- Furthest neighbor method

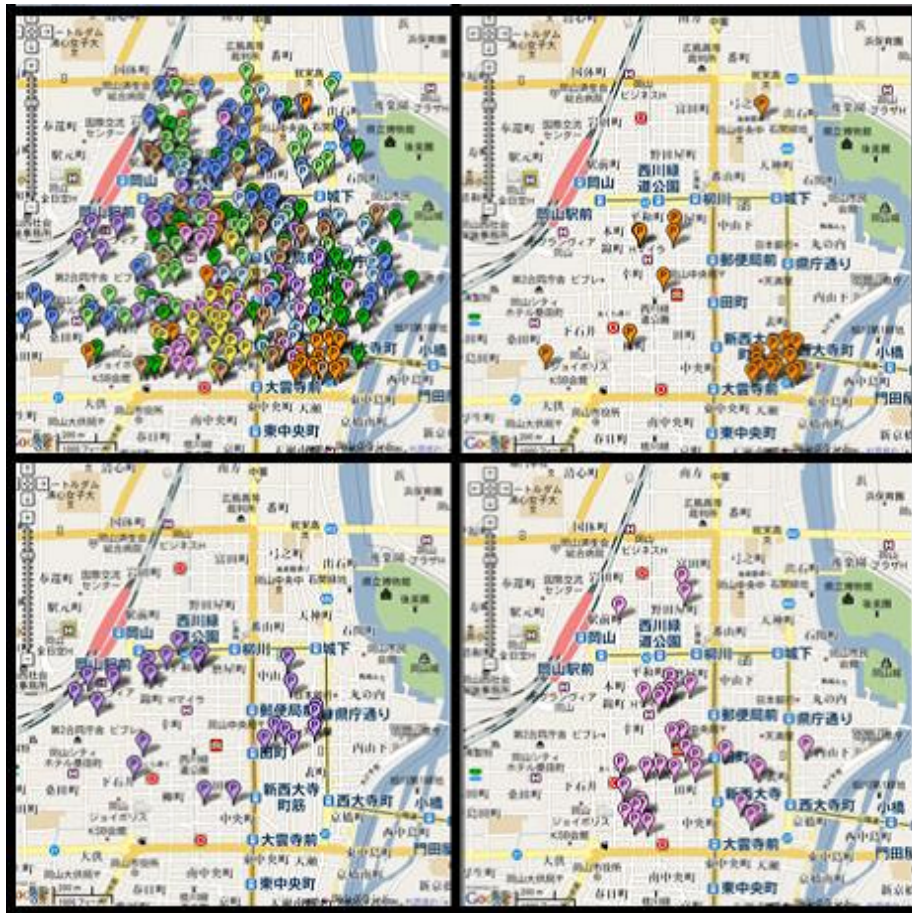


Figure 2: 9 groups (top left), group 1 (top right), group 7 (bottom left) and group 8 (bottom right)

- Group average method
- Centroid method
- Median method
- Ward method
- Flexible method

Because we use squared Euclidean distance for the definition of the dissimilarities and we need a simple dendrogram, we use the Ward method. We show the result of cluster analysis. Figure 1 shows dendrogram of the result.

There can be found some suitable numbers of clusters; 3, 5 or 9. We treat the result of 9 clusters to interpret some reasons from background. Figure 2 shows nine kinds of balloons that identify the groups (clusters), and characteristic group 1, 2 and 7. Additionally Table 3 shows an interpretation and background of each group of cluster.

Table 3: Interpolation and background of each group of cluster

| Group | Interpretation                               | Background                     |
|-------|--|--------------------------------|
| 1     | low price on all time in all date            | far from centre                |
| 2     | high price on afternoon                      | business and education         |
| 3     | high price on all time                       | busy street                    |
| 4     | low price on early morning                   | business                       |
| 5     | low price on early morning (higher than 4)   | business and much traffics     |
| 6     | higher price on all time in all date         | face to big street             |
| 7     | the highest price on all time in all date    | near station and shopping area |
| 8     | higher price on night and early morning      | restaurant and cafe            |
| 9     | the highest price on night and early morning | bar and pub                    |

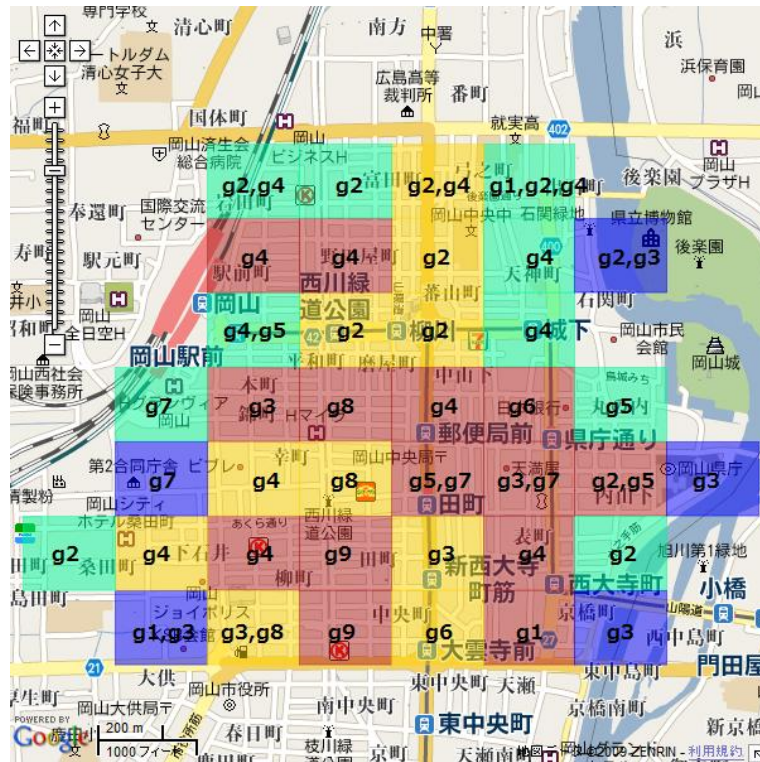


Figure 3: The representative cluster number over the colored mesh

### 2.3. Mesh

While we show the way of pointing car parking space in its longitude and latitude on the map of Okayama central area, there are another ways to show the results by area. For one of them we use grid cut area vertically in both longitude and latitude (that is called a mesh). There are some kinds of mesh with original code (mesh code), from big to small size, for example the first mesh covers 1 degree in longitude direction and 40 minutes in latitude direction, approximately 80km times 80km in Japan. At first we color a mesh (5<sup>th</sup> bigger mesh) in gradations from blue to red corresponding to small to big by a number of car parking spaces. Then we regard the populous group (cluster) as a representative cluster, that means the number of the representative cluster is biggest in each mesh. Figure 3 shows the representative cluster number over the colored mesh.

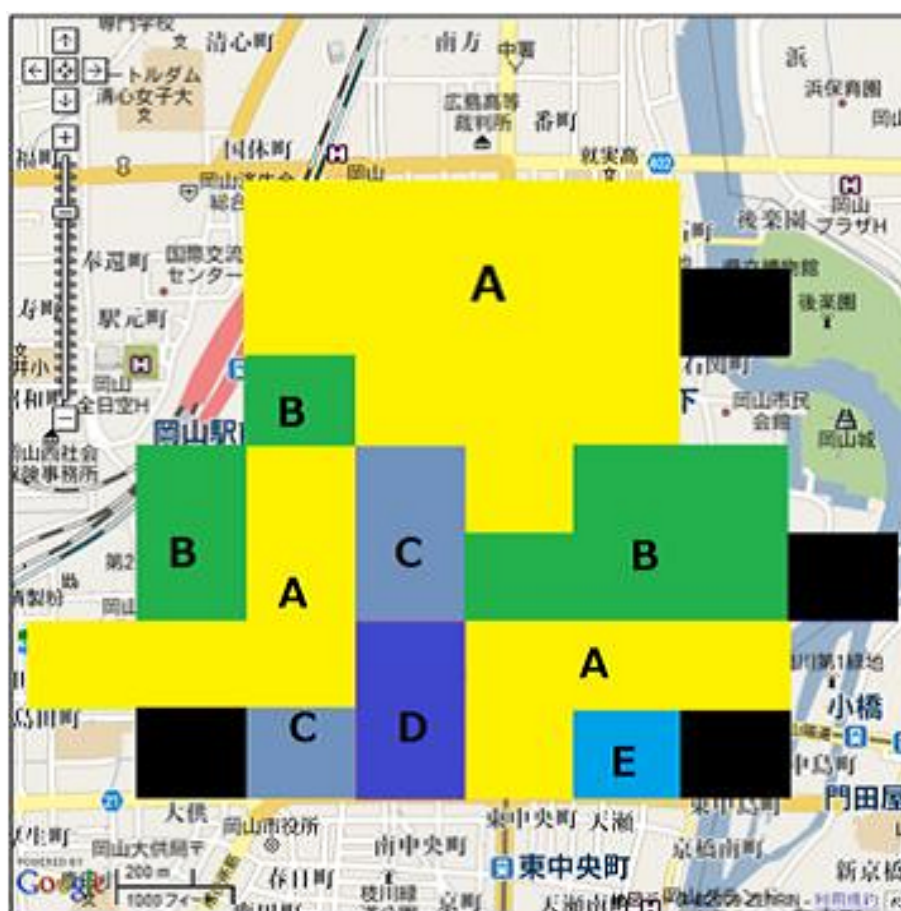


Figure 4: Combined 5 groups

Table 4: Interpolation and background of each group of cluster

| Group | Corresponding group | Background          |
|-------|---------------------|---------------------|
| A     | 2,3 and 4           | Standard district   |
| B     | 5,6 and 7           | Shopping area       |
| C     | 8                   | Restaurant district |
| D     | 9                   | Nightlife circuit   |
| E     | 1                   | Far from centre     |

We combine the newer five clusters from the result of nine clusters in an aspect of both the result of clusters' similarities and real distance of the mesh (Figure 4). Table 4 shows the explanation of each group and corresponding old group. This is only one result and might have arbitrariness. Therefore, we develop web application for users to interactively check the results.

### 3. Development of Web Application

It is difficult to consider a linkage between a particular car parking space and the corresponding cluster from the result of the cluster analysis in Section 2. Moreover it is a difficult problem of how many

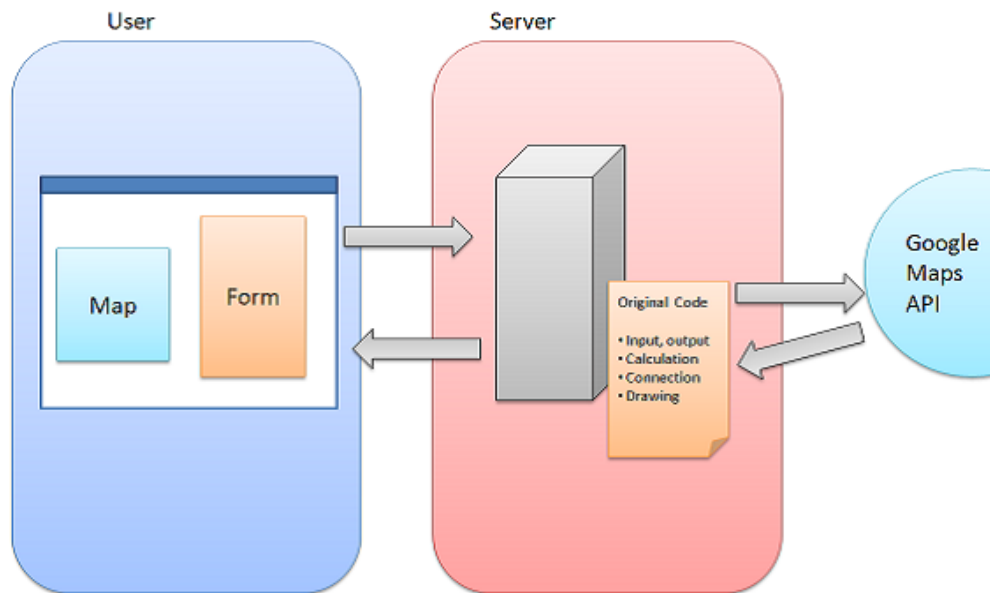


Figure 5: Flow of our system

clusters should we treat for the car parking data. In order to solve these two problems we develop the web application “MeshCodeViewer” as the second purpose of our study. This application facilitates the user to solve these problems through interactive handling.

### 3.1. Environment

We suppose the independence of system (OS or browser) as well as hardware (for example PC, cell phones or tablets) for the environment. For this purpose we chose HTML, Javascript and Google Maps API and we script our own functions. The directive of our web application is as follows URL:

<http://face.f7.ems.okayama-u.ac.jp/mcv/MeshCodeViewer.html>

### 3.2. Development and examples

Original code that is scripted for a class of viewing point data and Mesh data is as follows

- Input and output data
- Calculating position and density of Mesh
- Connecting to Google Maps API
- Drawing original Symbols

Figure 5 shows the flow of our system.

At first the user inputs local data (CSV (comma separated value) or TSV (tab separated value) is permitted) into the form area, then the user inputs a column number of longitude, latitude, and characteristic value. Additionally, the user can select options such as color or the size of balloon. Then the user selects the type of layer on map, point (balloon) or area (mesh). Then the data and

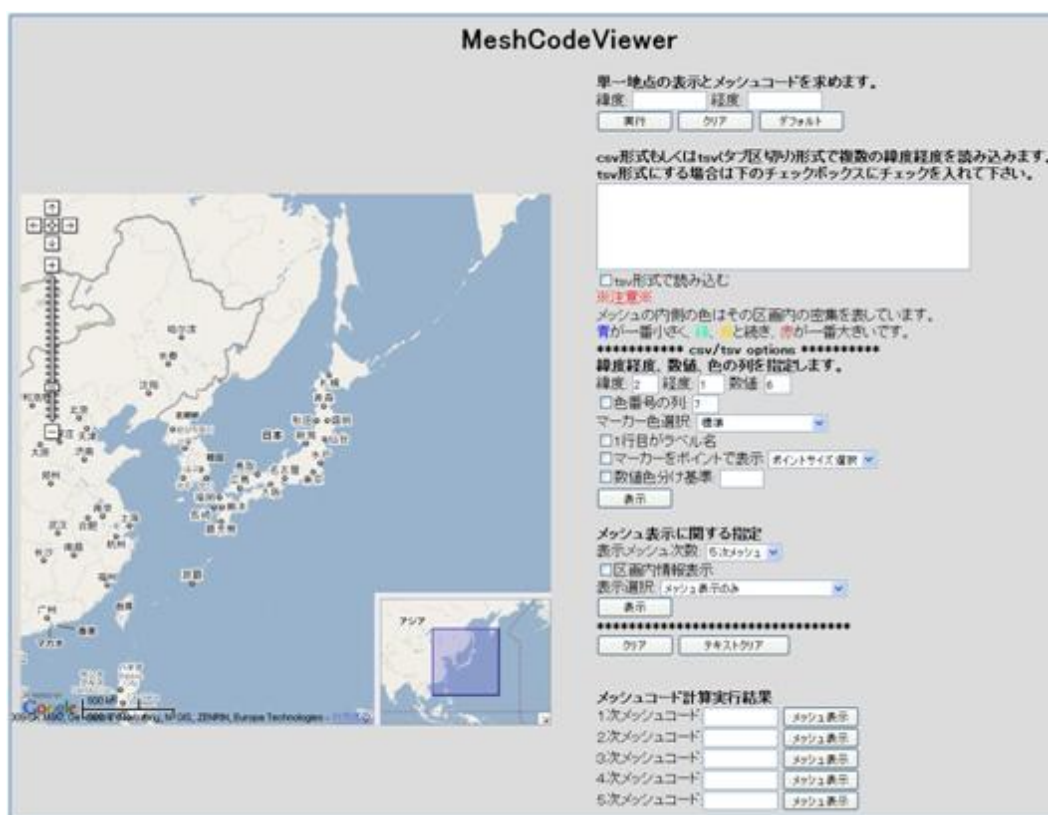


Figure 6: Top page of our system

selected option is submitted to the server to calculate and connect to Google Maps API. The result is back to the left side of browser. The Google Maps can handle the data in an interactive way such as zoom-in, zoom-out, and move to some places. Our script also can handle it to display information of the point by click in a case of point data, or can display density of point by color (a number of points increase as a color from blue to red) in a case of the mesh data. Figure 6 shows the top page of our system.

Figure 7 (left) shows an output of point data. When the user clicks any balloons, information of a point appears. Figure 7 (right) shows another output of colored mesh by a number of points.

#### 4. Concluding Remarks and Future Studies

In this study we applied cluster analysis to “Okayama Parking data” that includes 48 variables of each one for our fare structure of both weekdays and weekends. From the results, we recognized the characteristics of clusters such as shopping, business, restaurant and nightlife districts in the Okayama central area. The results showed an interesting relation between a result of cluster analysis without the location data for variables and location characteristics. We developed a web application to solve two problems, a linkage from original data and cluster, and a number of clusters. This application facilitated the user to solve these two problems as well as providing an initial step to analyze the Okayama parking data.





Figure 7: An output of point data (left) and the other output of mesh data (right)

We will consider about definition of variables in cluster analysis of Okayama parking data for our future studies such as the number of fare structures for each car parking space, a limitation cap price of it (for example only 1000 yen when parking over 6 hours to 24 hours), and the capacity of it (a number of cars can park there). Then we will extend to other areas by using some existing data bases. In addition, we will add functions of search and recommend car parking space for user who wants to find it.

## References

- Blackburn town centre renaissance (2007). *Blackburn Town Centre Parking Strategy and Transport Strategy Consultation document*, Available from: [http://www.blackburn.gov.uk/upload/pdf/Final\\_document\\_-\\_transport\\_and\\_parking.pdf](http://www.blackburn.gov.uk/upload/pdf/Final_document_-_transport_and_parking.pdf) (January 2011)
- MVA for South East of England Regional Assembly (2005). *Review of Regional Parking Standards, Report*, Available from: [http://www.southeast-ra.gov.uk/documents/transport/review\\_of\\_regional\\_parking\\_standards\\_v2.pdf](http://www.southeast-ra.gov.uk/documents/transport/review_of_regional_parking_standards_v2.pdf) (January 2011)
- Tanaka and Tarumi (1995). *A handbook of statistical analysis for windows - multivariate statistical analysis (in Japanese)*, Kyoritsu shuppan.

Received September 2010; Accepted March 2011