Comparing Urban Spatial Structures: A Model Selection Approach

K. Ito[1]

[1]Department of Architecture, Faculty of Science and Technology, Tokyo University of Science
2641 Yamazaki, Noda-shi, Chiba, 278-8510, JAPAN
Email: kaori@rs.noda.tus.ac.jp

1. Introduction

People gather together to form a society, of which the physical manifestation is a city. Each city's urban spatial structure can be considered as an underlying context of current apparent physical space. Considerable literature focusing on the spatial structure of cities has been published to date (Kostof 1993) and many researchers have investigated methods for measuring urban spatial structure (Anas et al. 1998), including spatial statistical approaches using land parcel-level point land use data (Cuthbert et al. 1998), fractal analysis using the shape of the built-up area (Frankhauser 2004), among others. In most cases, however, each approach only measures only one or very few aspects, such as the extent of clustering, and results cannot be compared between different models.

The aim of this research was to develop a method to compare different spatial structure models by using model selection criteria. In this paper, the perspective of comparing different model classes for feature distribution patterns by using the Minimum Description Length criterion is presented, and an experiment with artificial datasets is demonstrated.

For example, Fig. 1 shows the central areas of Tokyo, the present capital of Japan, and Kyoto, the ancient capital. Central Tokyo tends to have a concentric ring (originally spiral) structure that originated in Edo (Tokyo's former name) in the 17th century, while Kyoto tends to have a grid structure whose origin dates back to the late 8th century. There could be other types of urban spatial structures. Valley towns and ridge towns tend to have linear structures or structures with linear centers. Some villages have a relatively homogeneous structure with houses located a certain distance from each other. Many cities have a combination of two or more structures. In some cities, the urban spatial structure can be recognized at a glance, while other cities have an urban spatial structure that is more difficult to discern. The interest of this research is to measure the tendency for a city to have a particular urban spatial structure, and to distinguish which type using observed data such as the distribution of population, buildings, businesses or roads. The approach of this research is to select an

Figure 1. Map of Tokyo (above left) and Kyoto (above right) (source: Digital Japan)

appropriate model that best explains the observed data among given model classes.

2. Minimum Description Length Criterion

The Minimum Description Length (MDL) criterion is a relatively recent method of model selection proposed by Rissanen (1984), which, given limited observations, selects from among competing explanations of data. The MDL criterion selects the appropriate model that minimizes the code length for describing the data based on the assumption that "the more regularities there are, the more the data can be compressed" (Grünwald 2000).

Suppose that the locations of features such as houses or businesses are observed data generated by a certain structure. A model that better represents the underlying order, or urban spatial structure, could describe the locations of features using fewer symbols.

In the MDL principle, the code length for describing the data is given by two-part encoding: first, encoding the model itself, and second, encoding the data using the model. Optimization is selection of a model that minimizes the total code length, or the sum of the code length for describing the model and that for describing the data using the model, in the given model class, or even among different model classes. There is a trade-off relationship between the code lengths. A complex model, which increases the model code and decreases the data code, tends to cause over-fitting. Conversely, a rough model, which decreases the model code and increases the data code, tends to cause under-fitting.

3. Experiment with two model classes

In this paper, two different basic model classes were examined: the Cartesian coordinate rectangular partitioning model (CcRp model) (Ito 2006) and the polar coordinate angular radial partitioning model (PcARp model). The former might better explain a city with a relatively homogeneous urban spatial structure such as a grid pattern, while the latter might better explain a city with a relatively directional urban spatial structure or a multi-ring structure. Both model classes have a hierarchical structure of sub-regions corresponding to a binary tree. In the CcRp model, each node of the binary tree represents a rectangular sub-region in the square region, and each tree-splitting corresponds to a step of recursive divisions of a parent rectangle into two equal-sized child rectangles with horizontal and vertical cuts made alternately (Fig. 2 (a) and (c)). Likewise, in the PcARp model, each node of the binary tree represents a sub-region in the circular region, and each tree-splitting corresponds to a step of recursive divisions of a parent sub-region into two equal-areal child sub-regions with radial and angular cuts made alternately (Fig. 2 (b) and (c)). Sub-regions in the former model are a square or a 1:2 rectangle, while those in the latter model are not usually similar in shape with each other, but equal in area.

The location of a point in the region, which can be assumed to be any feature such as a house, is represented as a leaf node of a complete tree with depth $2M$ (Fig. 2 (c)). The location data is modeled based on a set of sub-regions in either model class, and the optimal set of sub-regions corresponding to the pruned or unpruned binary tree is searched.

The search for the optimal model within each model class proceeds under the following assumptions:

(1) the location is described by the probability distribution, and

(2) the parameters of the probability distribution vary from sub-region to sub-region.

Given a certain set of sub-regions, let $\mathbf{X}_i$ be a set of location data observed in the ith sub-region. Let $P(\mathbf{X}_i)$ be the probability that $\mathbf{X}_i$ is observed. In these terms, L, the total code length for describing the observed data under the set of sub-regions, is given by:

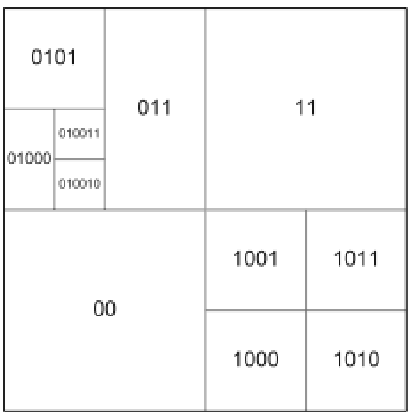
$$L = \left(2K - 1 + \sum_i r_i\right) + \sum_i\left(-\log_2 P(\mathbf{X}_i)\right) \text{ [bit]}, \tag{1}$$

where $2K$-1 is the code length for describing a binary tree with K leaves, r_i is the code length for describing the parameter in the ith sub-region and -$\log_2 P(\mathbf{X}_i)$ is the code length for describing $\mathbf{X}_i$, (Tsuchiya et al. 1996). The optimal model within each model class is obtained by minimizing L for possible sets of sub-regions. An algorithm using the recursive structure of a binary tree enables us to search efficiently for the global optimum. Moreover, the optimal from both classes is obtained by comparing the minimum L within the CcRp model class and the minimum L within the PcARp model class.

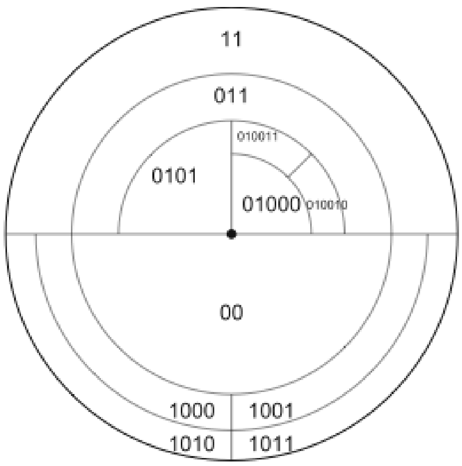
4. Experimental results and discussion

Three sets of artificial data with 10^4 points are generated: A. three Gaussian distributions, B. a circular distribution and C. a linear distribution (Fig. 3). The optimization results for each set of data within each model class are shown in Fig. 4. In the PcARp model, the entire region of the data is scanned to find the origin point that minimizes L. The maximum depth of a tree, $2M$, is set to be 16, and thus a complete tree has 2^{16} leaves.

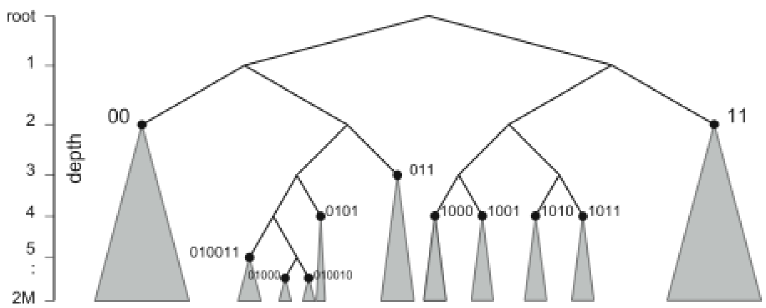
When the location is described point-by-point by a leaf of a complete tree, the code



(a) Set of sub-regions in the CcRp model



(b) Set of sub-regions in the PcARp model



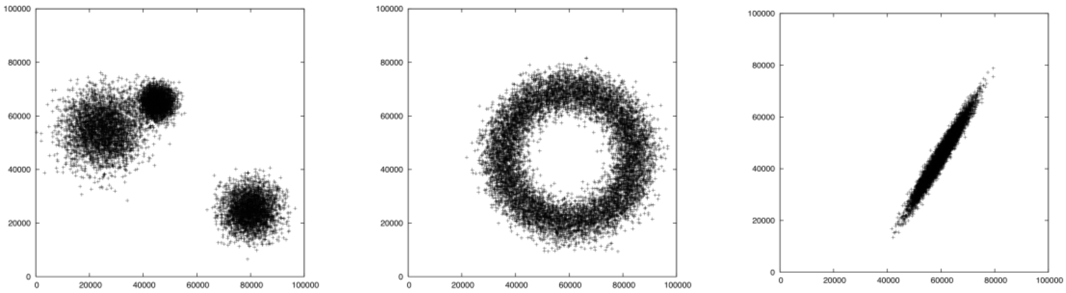
(c) Corresponding binary tree

Figure 2. Example of a set of sub-regions and corresponding binary tree

length per point is $2M = 16$ bits. The experimental results show that, based on the optimal models, the code length per point varies from 7.5 bits to 11.5 bits depending on spatial distribution and model class. Data set C is compressed the most in both model classes, which means the data was explained more efficiently than the other datasets. The PcARp model compresses all the sets of data slightly more effectively than the CcRp model. This indicates that every set of data is a little more reasonably explained based on a directional and/or ring structure than a homogeneous rectangular structure. However, in Fig. 4(b)B, the origin of the optimal polar coordinates does not correspond to the location of the center of the circular distribution. When the origin of the polar coordinate system is located at the center of the circular distribution, the code length is longer than when the origin is based on the optimal CcRp model (Fig. 5). A rather fine partition around the center in Fig. 5 may cause this longer code length. The recursive division with alternating radial and angular cuts may keep the area around the center from staying undivided. For data set C, the rate of improvement of the optimal PcARp model compared to the optimal CcRp model is higher than for the other datasets. The PcARp model has long, narrow sub-regions in both the central and peripheral areas, and is thus effective for a distribution like that of data set C, which extends in an oblique direction. Fig. 6 shows the details of the optimal models within the CcRp model class and the PcARp model class. In the optimal PcARp model, efficient division along the data distribution is observed. The number of sub-regions of the optimal PcARp model is less than two thirds of the number in the optimal CcRp model. Areas where the density changes rapidly are finely divided into many sub-regions, because a detailed description of the data is required. In some cases, as observed in the optimal CcRp model for data set C, this causes a division of neighboring sparse areas where a detailed description of the data is not necessary.

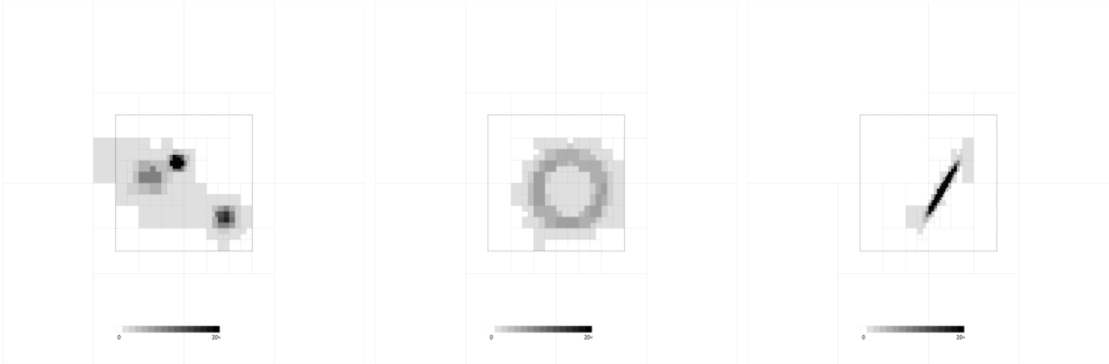
In this work-in-progress paper, a perspective and an experiment have been presented. Through the experiment, the spatial structural tendency of three sample datasets and the model selection behavior can be observed.

One of the current issues to be addressed is application of this concept to actual data. Comparison of cities such as Tokyo and Kyoto, as previously mentioned, is one option. In some cities, layers of different spatial structures might be observed. For example, the distribution of housing surrounding the inner city may be better explained by the PcARp model, while the distribution of retail agglomerations scattered over the region could be better explained by the CcRp model. Another issue is the development of other models that can explain urban spatial structures even better based on these results and applications to actual data. Ito (2006) indicated that this process of selecting the optimal model can also be considered as an optimization of aggregation units for geographically distributed features. The future directions of this research include contributions to this issue.



A. Three Gaussian distributions B. Circular distribution C. Linear distribution

Figure 3. Three artificial sets of data (size of each dataset is 10^4)



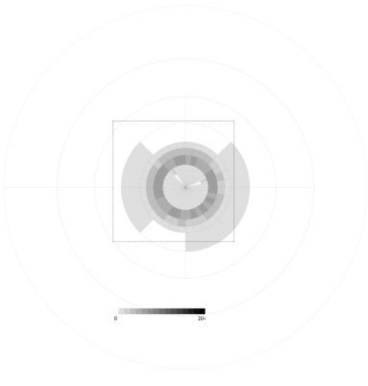
A. L = 10.29 (bits/point), 137 sub-regions B. L = 11.49 (bits/point), 131 sub-regions C. L = 7.66 (bits/points), 158 sub-regions

(a) Optimal Cartesian coordinate rectangular partitioning (CcRp) models

A. L = 10.26 (bits/point), 118 sub-regions B. L = 11.45 (bits/point), 101 sub-regions C. L = 7.54 (bits/point), 100 sub-regions

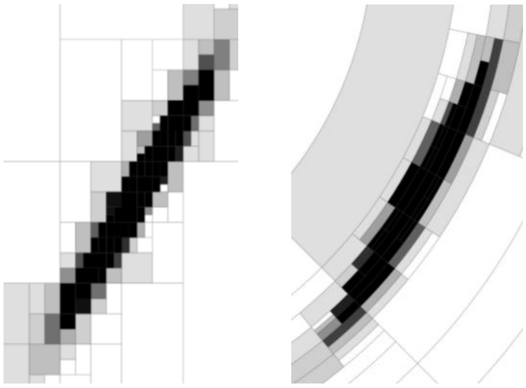
(b) Optimal polar coordinate angular radial partitioning (PcARp) models

Figure 4. Optimal model for each data within two model classes



L = 11.55 (bits/point), 149 sub-regions

Figure 5. Model with the coordinate origin is located at the distribution center



the CcRp model the PcARp model

Figure 6. Detailed view of the optimal models for data set C

References

Anas A, Arnott R and Small KA, 1998, Urban spatial structure. *Journal of Economic Literature*, 36, 1426-1464.

Cuthbert AL and Anderson WP, 1998, Using spatial statistics to examine the pattern of urban land development in Halifax-Dartmouth. *The Professional Geographer*, 54(4), 521 - 532.

Frankhauser P, 2004, Comparing the morphology of urban patterns in Europe: A fractal approach. In: Borsdorf A and Zambri P (eds), *European Cities: Insights on outskirts Report COST Action 10 Urban Civil Engineering*, 2, 79- 105.

Grünwald PD, Myung IJ and Pitt MA, 2000, *Advances in Minimum Description Length: Theory and applications*. The MIT Press.

Ito K, 2006, Designing optimal aggregation units for representing geographically distributed data using the MDL criterion. *GIScience 2006 Extended Abstracts*, 93-97.

Kostof S, 1993, *The city shaped: Urban patterns and meanings through history*. Bulfinch.

Rissanen J, 1984, Universal coding, information, prediction and estimation. *IEEE Trans. Information Theory*, IT-30(4): 629-636.

Tsuchiya H, Itoh S and Hashimoto T, 1996, An algorithm for designing a pattern classifier by using MDL criterion. *IEICE Trans. Fundamentals*, E79-A(6): 910-920.