Echelon Analysis for Spatial Structure of Data

Koji Kurihara
Okayama University, Department of Environmental and Mathematical Sciences
Okayama 700-8530, Japan, kurihara@ems.okayama-u.ac.jp

Wayne L. Myers
Environmental Resources Research Institute, The Pennsylvania State University
University Park, PA 16802, USA, wlm@psu.edu

Ganapati P. Patil
Center for Statistical Ecology and Environmental Statistics, The Pennsylvania State University
University Park, PA 16802, USA, gpp@stat.psu.edu

1. Introduction

Statistical analyses for spatial data are important features at the social and environmental sciences. Geographic information system (GIS) provides powerful tools to study the spatial structure of such data with some explicitness. However, it is very difficult to describe the objective manners based on such software. The echelon approach (Myers et al., 1997) is used to present the topological structure of spatial data in the systematic and objective manners. The echelons are derived from the changes in topological connectivity with decreasing surface level of spatial data. The purpose of this paper is to investigate the structure of meshed spatial data and mapped data based on the echelon approach. In addition, echelon approach is applied to ordered categorical data.

2. Echelon analysis for meshed spatial data and mapped data

The echelon analysis is based on the areas of relative high and low values of response variables for mapped or spatial data. For one horizontal dimension case, the hypothetical set of hillforms is divided to the same structured areas shown in the left side of Figure 1. The horizontal line shows the position (x) of spatial data, and the vertical line shows the value (h) of response variable for specified horizontal position. Thus, the data are given by the set of (x, h). The structure of these hillforms is given by the following echelon dendrogram shown in the right side of Figure 1.

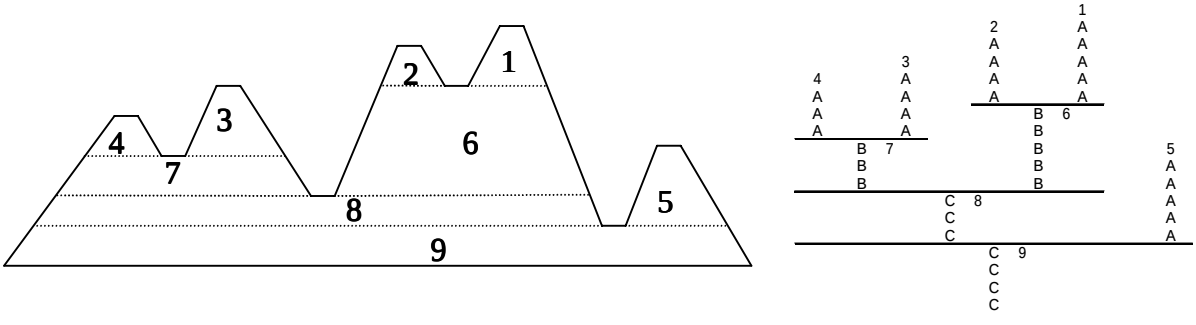

Figure 1. The hypothetical set of hillforms in one horizontal dimension and its echelon dendrogram.

The spatial data like remote sensing data are given as the pixels of digital values (h) over the

n-by–*m* array area $D_{ij}= \{(x, y)| x_{i-1} < x < x_i , y_{j-1} < y < y_j \}$, $i=1,2,\ldots,n$, $j=1,2,\ldots,m$. The spatial data for population data are obtained as the numbers of population (h) over the meshed area $D_{ij.}$ Thus, the form of these spatial data depends on (i, j, h). In general, the function $h=f(i, j)$ is not so simple and the structure for digital values is fairly complicated. Nevertheless, we can easily make the echelon dendrogram to study the structure of such data. For mapped data, the echelon approach is applied by making connective information table for the regional areas of interest.

3. Echelon analysis for ordered categorical data

Table 1 shows the contingency table for the numbers of death due to a traffic accident at Okayama prefecture in Japan. Note that some values are changed. Data are classified by the variables of age and situation. These classifications are ordered by the measure of the age and speed. Thus the values in this table are considered as spatially referenced data. In this situation, the echelon analysis is done by the following steps for the values over an 8-by-5 array.

Step1) Find the peaks: The value in the peak is larger than the value which is connected to the peak excluding the value in the same peak. There are four peaks in this 8-by-5 array. The first peak consists of the values of 18, 17 and 8. In the same procedure, we can find four peaks in this table.

Step2) Find the foundations of the peaks and foundations: The maximum value excluding the values of four peaks is 7. The value of 7 is a member of foundation between the peaks 1 and 2. With similar procedure, we can find other foundations. As a result, we can find the structure of this spatial data by the echelon dendrogram of Figure 2.

Table 1*. *The numbers of death by a traffic accident.

age	four-wheeled vehicle	two-wheeled vehicle		bicycle	pedestrian
		motorcycle	motor bike		
70-	2	2	5	8	14
60-69	12	3	3	11	12
50-59	17	0	2	2	7
40-49	7	1	1	1	2
30-39	18	0	1	2	2
20-29	17	4	1	1	2
10-19	8	4	4	1	0
-9	1	1	1	0	5

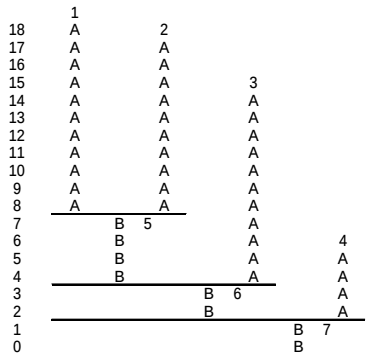

Figure 2. ***Echelon dendrogram for traffic accident data.***

REFERRENCES

Kurihara, K., Myers, W. L. and Patil, G. P. (2000). Echelon analysis of the relationship between population and land cover patterns based on remote sensing data. Community Ecology, 1, 103-122.

Myers, W. L., Patil, G. P. and Joly, K. (1997). Echelon approach to areas of concern in synoptic regional monitoring. Environmental and Ecological Statistics, 4, 131-152.

RESUME

L'approche d'échelon fournit la structure topologique des données spatical de la façon systématique et objective. Les échelons sont dérivés des changements de la connectivité topologique avec le lecel extérieur décroissant des données spatiales. Le but de cet article est au montrer l'approche d'échelon pour des données spatiales et des données catégoriques commandées.