Table cartogram generation as an optimization problem

Mao Li [a], Ryo Inoue [a, *]

[a] Graduate School of Information Sciences, Tohoku University, Sendai, Japan. li.mao.t6@dc.tohoku.ac.jp, rinoue@tohoku.ac.jp
* Corresponding author

Abstract:

A table cartogram, visualization of table-form data, is a rectangle-shaped table in which each cell is transformed to express the magnitude of positive weight by its area while maintaining the adjacency relationship of cells in the original table. Winter (2011) applies an area cartogram generation method of Gastner and Newman (2004) for their generation, and Evans et al. (2018) proposes a new geometric procedure. The rows and columns on a table cartogram should be easily recognized by readers, however, no methods have focused to enhance the readability. This study proposes a method that defines table cartogram generation as an optimization problem and attempts to minimize vertical and horizontal deformation. Since the original tables are comprised of regular quadrangles, this study uses quadrangles to express cells in a table cartogram and fixes the outer border to attempt to retain the shape of a standard table.

This study proposes a two-step approach for table cartogram generation with cells that begin as squares and with fixed outer table borders. The first step only adjusts the vertical and horizontal borders of cells to express weights to the greatest possible degree. All cells maintain their rectangular shape after this step, although the limited degree of freedom of this operation results in low data representation accuracy. The second step adapts the cells of the low-accuracy table cartogram to accurately fit area to weight by relaxing the constraints on the directions of borders of cells. This study utilizes an area cartogram generation method proposed by Inoue and Shimizu (2006), which defines area cartogram generation as an optimization problem. The formulation with vertex coordinate parameters consists of an objective function that minimizes the difference between the given data and size of each cell, and a regularization term that controls the changes of bearing angles. It is formulated as non-linear least squares, and is solved through the iteration of linear least squares by linearizing the problem at the coordinates of vertices and updating the estimated coordinates until the value of the objective function becomes small enough.

The first step is formulated as follows. Let N_v denote the number of vertical lines, N_h denote the number of horizontal lines, including the outer borders of the table, and x_i and y_j denote the coordinates of the vertical line i and horizontal line j, respectively. The lower-left corner of table (x_1, y_1) is placed at the origin $(0, 0)$ and D_{ij} denotes the weight of the cell enclosed by vertical lines i and $i+1$ and horizontal lines j and $j+1$. The total size of the table cartogram is given as a rectangle with a horizontal border of length $(N_v - 1) \times l$ and vertical border of length $(N_h - 1) \times l$, if each cell in the initial table is a square with sides of l. The value of l is decided to fit the total area to the total sum of the data. The optimization problem to obtain the positions of the horizontal and vertical lines that minimize the difference between the size and the given weight of each cell is represented by

$$\min_{\substack{x_2,\dots,x_{N_v-1} \\ y_2,\dots,y_{N_h-1}}} \sum_{i=1}^{N_v-1} \sum_{j=1}^{N_h-1} \left[(x_{i+1} - x_i)(y_{j+1} - y_j) - D_{ij} \right]^2 \ \text{s. t.} \ x_{i+1} - x_i > 0 \,, y_{j+1} - y_j > 0. \tag{1}$$

A low-accuracy table cartogram with rectangular cells is obtained by solving the non-linear least squares of Equation 1. This study solves the equation by linearizing it at the approximate coordinate values of the horizontal and vertical lines. First, the heights of rows are set according to the proportion of the weight given to each row over the total weight, and approximate y-coordinate values of horizontal lines (y_js) are set. Next, x-coordinates of vertical lines (x_is) are updated by solving the linear least squares by fixing y_j in Equation 1. Values of y_j are then updated by fixing x_is. The solution of Equation 1 can be obtained by repeating this procedure.

The second step deforms the shapes of cells in order to reduce the differences between their sizes and given weights, utilizing the continuous area cartogram construction method by Inoue and Shimizu (2006). Since outer borders should be fixed in table cartogram construction, unlike continuous area cartogram construction, the x-coordinates of vertices on the horizontal outer borders and y-coordinates of vertices on the vertical outer borders are not set as parameters in this study. Changes of border bearing angles are minimized to reduce unnecessary deformation of table cartogram shape to keep the directions of edges in table cartograms to their original horizontal or vertical directions. We suggest not updating the original bearing angles in the regularization term during iteration.

The proposed method is applied to the GDPs of countries and regions affected by the 1997 Asian financial crisis. The data set is published by the World Bank. Figure 1 shows the table cartograms created by the proposed and Inoue and

Shimizu methods, both modified to construct cartograms with fixed borders, and Table 1 indicates their calculation time and quantitative indices on representation accuracy and shape deformation.

The table cartogram created by the proposed method (Figure 1 (a)) has borders closer to horizontal and vertical than the one using the Inoue and Shimizu method (Figure 1 (b)). The evaluation indices in Table 1 also indicate that the table cartogram deformation of the proposed method is smaller than that of the previous method. The differences between bearing angles of borders in the transformed and original tables are evaluated by the RMSEs (root mean-square errors), in radians. Smaller values indicate less horizontal and vertical deviation. The proposed method produces less deviation than the Inoue and Shimizu method.

In summary, this study formulates table cartogram generation as an optimization problem. The proposed method is designed to produce table cartograms with smaller degrees of deformation by keeping borders close to horizontal and vertical. The proposed method uses a two-step approach; the first step optimizes the positions of vertical and horizontal borders and the second step transforms the cells with minimal changes of bearing angles. The GDP dataset application confirms that the proposed method achieves less deformation and small data representation error.

(a)

year \ country	1994	1995	1996	1997	1998	1999	2000
Indonesia	1.77	2.02	2.27	2.16	0.95	1.40	1.65
Malaysia	0.74	0.89	1.01	1.00	0.72	0.79	0.94
Philippines	0.64	0.74	0.83	0.82	0.72	0.83	0.81
Singapore	0.74	0.88	0.96	1.00	0.86	0.86	0.96
Thailand	1.47	1.69	1.83	1.50	1.14	1.27	1.26
Hong Kong	1.36	1.45	1.60	1.77	1.69	1.66	1.72
South Korea	4.56	5.56	5.98	5.58	3.74	4.85	5.62

(b)

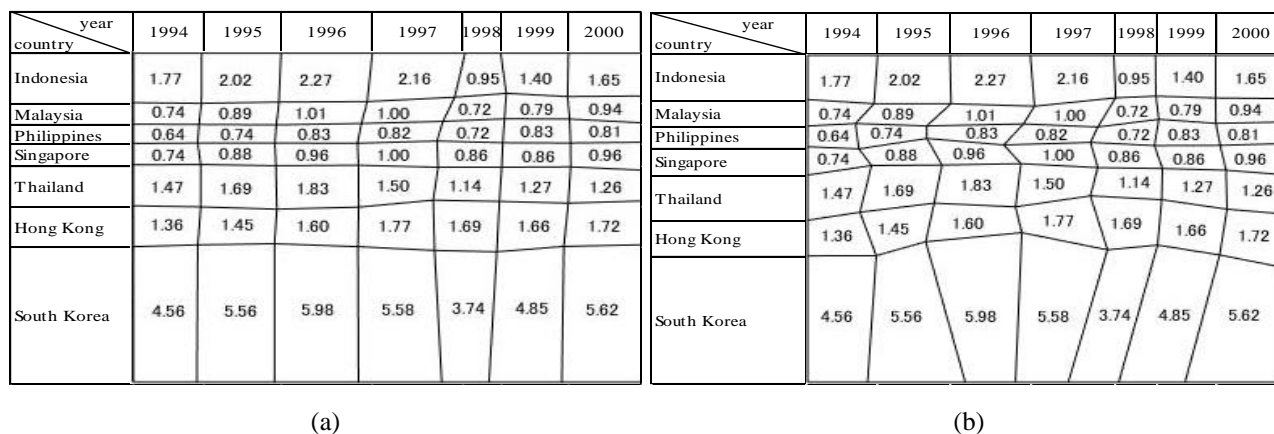| year \ country	1994	1995	1996	1997	1998	1999	2000
Indonesia	1.77	2.02	2.27	2.16	0.95	1.40	1.65
Malaysia	0.74	0.89	1.01	1.00	0.72	0.79	0.94
Philippines	0.64	0.74	0.83	0.82	0.72	0.83	0.81
Singapore	0.74	0.88	0.96	1.00	0.86	0.86	0.96
Thailand	1.47	1.69	1.83	1.50	1.14	1.27	1.26
Hong Kong	1.36	1.45	1.60	1.77	1.69	1.66	1.72
South Korea	4.56	5.56	5.98	5.58	3.74	4.85	5.62

Figure 1. Table cartograms of GDP (10 billion current USD) of affected countries and regions in the 1997 Asian financial crisis using (a) the proposed method and (b) the Inoue and Shimizu (2006) method. Inoue and Shimizu (2006) method are modified to create fixed outer border cartograms.

Methods	RMSEs between GDP of countries and size of cells (10 billion current USD)	RMSEs of bearing angles of edges between transformed and original tables (in rad)	Calculation time (s)*
The proposed method	2.19e-04	0.0583	7.44
Inoue and Shimizu (2006)	3.97e-04	0.1789	9.64

* These methods are implemented in MATLAB and tested on a computer with 2.93-GHz CPU.

Table 1. Data representation precision, table deformation, and calculation time using the two methods.

References:

Evans, W., Felsner, S., Kaufmann, M., Kobourov, S. G., Mondal, D., Nishat, R. I., and Verbeek, K. 2018. Table cartogram. Computational Geometry, 68: 174–185.

Gastner, M. T. and Newman, M. E. J. 2004. Diffusion-based method for producing density equalizing maps. Proceedings of the National Academy of Sciences of the United States of America, 101(20): 7499–7504.

Inoue, R. and Shimizu, E. 2006. A new algorithm for continuous area cartogram construction with triangulation of regions and restriction on bearing changes of edges. Cartography and Geographic Information Science, 33(2): 115–125.

Winter, M. J. 2011. Diffusion cartogram for the display of periodic table data. Journal of Chemical Education, 88(11): 1507–1510.