Title
A Multistart Heuristic Approach to Spatial Aggregation Problems

Journal
International Conference on GIScience Short Paper Proceedings, 1(1)

Authors
Xiao, Ningchuan
Jiang, Peixuan
Kim, Myung Jin
et al.

Publication Date
2016

DOI
10.21433/B3111650v4z6

Peer reviewed

A Multistart Heuristic Approach to Spatial Aggregation Problems

Ningchuan Xiao[1], Peixuan Jiang[2], Myung Jin Kim[3], Anuj Gadhave [4]

[1] Department of Geography, Ohio State University, Columbus, OH 43210
Email: xiao.37@osu.edu
[2] Esri, Redlands, CA, 92373
[3] Gyeonggi Institute of Science and Technology Promotion, Suwon, Gyeonggi, Korea
[4] Department of Computer Science and Engineering, Ohio State University, Columbus, OH 43210

Abstract

In this paper, we present a heuristic method that can be used to search for a diverse set of solutions to spatial aggregation problems. The algorithm is developed using a multistart strategy. Computational experiments are conducted to test the effectiveness of the algorithm.

Spatial data are often aggregated for different purposes. The census data in the United States, for example, are aggregated into levels such as blocks, block groups, and tracts. In another example, political redistricting requires the aggregation of spatial units into districts such that some objectives can be optimized. Though aggregation is a common exercise in the use of spatial data, it has been noted that many of the aggregations are arbitrary and may not provide effective spatial units for applications (Martin, 1998; Cockings and Martin, 2005), which often leads to the modifiable areal unit problem (Openshaw, 1983). Equally important is the multiplicity of spatial aggregations: given an aggregation of spatial units, many equivalent schemes may also exist. For example, there often exist many perfect political redistricting plans when population equality is the only objective (Kim and Xiao, 2016). Subsequently, it is important to explore a diverse set of aggregation schemes in order to fully understand the complexity of aggregation. Researchers have developed a wide range of methods that can be used to solve aggregation problems (Openshaw and Rao, 1995; Xiao, 2008). However these methods are designed for specific purposes and they generally do not aim to explore the complexity of aggregation.

The purpose of this paper is to develop a heuristic method that can be used to find, not one, but a diverse set of high quality solutions to an aggregation problem. This method first uses a search algorithm to find a set of good solutions. These solutions are stored in a pool and another algorithm is developed to improve the solutions in the pool by recombining them into new solutions. This method is heuristic, meaning it cannot guarantee optimal solutions be found, and we test its effectiveness using a set of benchmark problems.

We have tested the multistart algorithm with a wide range of data. Due to the page limit of this abstract, we discuss the first data set of Iowa congressional redistricting for year 2000. The Iowa Constitution dictates that the counties shall not be split for political redistricting purposes, which make it a problem of aggregating 99 spatial units into 5 regions. The official 2000 plan has an objective function value of 0.0080, and the literature has documented a number of redistricting

plans that can be considered as the benchmark for other algorithms: 0.0045 (Xiao, 2008), 0.0066 (Kim, 2011), 0.0011 (Guo and Jin, 2011), and 0.00079 (Kim, 2011).

Figure 1 includes the solutions found in a run with a pool size of 50, where the upper-left map shows the best solution found with an objective function value of 0.00017, which is better than the best solution in the literature, and the lower-right map the worst in the pool (with an objective function value of 0.0037, which is still an improvement over the official plan).

The multistart heuristic method presented in this paper is shown to be effective in generating aggregated regions with an objective of equal total weight (population) among the regions. We stress that the method is designed specifically to explore the multiplicity of spatial aggregation and the method can be used to generate a diverse set of aggregations. Though we only considered spatial contiguity and weight equality in the current form of this method, we believe it is possible to incorporate other constraints and objectives in this framework. Data sets of practical sizes (in hundreds of spatial units) were used to test this method. One of the future directions is to utilize high performance computing techniques to further speed up so that large data sets with thousands of spatial units can be practically processed.

References

Cockings, S. and D. Martin (2005). Zone design for environment and health studies using pre-aggregated data. *Social science & medicine 60*(12), 2729–2742.

Guo, D. and H. Jin (2011). iredistrict: Geovisual analytics for redistricting optimization. *Journal of Visual Languages & Computing 22*(4), 279–289.

Kim, M. and N. Xiao (2016). Contiguity-based optimization models for political redistricting problems. *International Journal of Applied Geospatial Research*, accepted.

Kim, M. J. (2011). *Optimization Approaches to Political Redistricting Problems*. Ph. D. thesis, The Ohio State University, Columbus, OH.

Martin, D. (1998). Optimizing census geography: the separation of collection and output geographies. *International Journal of Geographical Information Systems 12*(7), 673–685.

Openshaw, S. (1983). *The Modifiable Areal Unit Problem*. Norwich: Geo Books.

Openshaw, S. and L. Rao (1995). Algorithms for reengineering 1991 census geography. *Environment and Planning A 27*, 425–446.

Xiao, N. (2008). A unified conceptual framework for geographical optimization using evolutionary algorithms. *Annals of the Association of American Geographers 98*(4), 795–817.

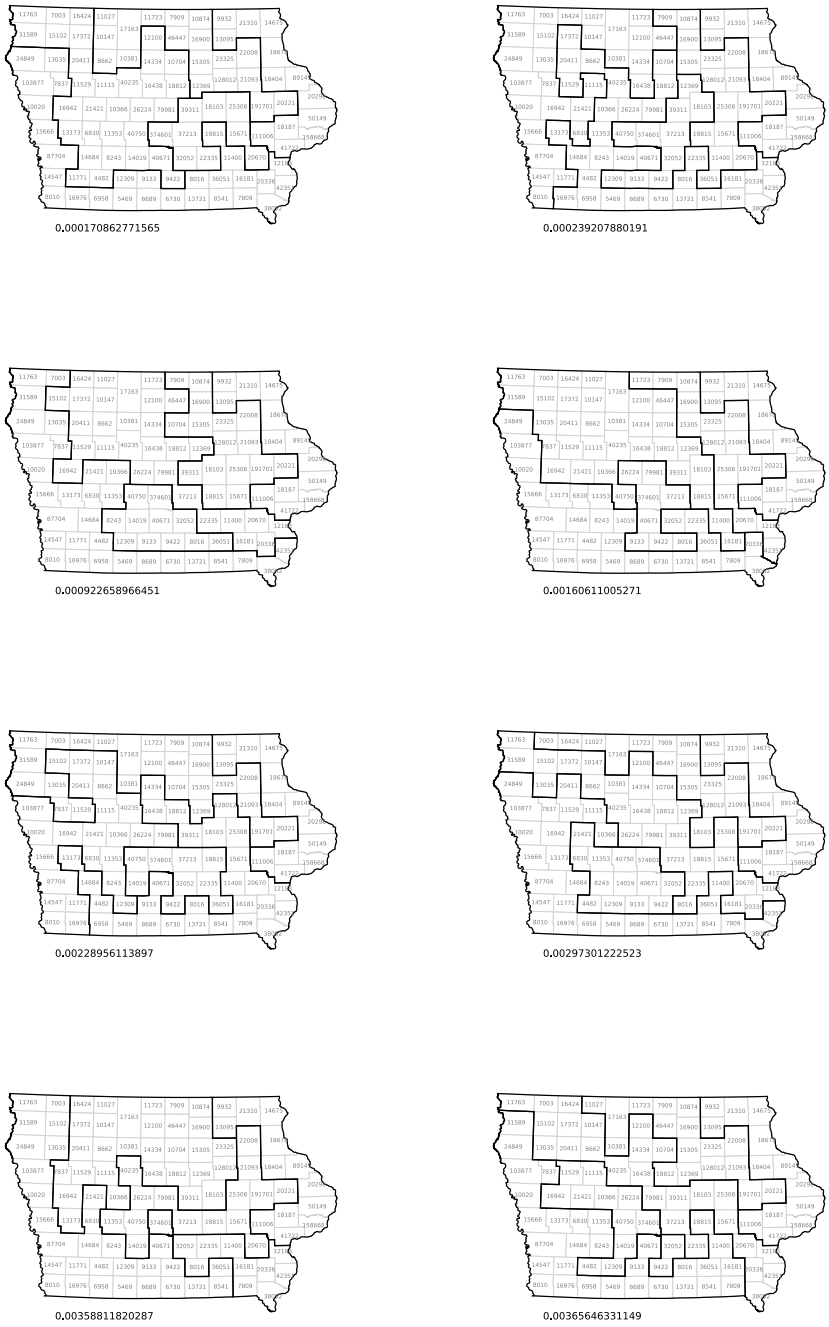

Figure 1: Political redistricting plans for Iowa (using 2000 census data).