

MapView - A Visualization Tool for Spatial Outlier Detection

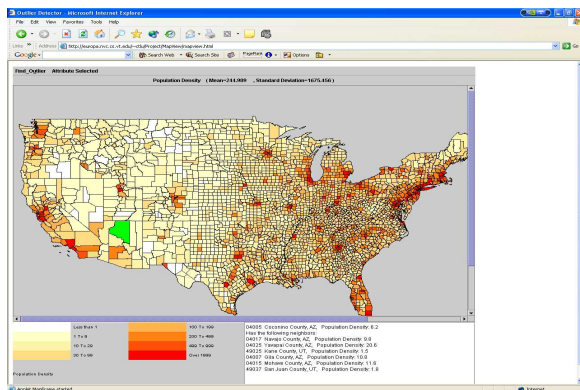
Chang-Tien Lu, Yufeng Kou, Hongjun Wang
Department of Computer Science
Virginia Polytechnic Institute and State University

Dechang Chen
Preventive Medicine and Biometrics
Uniformed Services University

1. INTRODUCTION

Spatial outliers are spatially referenced objects whose non-spatial attribute values are significantly different from those of other spatially referenced objects in their spatial neighborhoods. Informally, a spatial outlier is a local instability, or an extreme observation with respect to its neighboring values. MapView is a web-based spatial analytical software, which is designed to facilitate the observation and discovery of spatial outliers for the US census data. MapView supports the visualization of 11 different census attributes and provides the functionality of detecting local abnormality using various spatial outlier detection algorithms, such as Scatterplot, Moran Scatterplot, Z-value, Iterative Z-value, and Iterative Ratio.

MapView can efficiently and effectively discover spatial outlier counties and mark them with distinguishable color. In addition, users can click each county to see its corresponding attribute value and those of its neighboring counties. To overcome the deficiencies of the existent spatial outlier detection algorithms, we proposed and developed two new algorithms, iterative Z-value and iterative ratio, which can detect true outliers ignored by the existing algorithms and remove falsely detected spatial outliers.



2. SOFTWARE ARCHITECTURE

The Mapview system has a three-tier architecture, including Graphic User Interface (GUI), outlier detection algorithms, and data files. The GUI draws a US map using the geographical coordination information of each county. The outlier detection algorithms receive user query from the GUI, compute the spatial outliers from database files, and send the results back to the GUI for display. There are three data files, including polygon, county attribute, and neighborhood relationship.

Supported Spatial Outlier Detection Algorithms

- **Scatterplot** is a graph based outlier detection method. It shows attribute values on X-axis and the average of the attribute values in the neighborhood on the Y-axis [2]. A least square regression line is used to identify outliers. Nodes far away from the regression line are flagged as spatial outliers.
- **Moran Scatterplot** is a plot of normalized attribute value against the neighborhood average of normalized attribute values. It contains four quadrants. The upper left and the lower right quadrants indicate a spatial association of dissimilar values: low values surrounded by high value neighbors and high values surrounded by low value neighbors. Spatial outliers can be identified from these two quadrants.
- **Z-value** approach calculates the standardized difference between the attribute value of a point and the average attribute value of its neighbors. Those points with the standardized difference values greater than a pre-defined threshold will be flagged as spatial outliers.
- **Iterative Z-value** approach is the proposed new algorithm [1]. The key idea of iterative approach is to detect spatial outliers one by one. After one outlier is detected, its attribute value will be substituted with the average attribute value of its neighbors before next iteration begins.
- **Iterative Ratio** approach is similar to iterative Z-value approach, whereas it identifies outliers through the ratio between the attribute value of a point and the average attribute value of its neighbors [1]. Those points with the ratios (or the inverse of ratios) greater than a pre-defined threshold will be flagged as spatial outliers.

Access the MapView demo version

<http://europa.nvc.cs.vt.edu/~ctl/Project/MapView/index.htm>

3. REFERENCES

- [1] C. Lu, D. Chen, and Y. Kou. Algorithms for Spatial Outlier Detection. In *IEEE International Conference on Data Mining*, Nov 2003.
- [2] S. Shekhar, C. Lu, and P. Zhang. Detecting Graph-Based Spatial Outlier: Algorithms and Applications(A Summary of Results). In *Proc. of the Seventh ACM-SIGKDD Int'l Conference on Knowledge Discovery and Data Mining*, Aug 2001.