DETECTION OF LOCAL AND GLOBAL OUTLIERS IN SPATIAL DATA

M. Ernst, G. Haesbroeck

University of Liege, Department of Mathematics, Liege, Belgium

m.ernst@ulg.ac.be; g.haesbroeck@ulg.ac.be

Spatial data are characterized by statistical units, with known geographical positions, on which non spatial attributes are measured. Spatial data are frequently observed in many fields (geography, social sciences,...) and may contain two types of atypical observations: global and/or local outliers (see [2] for more details).
The attribute values of a global outlier are outlying with respect to the values taken by the majority of the data points while the attribute values of a local outlier are extreme when compared to those of its neighbors.
Classical outlier detection techniques may be used to find global outliers as the geographical positions of the data is not taken into account in this search. The detection of local outliers is more complex especially when there are more than one non spatial attribute. In this talk, some existing techniques for detecting local outliers in multivariate spatial data will be outlined and two new procedures will be defined. The different procedures will be illustrated and compared on real and simulated examples.
Among the existing techniques, focus will be on the proposals of Filzmoser et al. [1] and Lu et al. [3]. These two approaches are based on Mahalanobis or robust distances computed in each neighborhood using a common estimation of the covariance matrix.
Plugging a common and global estimation of the covariance matrix in the local distances is justified by the fact that the sample size might be small with respect to the number of variables when the data are restricted to the neighborhood. However, imposing the same structure to each neighborhood does not seem to be the most appropriate approach. A natural way to overcome this limitation is to estimate the covariance matrix using a regularized estimator. This yields the first new proposal for detecting local outliers. The second proposal measures outlyingness using depths and finds local outliers by detecting those observations whose majority of neighbors are lying in "far" depth contours.

References

[1] Filzmoser, P., Ruiz-Gazen, A and Thomas-Agnan, C., Identification of local multivariate outliers, (2012), *submitted*.

[2] Haslett, J., Brandley, R., Craig, P., Unwin, A. et Wills, G., Dynamic Graphics for Exploring Spatial Data With Applications to Locating Global and Local Anomalies, *The American Statistician*, **45**, (1991), 234-242.

[3] Lu, C.T., Chen, D. and Kou, Y., Multivariate spatial outlier detection, *International Journal on Artificial Intelligence Tools*, **13**, (2004), 801–812.