Delineating Spatial Clusters with Artificial Neural Networks

Gladston J. P. Moreira, Ricardo H.C. Takahashi, Luiz Duczmal
Universidade Federal de Minas Gerais, Brazil

OBJECTIVE

Multiple or irregularly shaped spatial clusters are often found in disease or syndromic surveillance maps. We develop a novel method to delineate the contours of spatial clusters, especially when there is not a clearly dominating primary cluster, through artificial neural networks. The method may be applied either for maps divided into regions or point data set maps.

BACKGROUND

The solution of computationally hard problems with artificial neural networks normally demands the use of a multi-layer network structure. The multi-layer topology processes the data through consecutive steps, for each one of the layers. It has been shown in the literature that feed-forward neural networks, such as Multi-Layer Perceptrons (MLP), are universal multi-variable function approximators [1]. The spatial scan statistic [2] is the usual measure of the strength of a cluster.

METHODS

A MLP is trained as a function that approximates the spatial scan statistic for the whole map domain. The solution delineates regions of the map belonging to distinct level curves values, and those constitute the estimates of the primary and the several possible secondary clusters. For maps divided into *m* regions the scan statistic is evaluated for each region taken individually, where each region is identified by the geographic coordinates of its centroid. We start defining a MLP artificial neural network with training set size *m*. The geographic coordinates and the scan values are, respectively, the net input and the desired output. A MLP with topology 2-*n*-1 is trained by the Levenberg-Marquardt backpropagation [2,3]. Following the training phase, the scan function evaluation is extended for the whole domain of geographic coordinates. The areas lying inside the level curves above a certain threshold value are thus considered the most likely clusters. The balance between bias (rigidity) and variance (flexibility) of the net is obtained by means of its sizing [4]. Larger topologies structures allow more flexibility but less bias. A similar approach is used for point data set maps.

RESULTS

Figures 1 and 2 display the scan function level curves for breast cancer clusters in the Northeast United States in 1988-1992 [5]. A topology 2-50-1 was used in Figure 1, and a topology 2-100-1 was used in Figure 2. We obtain smoother clusters in Figure 1, when less variance and more bias are allowed. In Figure 2 we obtain otherwise sharply delineated, more irregularly shaped clusters, indicating less bias, due to the larger variance allowed.

CONCLUSIONS

The artificial neural networks method is a fast and flexible algorithm for spatial cluster delineation. The irregularity of the cluster shapes can be adjusted varying the parameters of the neural network.

Figure 1 – Smooth clusters: low variance, high bias.

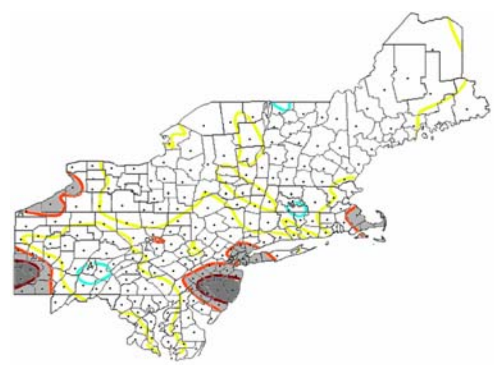

Figure 2 – Sharp clusters: high variance, low bias.

REFERENCES

[1] Kulldorff M, 1997. A Spatial Scan Statistic, *Comm. Statist. Theory Meth.*, 26(6), 1481-1496.

[2] M. H. Fun and M.T. Hagan, 1996. Levenberg-Marquardt training for modular networks. *In Proceedings of the IEEE International Conference on Neural Networks*, 1, 468-473.

[3] D. E. Rumelhart, G. E. Hinton, and R. J. Williams,1986. Learning representations by back-propagating errors. *Nature***,** 323:533-536.

[4] Teixeira, R.A., Braga, A.P., Takahashi, R.H.C., Saldanha, R.R., 2000. Improving generalization of MLPs with multi-objetive optimization. *Neurocomputing*, v. 35, p. 189-194.

[5]Kulldorff M., Feuer E.J., Miller B.A., Freedman L.S., 1997. Breast cancer clusters in the Northeast United States: a geographic analysis. *American Journal of Epidemiology*, 146:161-170.

Corresponding author: duczmal@est.ufmg.br