COACTION PUBLISHING

Spatial cluster detection in case-control datasets with the Autonomous Leaves Graph

Nara Moreira, Denise Burgarelli, Luiz Duczmal* and Flávia C. O. Magalhães

Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil

Objective

To detect multiple significant spatial clusters of disease in case-control point event data using the Autonomous Leaves Graph (ALG) and the spatial scan statistic.

Introduction

Ordering-based approaches (1,2) and quadtrees (3) have been introduced recently to detect multiple spatial clusters in point event datasets. The ALG (4) is an efficient graph-based data structure to handle the communication of cells in discrete domains. This adaptive data structure was favorably compared to common tree-based data structures (quadtrees). An additional feature of the ALG data structure is the total ordering of the component cells through a modified adaptive Hilbert curve, which links sequentially the cells (the orange curve in the example of Fig. 1).

Methods

We combine ordering-based approaches with the ALG structure to identify multiple clusters in case-control datasets, in a five-step procedure. In the first step, we subdivide adaptively the domain into square cells (blue squares in Fig. 1), with controls (pink points) and cases (little dotted black squares). In the second step, the cell's ordering given by the Hilbert curve is used to sequentially join cells with the highest proportion of cases over controls. This produces loose groups of higher than average rates of disease. The adjacent groups thus formed are themselves united, according to certain criteria, into larger groups in the third step, forming the cluster candidates. In the fourth step, Kulldorff's spatial scan statistic is computed for each group, and the clusters are ranked. Finally the cases are randomized, and steps 1–4 are repeated. That is done hundreds of times to compute the significance of each cluster. Only significant clusters are reported.

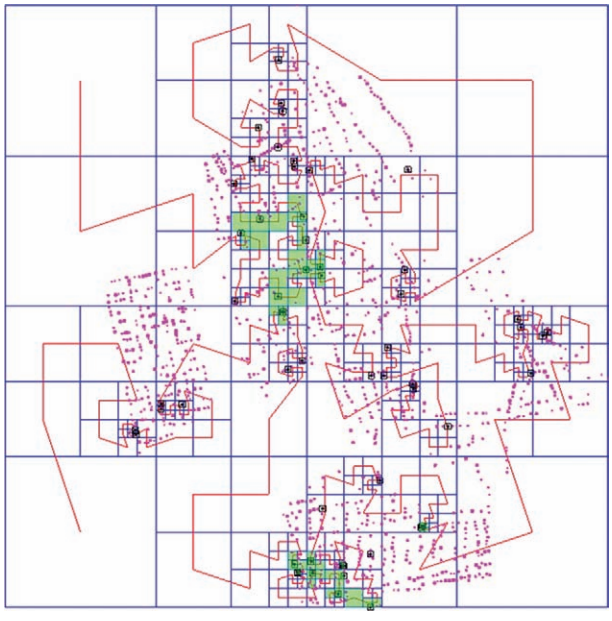

Fig. 1. The three most likely clusters of dengue fever.

Results

As an application, we find clusters of dengue fever for Lassance City, in southeast Brazil, 2010 (5). Fig. 1 shows the three significant clusters found, displayed as the three green patches.

Conclusions

The previous adaptive subdivision of the domain is essential to define more homogeneous regions within the study area.

Also, instead of just applying some ordering-based approach, our method introduces an intermediate step (step 3) to combine the separated regions of high incidence. Those two features produce more reliable clusters, compared with the usual ordering-based methods.

Keywords

Spatial cluster; autonomous leaves graph; ordering-based; spatial scan statistic; quadtree

Acknowledgments

We thank CNPq and Fapemig.

References

1. Demattei C, Molinari N, Daures J. Arbitrarily shaped multiple spatial cluster detection for case event data. Comput Stat Data Anal. 2007;51:3931–45.
2. Demattei C, Cucala L. Multiple spatio-temporal cluster detection for case event data: an ordering-based approach. Comm Stat Theor Meth. 2011;40:358–72.
3. Yiannakoulias N, Karosas A, Schop DP, Svenson LW, Hodgson JM. Using Quad trees to generate grid points for applications in geographic disease surveillance. Adv Dis Surveill. 2007;3:2.
4. Burgarelli D, Kishinhevsky M, Biezuner RJ. A new adaptive mesh refinement strategy for numerically solving evolutionary PDE's. J Comput Appl Math. 2006;196:115–31.
5. Duczmal LH, Moreira GJP, Burgarelli D, Takahashi RHC, Magalhães FCO. Voronoi distance based prospective space-time scans for point data sets: a dengue fever cluster analysis in a southeast Brazilian town. Int J Health Geograph. 2011;10:29.

***Luiz Duczmal**
E-mail: duczmal@ufmg.br

Citation: Emerging Health Threats Journal 2011, **4**: 11197 - DOI: 10.3402/ehtj.v4i0.11197