Spatial cluster detection through constrained dynamic programming

Gladston Prates Moreira[1], Luiz Duczmal[2]*, Luis Paquete[3] and Ricardo H. C. Takahashi[2]

[1]Universidade Federal de Ouro Preto, Ouro Preto, MG, Brazil; [2]Universidade Federal de Minas Gerais, Belo Horizonte, Brazil; [3]Universidade de Coimbra, Coimbra, Portugal

Objective

We propose a fast, exact algorithm to make detection and inference of arbitrarily shaped connected spatial clusters in aggregated area maps based on constrained dynamic programming.

Introduction

The spatial scan statistic (1) is the most used measure for cluster strenght. The evaluation of all possible subsets of regions in a large dataset is computationally infeasible.

Many heuristics have appeared recently to compute approximate values that maximizes the logarithm of the likelihood ratio. The Fast Subset Scan (2) finds exactly the optimal irregularly spatial cluster; however, the solution may not be connected.

The spatial cluster detection problem was formulated as the classic knapsack problem (3) and modeled as a biobjective unconstrained combinatorial optimization problem.

Dynamic programming relies on the principle that, in an optimal sequence of decisions or choices, each subsequence must also be optimal. During the search for a solution, it avoids full enumeration by pruning early partial decision solutions that cannot possibly lead to optimal solutions.

Methods

We propose a novel method, the Geographical Dynamic Scan (GDScan) to find optimal connected clusters. It employs an adaptation of the Nemhauser–Ullman algorithm for the 0–1 knapsack problem (4). We minimize a biobjective vector function $F(z) = (-C(z), N(z))$, where $C(z)$ and $N(z)$ are the number of cases and the population of the candidate cluster z, respectively. Then, we show that the solution which maximizes the spatial scan statistic is included in the set of nondominated solutions of F (the Pareto set), showing that the dynamic programming algorithm allows to solve the unconstrained maximization of the scan statistic for any given spatial dataset. However, this is typically not sufficient to solve practical spatial detection problems.

The dynamic programming algorithm is thus modified to consider (i) a geographical proximity constraint and (ii) a connectivity constraint, for each region j. Assuming that the geographical proximity of a region j contains k regions, the geographical dynamic scan guarantees the optimal solution within the collection of 2ˆk subsets, searching for only a small number of subsets, which depends almost linearly on k, on average.

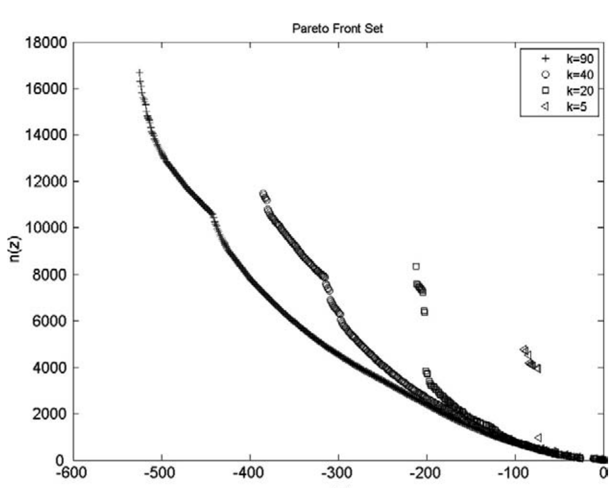

Fig. 1. The nondominated sets of cluster solutions.

Results

We conducted numerical simulations showing that GDScan has good power of detection, sensitivity and positive predicted value.

An application is shown for the dataset of Chagas' disease cases in the population at risk of puerperal women in Minas Gerais state, Brazil, in 2006 (5). Fig. 1 shows the nondominated sets of solutions obtained by GDScan for neighborhood sizes of 5, 20, 40 and 90.

Conclusions

GDScan is a fast and an efficient method to detect connected arbitrarily shaped disease clusters in aggregated area maps.

Keywords

Disease cluster; dynamic programming; spatial scan statistic

Acknowledgments

The authors thank the agencies CNPq, Fapemig and Capes.

References

1. Kulldorff M. A spatial scan statistic. Commun Stat Theor Meth. 1997;26:1481–96.
2. Neill DB. Fast and flexible outbreak detection by linear-time subset scanning. Adv Dis Surveill. 2008;5:48.
3. Cançado ALF. Spatial clusters detection through multiobjective optimization (in portuguese). Ph.D. thesis, Universidade Federal de Minas Gerais, Brazil; 2009.
4. Nemhauser GL, Ullmann Z. Discrete dynamic programming and capital allocation. Manage Sci. 1969;15:494–505.
5. Cançado ALF, Duarte AR, Duczmal L, Ferreira SJ, Fonseca CM, Gontijo ECDM. Penalized likelihood and multi-objective spatial scans for the detection and inference of irregular clusters. Int J Health Geograph. 2010;9:55.

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

Citation: Emerging Health Threats Journal 2011, **4**: 11189 - DOI: 10.3402/ehtj.v4i0.11189