Optimizing simultaneously the geometry and the internal cohesion of clusters

Anderson R. Duarte, Luiz Duczmal, Sabino J. Ferreira, André L.F. Cançado
Universidade Federal de Minas Gerais

OBJECTIVE

We introduce a novel spatial scan algorithm for finding irregularly shaped disease clusters maximizing simultaneously two objectives: the regularity of shape and the internal cohesion of the cluster.

BACKGROUND

Irregularly shaped cluster finders frequently end up with a solution consisting of a large zone z spreading through the map, which is merely a collection of the highest valued regions, but not a geographically sound cluster. One way to amenize this problem is to introduce penalty functions to avoid the excessive freedom of shape of z. The compactness penalty K(z) [1] is a function used to reduce the scan value of irregularly shaped clusters, based on its geometric shape. Another penalty is the cohesion function C(z), a measure of the absence of weak links, or under-populated regions within the cluster which disconnect it when removed. It was mentioned in [1] that such weak links could be responsible for a diminished power of detection in cluster finder algorithms. Methods using those penalty functions present better performance. The geometric compactness is not entirely satisfactory, although, because it has a tendency to avoid potentially interesting irregularly shaped clusters, acting as a low-pass filter. The cohesion function penalty method, although, has slightly less specificity.

METHODS

We developed a novel method combining the compactness function and the cohesion function in a multi-objective optimization algorithm. Specifically, we maximize simultaneously LLR(z).K(z) and LLR(z).C(z), where LLR(z) is the spatial scan statistic [3]. The solution presented is a Pareto-set, consisting of all the clusters found which are not simultaneously worse in both objectives. The significance evaluation is conducted in parallel for all the clusters in the Pareto-set through a Monte Carlo simulation, determining the best cluster solution.

RESULTS

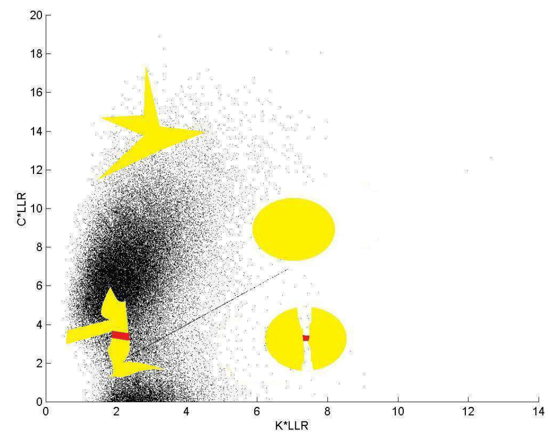

Figure 1 - Clusters which are at the same time irregularly shaped *and* weakly connected tend to stay close to the origin in this LLR(z).K(z) *versus* LLR(z).C(z) graph. Weakly connected, regularly shaped clusters stay at the lower right part, as opposed to strongly connected, irregularly shaped clusters, which stay at the upper left part.

We conduct numerical simulations showing that our method has comparatively good power of detection for those clusters which are not simultaneously irregularly shaped *and* have weak links.

CONCLUSIONS

Our method distinguishes clearly those clusters which are worse both from the geometric and the topological viewpoint (Figure 1). From a geographic perspective, those clusters are undesirable, and it makes sense to penalize them, and at the same time allowing those clusters which are strong in at least one of these two objectives.

REFERENCES

[1]Duczmal L, Kulldorff M, Huang L. Evaluation of spatial scan statistics for irregularly shaped disease clusters, *J. Comput. Graph. Stat.*, 2006;15, 428-442

[2]Duczmal L, Cançado ALF, Takahashi RHC. Geographic Delineation of Disease Clusters through Multi-Objective Optimization, *J. Comput. Graph. Stat.*, 2008; 17, 243-262

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

Corresponding author: duczmal@ufmg.br