Scalable detection of irregular disease clusters using soft compactness constraints

Skyler Speakman*, Edward McFowland III, Sriram Somanchi and Daniel B. Neill

Event and Pattern Detection Laboratory, Carnegie Mellon University, Pittsburgh, PA, USA

Objective

We present a new method for efficiently and accurately detecting irregularly shaped outbreaks by incorporating 'soft' constraints, rewarding spatial compactness and penalizing sparse regions.

Introduction

The spatial scan statistic (1) detects significant spatial clusters of disease by maximizing a likelihood ratio statistic $F(S)$ over a large set of spatial regions, typically constrained by shape. The fast localized scan (2) enables scalable detection of irregular clusters by searching over proximity-constrained subsets of locations, using the linear-time subset scanning (LTSS) property to efficiently search over all subsets of each location and its $k-1$ nearest neighbors. However, for a fixed neighborhood size k, each of the 2^k subsets are considered equally likely, and thus the fast localized scan does not take into account the spatial attributes of a subset. Hence, we wish to extend the fast localized scan by incorporating soft constraints, which give preference to spatially compact clusters while still considering all subsets within a given neighborhood.

Methods

For a given local neighborhood with center location s_c and size k, we place a bonus or penalty $\triangle_i = h(1 - 2d_i/r)$ on each location s_i, where d_i is that location's distance from the center, r is the neighborhood radius and h is a constant representing the strength of the compactness constraint. Each $\triangle_i$ can be interpreted as the prior log-odds that s_i will be affected, and thus the center location ($d_i = 0$, $\triangle_i = h$) is e^h times as likely as its $(k-1)$th nearest neighbor ($d_i = r$, $\triangle_i = -h$). We demonstrate that the penalized score function $F'(S) = F(S) + \Sigma s_i \in S\ \triangle_i$ can be efficiently maximized over all subsets S for each neighborhood. To do so, we show that $F(S)$ can be written as an additive function (sum over locations) conditioned on the relative risk in region S, and therefore $F'(S)$ is additive given the risk as well. We then jointly maximize $F'(S)$ over all subsets S and all values of the risk.

Results

The penalized subset scan was evaluated using emergency department (ED) data from 97 Allegheny County zip codes. We compared detection power and spatial accuracy, with and without compactness constraints, on synthetic, spatially localized outbreaks injected into the ED data. Our results show that including compactness constraints allows the penalized subset scan methods to detect outbreaks earlier and improve spatial accuracy, as compared to the unpenalized fast localized scan and circular scan, over a wide range of neighborhood sizes (Fig. 1). Without compactness constraints, the method averaged 7.6 days to detect and an overlap coefficient of 54.9% for neighborhood size of $k = 10$, but detection performance degraded rapidly for smaller or larger values of k. Performance of the compactness-constrained methods was less dependent on
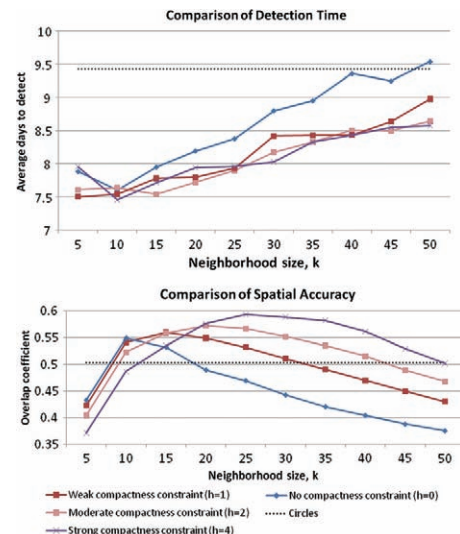

Fig. 1. Comparison of detection performance for methods with and without soft compactness constraints.

neighborhood size, requiring 7.4 days to detect and achieving an overlap coefficient of 59.3% for well-chosen parameter values.

Conclusions

Our results demonstrate that the incorporation of soft compactness constraints substantially improves the timeliness and accuracy of outbreak detection. Our new approach based on additive linear-time subset scanning enables efficient maximization of penalized scan statistics over subsets of the data, thus improving detection of irregular clusters.

Keywords

Outbreak detection; spatial scan; penalized subset scan

References

1. Kulldorff M. A spatial scan statistic. Commun Stat Theor Meth. 1997;26:1481–96.
2. Neill DB. Fast subset scan for spatial pattern detection. J R Stat Soc (Ser B). 2011;to appear.

*Skyler Speakman
E-mail: skylerspeakman@gmail.com