ISDS 2014 Conference Abstracts

StarScan: A Novel Scan Statistic for Irregularly-Shaped Spatial Clusters

Sriram Somanchi*, David Choi and Daniel B. Neill

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

Objective

We present StarScan, a novel scan statistic for accurately detecting irregularly-shaped disease outbreaks. StarScan maximizes a penalized log-likelihood ratio statistic, allowing the radius around a central location to vary as a function of the angle and applying a penalty proportional to the total change in radius.

Introduction

Kulldorff's spatial scan statistic[1] detects significant spatial clusters of disease by maximizing a likelihood ratio statistic over circular spatial regions. The fast localized subset scan[2] enables scalable detection of proximity-constrained subsets and increases power to detect irregularly-shaped clusters, However, unconstrained subset scanning within each circular neighborhood[2], may not necessarily capture the pattern of interest, and is too under-constrained for use with case/control point data. Thus we propose the star-shaped scan statistic (StarScan), a novel method that efficiently maximizes the log-likelihood ratio over irregularly-shaped clusters, while incorporating soft constraints on smoothness. More precisely, we allow the radius of the cluster around a center location to vary along with angle, and penalize proportional to the total change in radius.

Methods

We propose a dynamic programming based solution to find optimal clusters, with penalty terms introduced to control smoothness in the radius of the cluster. Our computationally efficient StarScan algorithm uses the key observation[3] that the log-likelihood ratio score may be written as an additive function, summing over all data elements, when conditioning on the relative risk value q. Given a region S, the log-likelihood ratio score F(S) is given by maximizing over the whole range of relative risk values. Let the score of a region for a given relative risk be given by $F(S \mid q)$, and let R(S) be its total change in the radius, for a given center location, to form the cluster. We use dynamic programming to find the optimal region S that maximizes the penalized score $F'(S \mid q) = F(S \mid q) - \lambda R(S)$, where the constant λ represents the amount of penalization for a given change in radius. We find the optimal penalized score $F'(S^*)$ and corresponding optimal subset S^*, by either grid search (evaluating a range of possible values of q) or using branch and bound techniques in order to find the optimal value of q.

Results

StarScan was compared to the circular scan[1] and fast subset scan[2] on simulated respiratory outbreaks and bioterrorist anthrax attacks injected into real-world Emergency Department data from Allegheny County, PA. Given a small amount of labeled training data, StarScan learns appropriate penalties for both compact and elongated clusters, resulting in improved detection performance. For irregularly shaped injects, StarScan improves performance both in terms of increasing the spatial overlap between true and detected regions, and increasing detection power as measured by the average number of days to detection at a fixed false positive rate. Finally, we show that StarScan generalizes both circular scan (for large λ) and fast localized subset scan (for $\lambda \rightarrow 0$).

Conclusions

StarScan generalizes the traditional, circular spatial scan statistic[1] by allowing the radius of the cluster around a center location to vary continuously with the angle, but penalizes the log-likelihood ratio score proportional to the total change in radius. This penalization allows StarScan to find irregularly-shaped clusters more accurately than either the circular scan or unconstrained fast subset scan, both of which are shown to be special cases of StarScan with appropriate choices of penalty.
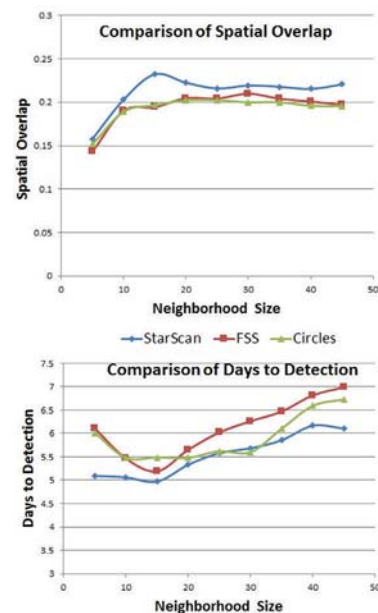
Comparison of Detection Performance

Keywords

Biosurveillance; Scan statistics; Dynamic programming

References

1. Kulldorff M. A spatial scan statistic, Communications in Statistics, Theory and Methods, 1997.
2. Neill DB. Fast subset scan for spatial pattern detection. Journal of the Royal Statistical Society (Series B: Statistical Methodology) 74(2): 337-360, 2012.
3. Speakman S, Somanchi S, McFowland E, Neill DB. Penalized fast subset scanning. Under review, 2014.

*Sriram Somanchi
E-mail: somanchi@cmu.edu