OJPHI

ISDS 2016 Conference Abstracts

ISDS INTERNATIONAL SOCIETY FOR DISEASE SURVEILLANCE

Support Vector Subset Scan for Spatial Outbreak Detection

Dylan Fitzpatrick*, Yun Ni and Daniel B. Neill

Heinz College/Machine Learning Department, Carnegie Mellon University, Pittsburgh, PA, USA

Objective

We present the support vector subset scan (SVSS), a new method for detecting localized and irregularly shaped patterns in spatial data. SVSS integrates the penalized fast subset scan[3] with a kernel support vector machine classifier to accurately detect disease clusters that are compact and irregular in shape.

Introduction

Neill's fast subset scan[2] detects significant spatial patterns of disease by efficiently maximizing a log-likelihood ratio statistic over subsets of locations, but may result in patterns that are not spatially compact. The penalized fast subset scan (PFSS)[3] provides a flexible framework for adding soft constraints to the fast subset scan, rewarding or penalizing inclusion of individual points into a cluster with additive point-specific penalty terms. We propose the support vector subset scan (SVSS), a novel method that iteratively assigns penalties according to distance from the separating hyperplane learned by a kernel support vector machine (SVM). SVSS efficiently detects disease clusters that are geometrically compact and irregular.

Methods

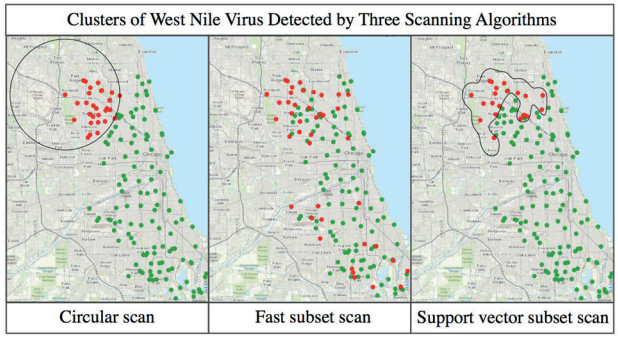
Speakman[3] observes that for a fixed value of relative risk q, the log-likelihood ratio for the exponential family of expectation-based scan statistics can be written as an additive set function over all data elements. This property enables addition of element-specific penalty terms to the log-likelihood ratio, interpreted as the prior log-odds of including a data point in the cluster. We propose an iterative method for setting the penalty terms which leads to spatially compact clusters, alternately running PFSS to obtain an optimal subset and training a kernel SVM to maximize the margin between points within and outside of the subset. On each iteration of PFSS, penalties are assigned based on distance to the SVM decision boundary. We apply random restarts across the penalty space to approach a global optimum in the non-convex SVSS objective function.

Results

We demonstrate detection of disease clusters in mosquito pools tested for West Nile Virus (WNV), using data made publicly available by the Chicago Department of Public Health through the City of Chicago Data Portal. In comparison to the circular scan[1], which detects circular patterns with elevated WNV, SVSS has improved power to detect disease clusters that are elongated or irregular in shape. For example, the top WNV cluster detected by SVSS roughly conforms to sections of two major rivers in North Chicago, overlapping significant portions of the forest preserves adjacent to these rivers. The unconstrained fast subset scan[2] has high detection power for subtle and irregular disease clusters, but finds patterns that are spatially sparse and intermingled with non-anomalous points. SVSS rewards patterns with spatial coherence, detecting clusters that are compact and separated from non-anomalous points while maintaining power to detect slight but significant increases in detected rates of WNV.

Conclusions

SVSS introduces soft spatial constraints to the fast subset scan[2] in the form of penalties to the log-likelihood ratio statistic, learned iteratively based on distance to a high-dimensional SVM decision boundary. These constraints give SVSS greater power to detect spatially compact and irregular patterns of disease.



Clusters of West Nile Virus detected by three scanning algorithms.

Keywords

Outbreak detection; Cluster detection; Spatial analysis; Machine learning; Subset scan

Acknowledgments

This work was partially supported by NSF grant IIS-0953330.

References

1. Kulldorff M. A spatial scan statistic. Commun Stat Theory Methods. 1997; 26(2): 1481-1496.
2. Neill DB. Fast subset scan for spatial pattern detection. J R Stat Soc Series B Stat Methodol. 2012; 74(2): 337-360.
3. Speakman, S, Somanchi, S, McFowland, E III, Neill, DB. Penalized fast subset scanning. J Comp Graph Stat. 2016; 25(2): 382-404.

***Dylan Fitzpatrick**
E-mail: djfitzpa@andrew.cmu.edu