Hot-Spot Geoinformatics for Digital Governance

G. P. Patil
The Pennsylvania State University, USA

R. Acharya
The Pennsylvania State University, USA

R. Modarres
George Washington University, USA

W. L. Myers
The Pennsylvania State University, USA

S. L. Rathbun
University of Georgia, USA

INTRODUCTION

Geoinformatic surveillance for spatial and temporal hot-spot detection and prioritization is crucial in the 21st century. A hot spot may be any unusual phenomenon, anomaly, aberration, outbreak, elevated cluster, or critical area. Government agencies require hot-spot delineation and prioritization for monitoring, etiology, management, or early warning. Responsible factors may be natural, accidental, or intentional, with relevance to both infrastructure and security.

This article describes multidisciplinary research based on novel methods for hot-spot detection and prioritization, driven by a diverse variety of case studies of interest to agencies, academia, and the private sector. These case studies concern critical societal issues, such as public health, ecosystem health, biodiversity and threats to biodiversity, emerging infectious disease, water management and conservation, carbon sources and sinks, persistent poverty, environmental justices, crop pathogens, invasive-species management, biosurveillance, biosecurity, disease biogeoinformatics, social networks, sensor networks, hospital networks and syndrome surveillance, video mining, early warning, tsunami inundation, remote sensing, and disaster management.

Our approach has involved an innovation of the popular circle-based spatial scan statistic. In particular, it employs the notion of an upper level set (ULS) and is accordingly called the upper level set scan statistic system, pointing to the next generation of sophisticated analytical and computational systems, effective for the detection of arbitrarily shaped hot spots along spatiotemporal dimensions. It also involves a novel prioritization scheme based on multiple indicators and stakeholder criteria without having to reduce indicators to a single index using Hasse diagrams and partially ordered sets. It is accordingly called the poset prioritization and ranking system (see Patil & Taillie, 2004a, 2004b).

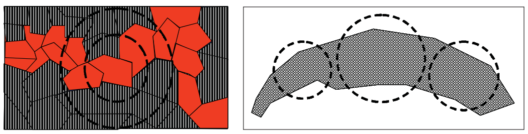
The following Web sites have additional information.

1. http://www.stat.psu.edu/hotspots/
2. http://www.stat.psu.edu/~gpp/
3. http://www.digitalgovernment.org/news/stories/2004/1104/1104_hotspots_heyman.jsp

UPPER LEVEL SET HOT-SPOT SCAN STATISTIC SYSTEM

Patil and Taillie (2004a, 2004b) introduce an innovation of the health-area-popular circle-based spatial and spatiotemporal scan statistic. It employs the notion of an upper level set, and is accordingly called the upper level set scan statistic, pointing to a sophisticated analytical and computational system as the next generation of the present-day popular SaTScan (Kulldorff, 1997, 2001; Kulldorff & Nagarwalla, 1995; Kulldorff, Rand, Gherman, Williams, & Defrancesco, 1998; Mostashari, Kulldorff, & Miller, 2002; Waller, 2002).

Figure 1. Limitations of circular scanning windows. (Left) An irregularly shaped cluster—perhaps a cholera outbreak along a winding river floodplain. Small circles miss much of the outbreak and large circles include many unwanted cells. (Right) Circular windows may report a single irregularly shaped cluster as a series of small clusters.



Background Theory of Scan Statistics

The spatial scan statistic concerns the following situation: A region R of Euclidian space is tessellated or subdivided into cells, which will be denoted by the symbol a. Data is available in the form of a count Y_a on each cell a. In addition, a size value A_a is associated with each cell. The cell sizes A_a are regarded as fixed and known, while the cell counts Y_a are independent random variables. Two distributional settings are commonly studied:

- **Binomial:** The size $A_a=N_a$ is a positive integer and Y_a~Binomial (N_a,p_a), where p_a is an unknown parameter attached to cell a with $0<p_a<1$.
- **Poisson:** The size A_a is a positive real number and Y_a~Poisson $(\lambda_a A_a)$, where $\lambda_a>0$ is an unknown parameter attached to cell a.

Each distributional model has a simple interpretation. For the binomial, N_a people reside in cell a and each person contracts a certain disease independently with probability p_a. The cell count Y_a is the number of diseased people. For the Poisson, A_a is the size (e.g., area or some adjusted population size) of the cell a, and Y_a is a realization of a Poisson process with intensity λ_a. In each scenario, the responses Y_a are independent; it is assumed that spatial variability can be accounted for by cell-to-cell variation in model parameters.

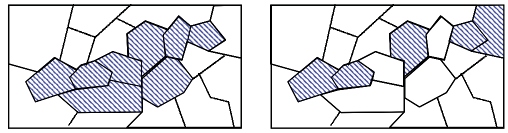
The spatial scan statistic seeks to identify hot spots or clusters of cells having an elevated response with respect to the remainder of the region. Elevated response means large values for the rates (or intensities),

$$G_a=Y_a / A_a,$$

instead of the raw counts Y_a. The scan statistic easily accommodates other adjustments, such as for age or gender.

A collection of cells from the tessellation should satisfy several geometric properties before it can be considered as a candidate for a hot-spot cluster. First, the union of the cells should comprise a geographically connected subset of the region R (Figure 2). Such collections of connected cells will be referred to as zones Z and the set of all zones is denoted by Ω. Second, the zone should not be excessively large. Otherwise, the zone instead of its exterior would constitute the background. This restriction is generally achieved by limiting the search for hot spots to zones comprising of less than, say, 50% of the region.

Figure 2. A tessellated region. The collection of shaded cells in the left-hand diagram is connected and, therefore, constitutes a zone in Ω. The collection on the right is not connected.



The notion of a hot spot is inherently vague and lacks any a priori definition. There is no true hot spot in the statistical sense of a true parameter value. A hot spot is instead defined by its estimate, provided the estimate is statistically significant. To this end, the scan statistic adopts a hypothesis testing model in which the hot spot occurs as an unknown zonal parameter in the statement of the alternative hypothesis.

The traditional spatial scan statistic uses expanding circles to determine a reduced list Ω_0 of candidate zones Z. By their very construction, these candidate zones tend to be compact in shape and may do a poor job of approximating actual clusters. The reduced parameter space of the circular scan statistic is determined entirely by the geometry of the tessellation and does not involve the data in any way. We propose a scan statistic that takes an adaptive point of view in which Ω_0 depends very much upon the data. Furthermore, Ω_0 induces a tree structure useful for visualization and expressing uncertainty of hot-spot clusters in the form of a hot-spot confidence set on the tree.

Although the traditional spatial scan statistic is applicable only to tessellated data, the ULS approach has an abstract graph (i.e., vertices and edges) as its starting point. Accordingly, this approach can also be applied to data defined over networks, such as subway, water, or highway systems. There is complete flexibility regarding the definition of adjacency. For example, one may declare two cells as adjacent if (a) their boundaries have at least one point in common, (b) their common boundary has positive length, or (c) in the case of a drainage network, the flow is from one cell to the next.

Related Content

Evaluating Citizen Adoption and Satisfaction of E-Government

Craig P. Orgeronand Doug Goodman (2011). *International Journal of Electronic Government Research (pp. 57-78).*

www.irma-international.org/article/evaluating-citizen-adoption-satisfaction-government/56099

Online Presence of the Members of the Turkish Parliament: Evaluation of the Turkish MP Web Sites

Mete Yildiz, Ahu Sumbasand Kadir Dede (2012). *E-Parliament and ICT-Based Legislation: Concept, Experiences and Lessons (pp. 164-179).*

www.irma-international.org/chapter/online-presence-members-turkish-parliament/60638

Open Government and Bureaucratic Secrecy in the Developing Democracies: Africa in Perspective

Luke A. Amadiand Prince Igwe (2018). *Proliferation of Open Government Initiatives and Systems (pp. 1-28).*

www.irma-international.org/chapter/open-government-and-bureaucratic-secrecy-in-the-developing-democracies/195687

Employees' Perceptions of Biometric Technology Adoption in E-Government: An Exploratory Study in the Kingdom of Saudi Arabia

Thamer Alhussainand Steve Drew (2012). *Digital Democracy: Concepts, Methodologies, Tools, and Applications (pp. 1120-1133).*

www.irma-international.org/chapter/employees-perceptions-biometric-technology-adoption/67652

The Impact of Institutions on Interorganizational IT Projects in the Mexican Federal Government

Luis F. Luna-Reyes, J. Ramon Gil-Garciaand Mireya Estrada-Marroquin (2008). *International Journal of Electronic Government Research (pp. 27-42).*

www.irma-international.org/article/impact-institutions-interorganizational-projects-mexican/2049