



## A Bayesian network model for spatial event surveillance

Xia Jiang<sup>a,\*</sup>, Daniel B. Neill<sup>b</sup>, Gregory F. Cooper<sup>a</sup>

<sup>a</sup> Department of Biomedical Informatics, University of Pittsburgh, Parkvale Building M183, 200 Meyran Ave., Pittsburgh, PA 15260, United States

<sup>b</sup> H.J. Heinz III College, School of Public Policy and Management, Carnegie Mellon University, Pittsburgh, PA 15213, United States

### ARTICLE INFO

#### Article history:

Received 19 March 2008

Received in revised form 8 August 2008

Accepted 5 January 2009

Available online 13 January 2009

#### Keywords:

Bayesian network

Spatial cluster detection

Event surveillance

Spatial event surveillance

Outbreak detection

### ABSTRACT

Methods for spatial cluster detection attempt to locate spatial subregions of some larger region where the count of some occurrences is higher than expected. Event surveillance consists of monitoring a region in order to detect emerging patterns that are indicative of some event of interest. In spatial event surveillance, we search for emerging patterns in spatial subregions.

A well-known method for spatial cluster detection is Kulldorff's [M. Kulldorff, A spatial scan statistic, *Communications in Statistics: Theory and Methods* 26 (6) (1997)] spatial scan statistic, which directly analyzes the counts of occurrences in the subregions. Neill et al. [D.B. Neill, A.W. Moore, G.F. Cooper, A Bayesian spatial scan statistic, *Advances in Neural Information Processing Systems (NIPS)* 18 (2005)] developed a Bayesian spatial scan statistic called BSS, which also directly analyzes the counts.

We developed a new Bayesian-network-based spatial scan statistic, called BNetScan, which models the relationships among the events of interest and the observable events using a Bayesian network. BNetScan is an entity-based Bayesian network that models the underlying state and observable variables for each individual in a population.

We compared the performance of BNetScan to Kulldorff's spatial scan statistic and BSS using simulated outbreaks of influenza and cryptosporidiosis injected into real Emergency Department data from Allegheny County, Pennsylvania. It is an open question whether we can obtain acceptable results using a Bayesian network if the probability distributions in the network do not closely reflect reality, and thus, we examined the robustness of BNetScan relative to the probability distributions used to generate the data in the experiments. Our results indicate that BNetScan outperforms the other methods and its performance is robust relative to the probability distribution that is used to generate the data.

© 2009 Elsevier Inc. All rights reserved.

## 1. Introduction

Methods for spatial cluster detection attempt to locate spatial subregions of some larger region where the count of some occurrences is higher than expected. As an example, we may look for clusters of certain kinds of trees in a forest; other applications of spatial cluster detection include mining astronomical data, medical imaging, and military surveillance [1]. Event surveillance consists of monitoring a region in order to detect emerging patterns that are indicative of some event of interest. As examples, we may look for emerging patterns that are indicative of a disaster that is in its early stages of development. Examples of such disasters include hurricanes, terrorist attacks, and outbreaks of diseases. In spatial event surveillance, we search for emerging patterns in spatial subregions. Spatial cluster detection is one statistical technique used for spatial event surveillance.

\* Corresponding author.

E-mail addresses: [xjiang@cbmi.pitt.edu](mailto:xjiang@cbmi.pitt.edu) (X. Jiang), [neill@cs.cmu.edu](mailto:neill@cs.cmu.edu) (D.B. Neill), [gfc@pitt.edu](mailto:gfc@pitt.edu) (G.F. Cooper).

A well-known method for spatial cluster detection is Kulldorff's [1] spatial scan statistic, which has been implemented as freeware in the SaTScan™ software package [2]. This method directly analyzes the counts of occurrences in the various subregions. Neill et al. [3] developed a Bayesian spatial scan statistic called BSS, which also directly analyzes the counts. When the clusters of interest are clusters of the events we observe, then directly analyzing the counts is likely to be best. However, in event surveillance, and disease-outbreak detection in particular, we ordinarily observe events that are *related* to the occurrences of interest. As an example, when we are interested in whether there is an outbreak of a certain disease, we observe individuals with various symptoms of the disease rather than the disease infections themselves. Instead of using a summary statistic, we develop a Bayesian-network-based spatial scan statistic, called BNetScan, which models the relationships among the events of interest and the observable events using a Bayesian network. This network is then used to determine the posterior probability of each subregion containing a cluster. A strength of this method is that it can model multiple causes of the clusters. For example, in disease-outbreak detection it can model any number of possible disease outbreaks using a single Bayesian network.

An important question is whether we can obtain acceptable results using a Bayesian network if the probability distributions in the network, which are often obtained from limited data and/or subjective judgment, do not closely reflect reality. We describe the results of experiments that test the robustness of BNetScan relative to changes in the probability distributions used to generate the experimental data. In particular, we examine the ability of BNetScan to detect outbreaks of influenza and cryptosporidiosis using simulated cases injected into real Emergency Department data from Allegheny County, Pennsylvania. As a point of comparison, we included SaTScan™ and BSS in the study, and we evaluated these systems both when they took advantage of the probabilistic information in BNetScan (i.e., which chief complaints are most commonly associated with each outbreak type) and when they did not. In this way, we could learn whether it is simply the use of probabilistic information that accounts for a certain performance level, or whether that performance may be due to the use of a Bayesian network model.

In the remainder of this section, we provide background on spatial cluster detection. Section 2 describes the two methods, SaTScan™ and BSS, which we compare to BNetScan. In Section 3 we develop our new Bayesian-network-based spatial scan statistic method, BNetScan. Section 4 presents our experiments and their results. Finally, Section 5 provides a discussion of the results.

### 1.1. Spatial cluster detection

In spatial cluster detection, the goal is to identify the location, shape, and size of possible clusters, and to determine how likely it is that a cluster is due to the event with which we are concerned (e.g., a disease outbreak) versus how likely it is that the cluster is merely a chance occurrence. When doing spatial cluster detection, we first enumerate the subregions of some geographical region  $G$ . For example, Kulldorff [1] places a circular window over the region and lets the center of the circle move over the region. For each center, the radius of the circle is varied. Neill et al. [3] represent the entire region by a grid and search over rectangular subregions of the grid. Inherent in these methods is that we assume that the entire region  $G$  is composed of cells. For example, if we cover  $G$  with an  $m \times n$  grid, each grid element is a cell. A subregion of  $G$  is the union of any number of cells. Since the number of such subregions is exponentially large, Neill et al. [3] search only over subregions which are axis-aligned rectangles. We take this same approach in this paper.

Classical methods for spatial cluster detection detect clusters based solely on counts of occurrences in the various subregions. Kulldorff [1,4] developed the well-known frequentist method called the spatial scan statistic. The scan statistic was first proposed by Naus [5] as a solution to the multiple hypothesis testing problem. Scan statistics have been used to find clusters of chronic diseases such as breast cancer [6] and leukemia [7]. They have also been used to detect clusters of work-related hazards [8] and West Nile virus [9]. As mentioned earlier, Kulldorff implemented the spatial scan statistic in his SaTScan™ software [2]. Jung et al. [10] develop a version of the spatial scan statistic that considers multinomial variables whose values are ordinal. Neill et al. [3] developed a Bayesian spatial scan statistic, and Neill et al. [11] developed a multivariate version of the Bayesian spatial scan statistic.

The methods discussed so far are non-temporal. That is, if we are looking for patterns that emerge with time, they look only at data from the most recent time period. A temporal method would detect a cluster based not only on data from the most recent time period, but also on data from previous time periods. Kulldorff et al. [12] developed a temporal version of the spatial scan statistic that looks at three-dimensional cylinders, while Neill et al. [13] developed a temporal version of the spatial scan statistic that can detect emerging clusters with incidence rates that increase over time.

In general, spatial cluster detection does not necessarily entail the notion of time. In some applications we many want to determine whether there is currently a cluster without concern for whether the counts are changing with time. For example, we may want to investigate whether there is a cluster of a particular type of star in space. Next we discuss spatial event surveillance, which does concern counts changing with time.

### 1.2. Spatial event surveillance

In event surveillance we try to detect emerging patterns that are indicative of some forthcoming event or an event that is in its early stages. A classical example is disease-outbreak detection. A disease outbreak detection system monitors a region each day to see if some pattern has emerged that is indicative of a disease outbreak. For example, a *Cryptosporidium* infection

متن کامل مقاله

دریافت فوری ←

**ISI**Articles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات