Information Fusion 13 (2012) 117–128

Contents lists available at ScienceDirect

Information Fusion

journal homepage: www.elsevier.com/locate/inffus

Optimizing biosurveillance systems that use threshold-based event detection methods

Ronald D. Fricker Jr. *, David Banschbach

Naval Postgraduate School, Operations Research Department, Monterey, CA, USA

ARTICLE INFO

Article history:
Received 20 October 2008
Received in revised form 24 November 2009
Accepted 2 December 2009
Available online 4 January 2010

Keywords:
Biosurveillance
Syndromic surveillance
Bioterrorism
Public health
Optimization
Shewhart chart

ABSTRACT

We describe a methodology for optimizing a threshold detection-based biosurveillance system. The goal is to maximize the system-wide probability of detecting an "event of interest" against a noisy background, subject to a constraint on the expected number of false signals. We use nonlinear programming to appropriately set detection thresholds taking into account the probability of an event of interest occurring somewhere in the coverage area. Using this approach, public health officials can "tune" their biosurveillance systems to optimally detect various threats, thereby allowing practitioners to focus their public health surveillance activities. Given some distributional assumptions, we derive a one-dimensional optimization methodology that allows for the efficient optimization of very large systems. We demonstrate that optimizing a syndromic surveillance system can improve its performance by 20–40%.

Published by Elsevier B.V.

1. Introduction

Biosurveillance is the practice of monitoring populations – human, animal, and plant – for the outbreak of disease. Often making use of existing health-related data, one of the principle objectives of biosurveillance systems has been to give early warning of bioterrorist attacks or other emerging health conditions [4]. The Centers for Disease Control and Prevention (CDC) as well as many state and local health departments around the United States are developing and fielding *syndromic surveillance* systems, one type of biosurveillance.

A *syndrome* is "A set of symptoms or conditions that occur together and suggest the presence of a certain disease or an increased chance of developing the disease" [17]. In the context of syndromic surveillance, a syndrome is a set of non-specific pre-diagnosis medical and other information that may indicate the health effects of a bioterrorism agent release or natural disease outbreak. See, for example, Syndrome Definitions for Diseases Associated with Critical Bioterrorism-associated Agents [3]. The data in syndromic surveillance systems may be clinically well-defined and linked to specific types of outbreaks, such as groupings of ICD-9 codes from emergency room "chief complaint" data, or only vaguely defined and perhaps only weakly linked to specific types of outbreaks, such as over-the-counter sales of cough and cold medication or absenteeism rates.

Since its inception, one focus of syndromic surveillance has been on early event detection: gathering and analyzing data in advance of diagnostic case confirmation to give early warning of a possible outbreak. Such early event detection is not supposed to provide a definitive determination that an outbreak is occurring. Rather, it is supposed to signal that an outbreak *may* be occurring, indicating a need for further evidence or triggering an investigation by public health officials (i.e., the CDC or a local or state public health department). See Fricker [10,9] and Fricker and Rolka [11] for more detailed exposition and discussion.

BioSense and EARS are two biosurveillance applications currently in use. The first is a true system, in the sense that it is comprised of dedicated computer hardware and software that collect and evaluate data routinely submitted from hospitals. The second is a set of software programs that are available for implementation by any public health organization.

- *BioSense* was developed and is operated by the National Center for Public Health Informatics of the CDC. It is intended to be a United States-wide electronic biosurveillance system. Begun in 2003, BioSense initially used Department of Defense and Department of Veterans Affairs outpatient data along with medical laboratory test results from a nationwide commercial laboratory. In 2006, BioSense began incorporating data from civilian hospitals as well. The primary objective of BioSense is

* Corresponding author. Tel.: +1 831 656 3048; fax: +1 831 656 2595.
E-mail address: rdfricke@nps.edu (R.D. Fricker Jr.).

1566-2535/$ - see front matter Published by Elsevier B.V.
doi:10.1016/j.inffus.2009.12.002

to "expedite event recognition and response coordination among federal, state, and local public health and health care organizations" [10,5,22,23]. As of May 2008, BioSense was receiving data from 563 facilities [7].
- *EARS* is an acronym for Early Aberration Reporting System. Developed by the CDC, EARS was designed to monitor for bioterrorism during large-scale events that often have little or no baseline data (i.e., as a short-term drop-in surveillance method) [6]. For example, the EARS system was used in the aftermath of Hurricane Katrina to monitor communicable diseases in Louisiana, for syndromic surveillance at the 2001 Super Bowl and World Series, as well as at the Democratic National Convention in 2000 [24,15]. Though developed as a drop-in surveillance method, EARS is now being used on an on-going basis in many syndromic surveillance systems.

A characteristic of some syndromic surveillance systems is that the data collection locations (typically hospitals and clinics) are in fixed locations that may or may not correspond to a particular threat of either natural disease or bioterrorism. In order to provide comprehensive population coverage, syndromic surveillance system designers and operators are inclined to enlist as many hospitals and clinics as possible. However, as the sources and types of data being monitored proliferate in a biosurveillance system, then so do the false positive signals from the systems. Indeed, false positives have become an epidemic problem for some systems. As one researcher [21] said, "...most health monitors... learned to ignore alarms triggered by their system. This is due to the excessive false alarm rate that is typical of most systems – there is nearly an alarm every day!"

Our research provides a methodology which, if implemented, would allow public health officials to "tune" their biosurveillance systems to optimally detect various threats while explicitly accounting for organizational resource constraints available for investigating and adjudicating signals. This allows practitioners to focus their public health surveillance activities on locations or diseases that pose the greatest threat at a particular point in time. Then, as the threat changes, using the same hospitals and clinics, the system can subsequently be tuned to optimally detect other threats. With this approach large biosurveillance systems are an asset.

The methodology assumes spatial independence of the data and temporal independence of the signals. The former is achieved by monitoring the residuals from some sort of model to account for and remove the systematic effects present in biosurveillance data. The assumption is that, while it is likely that raw biosurveillance data will have spatial correlation, once the systematic components of the data are removed the residuals will be independent. The latter is achieved by employing detection algorithms that only depend on data from the current time period.

It is worth emphasizing that our focus is on how to optimally set threshold levels for detection in an *existing* system, rather than how to design a new system. This is something of a unique problem for syndromic surveillance systems, meaning that in many other types of sensor systems, one might design a system for a specific, unchanging threat or change the location of the sensors to respond to a changing threat. But in syndromic surveillance systems, where we can think of each hospital or clinic as a fixed biosurveillance "sensor" for a particular location or population, the sensor locations cannot be changed. Part of the solution is to adjust the way the data from the sensors are monitored.

1.1. Threshold detection methods

In this work, we define a *threshold detection* method as an algorithm that generates a binary output, signal or no signal, given that some function of the input or inputs exceed a pre-defined threshold level. In addition, for the methods we consider, inputs come in discrete time periods and the decision to signal or not is based only on the most recent input or inputs. That is, the methods do not use historical information in their signal determination; they only use the information obtained at the current time period.

In the quality control literature, the *Shewhart chart* is such a threshold detection method. At each time period a measurement is taken and plotted on a chart. If the measurement exceeds a pre-defined threshold a signal is generated. However, if the measurement does not exceed the threshold then the process is repeated at the next time period, and continues to be repeated until such time as the threshold is exceeded. See Shewhart [20] or Montgomery [19] for additional detail. A sonar detection algorithm based on signal excess is also an example of threshold detection. See Washburn [26] and references therein for a discussion.

Threshold detection methods are subject to errors, either signalling that an event of interest occurred when it did not, or failing to signal when in fact the event of interest did occur. In classical hypothesis testing, these errors are referred to as *Type I* and *Type II* errors, respectively. A Type I error is a false signal and a Type II error is a missed detection. In threshold detection, setting the threshold requires making a trade-off between the probability of false signals and the probability of a missed detection. A *receiver operating characteristic* (or ROC) curve is a plot of the probability of false signal versus probability of detection (one minus the probability of a missed detection) for all possible threshold levels. See Washburn [26, Chapter 10] and the references therein for additional discussion.

1.2. Optimizing sensor systems

Optimizing a system of threshold detection-based sensors, in the sense of maximizing the probability of detecting an event of interest somewhere in the region being monitored by the system, subject to a constraint on the expected number of system-wide false signals, to the best of our knowledge, has not been done. Washburn [26, Chapter 10.4] introduces the idea of optimizing the threshold for a single sensor, parameterizing the problem in terms of the cost of a missed detection and the cost of a false signal, and seeks to minimize the average cost "per look". He concludes that "In practice, the consequences of the two types of error are typically so disparate that it is difficult to measure c_1 [cost of a missed detection] and c_2 [cost of a false signal] on a common scale. For this reason, the false alarm probability is typically not formally optimized in practice".

Kress et al. [18] develop a methodology for optimizing the employment of non-reactive arial sensors. In their problem the goal is to optimize a mobile sensor's search path in order to identify the location or locations of fixed targets with high probability. By dividing the search region into a grid of cells, Kress et al. use a Bayesian updating methodology combined with an optimization model that seeks to maximize the probability of target location subject to a constraint on the number of looks by the sensors. Their work differs from ours in a number of important respects, including that their sensors can have multiple looks for a target, there may be multiple targets present, and the use of Bayesian updating to calculate the probability of a target being present in a particular grid cell. In contrast, in our problem the sensors are fixed, they can only take one look per period, and at most one "event of interest" can occur in any time period.

One active area of research is how to combine threshold rules for systems of sensors in order to achieve high detection rates and low false positive rates compared to the rates for individual sensors. For example, Zhu et al. [28] consider a system of threshold detection sensors for which they propose a centralized "threshold-

Download English Version:

https://daneshyari.com/en/article/528244

Download Persian Version:

https://daneshyari.com/article/528244

Daneshyari.com