A Bayesian Algorithm for Detecting CDC Category A Outbreak Diseases from Emergency Department Chief Complaints

Gregory F. Cooper, John N. Dowling, John D. Levander, Peter Sutovsky
Department of Biomedical Informatics, University of Pittsburgh

OBJECTIVE

This paper describes a Bayesian algorithm for diagnosing the CDC Category A diseases, namely, anthrax, smallpox, tularemia, botulism and hemorrhagic fever, using emergency department chief complaints. The algorithm was evaluated on real data and on semi-synthetic data, and this paper summarizes the results of that evaluation.

BACKGROUND

A number of outbreak detection algorithms have been developed using syndromic data as input [1]. To our knowledge, no algorithm, including Bayesian algorithms, has been reported to diagnose all of the CDC Category A diseases [2]. This paper describes such an algorithm.

METHODS

We extended the PANDA Bayesian network system in [3], which only monitored for windborne anthrax outbreaks, to model the CDC Category A diseases. This augmented system, which we call PANDA+, takes as input a time series of emergency department (ED) chief complaints, and each hour it outputs the posterior probability of (1) each CDC Category A disease and (2) several additional diseases, namely, influenza, crytosporidiosis, hepatitis A and asthma.

We performed two experiments using PANDA+. First, we used a previously developed anthrax outbreak simulator (BARD), along with chief complaint probabilities taken from PANDA+, to generate 96 outbreak scenarios [3]. Each scenario consisted of a time series of outbreak chief complaints that were overlaid on a time series of real ED chief complaints. In a second experiment, we evaluated the ability of PANDA+ to detect a laboratory validated outbreak of influenza in Pittsburgh, PA in 2003.

RESULTS

Figure 1 shows the results of the anthrax experiment, which are comparable in performance to those obtained by the anthrax-specific version of PANDA described in [3]. For a simulated release of 1 kg, the PANDA+ detection time was 35 hours after release, when the false positive rate (fpr) was zero. For the version of PANDA in [3], the detection time was 46 hours.

The results of the influenza experiment are shown in Figure 2, which indicates a detection time of 4 hours when the fpr was 0.001/hour (≅ 1/month).

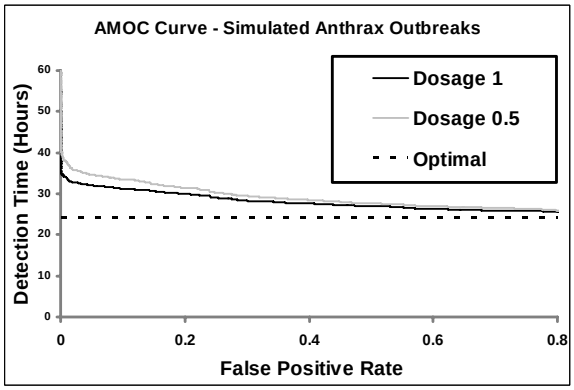

Figure 1. An AMOC plot for the detection of anthrax in a simulated outbreak due to a windborne-release of anthrax spores. The false positive rate per hour is shown.

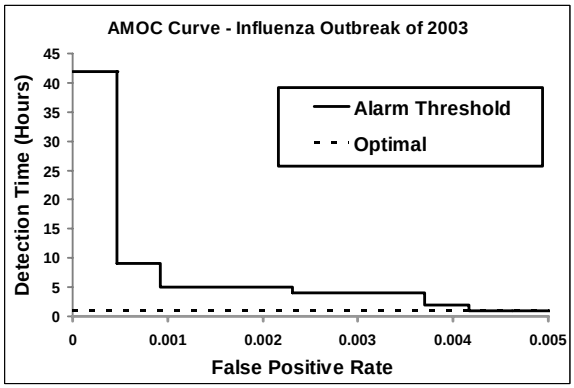

Figure 2. An AMOC plot for the detection of influenza in a real outbreak of influenza in Pittsburgh, Pennsylvania in 2003.

CONCLUSIONS

PANDA+ performed well in detecting a previous real outbreak and a simulated outbreak, and in determining the identity of the underlying outbreak diseases. In future work, we plan to evaluate its ability to detect the other CDC Category A diseases. We also plan extensions that include expanding its spatial and temporal model, the evidence used as input, and the diseases modeled as potential causes of outbreaks.

REFERENCES

[1] Wagner MM, Moore AW, Aryl R (eds). Handbook of Biosurveillance (Elsevier, 2006).

[2] http://www.bt.cdc.gov/agent/agentlist-category.asp

[3] Cooper GF, Dash DH, Levander JD, Wong WK, Hogan WR, Wagner MM. Bayesian biosurveillance of disease outbreaks. In: *Proceedings of the Conference on Uncertainty in Artificial Intelligence* (2004) 94-103.

Further Information: Gregory Cooper, gfc@cbmi.pitt.edu, www.cbmi.pitt.edu/panda