An Outbreak Detection Algorithm that Efficiently Performs Complete Bayesian Model Averaging Over All Possible Spatial Distributions of Disease

Yanna Shen[1,2], Weng-Keen Wong[3], John D. Levander[2] and Gregory F. Cooper[1,2]

[1]Intelligent Systems Program and [2]Department of Biomedical Informatics, University of Pittsburgh; [3]School of Electrical Engineering and Computer Science, Oregon State University

OBJECTIVE

We introduce a disease-outbreak detection algorithm that performs complete Bayesian Model Averaging (BMA) over all possible spatial distributions of disease, yet runs in polynomial time.

BACKGROUND

Many disease-outbreak detection algorithms, such as control chart methods, use frequentist statistical techniques. We describe a Bayesian algorithm that uses data D consisting of current day counts of some event (e.g., emergency department (ED) chief complaints of respiratory disease) that are tallied according to demographic area (e.g., zip codes).

METHODS

Assume there are r zip codes in a region being monitored. Let i, $1 \le i \le r$, represent the index of a specific zip code. We use OB_i and NOB_i to represent the disease outbreak states for zip code i, namely, *outbreak* and *non-outbreak*. Therefore, we have a total of 2^r possible outbreak states in the region. If we perform complete BMA over all of these states in a brute-force way, the time complexity is exponential in r. By factoring the states, we can derive a polynomial time, spatial BMA detection algorithm, which we call SBMA.

Let q be the probability that a zip code has an outbreak (i.e., $P(OB_i \mid q) = q$). We use $P(q)$ to represent our belief about q. Let OB denote the state that at least one zip code (among the total of r zip codes) has an outbreak present; let NOB represent the state that no zip code has an outbreak, where $P(NOB) = \alpha$. When $0 < q \le 1$, we model that $P(q) = 1 - \alpha$. We derive the joint probability of the data (D) and the *outbreak* state (OB) using Eq. (1).

$$P(D, OB) = \int_{0+}^{1} P(q)\left[\prod_{i=1}^{r}[P(D_i, NOB_i \mid q) + P(D_i, OB_i \mid q)] - \prod_{i=1}^{r} P(D_i, NOB_i \mid q)\right]dq \quad (1)$$

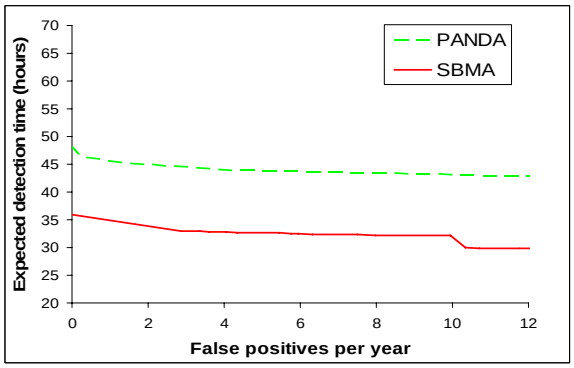
where D_i represents the "respiratory" state (true or false) of every person in zip code i who came to the ED in the last 24 hours, $P(D_i, OB_i \mid q) = qP(D_i \mid OB_i)$ and $P(D_i, NOB_i \mid q) = (1 - q)P(D_i \mid NOB_i)$. We model $P(D_i \mid OB_i)$ and $P(D_i \mid NOB_i)$ as Binomial-Beta hierarchical models; details about those models are provided at [2]. In Eq. (1), the outer bracket calculates $P(D, OB \mid q)$, which takes linear time in r. We use numerical integration to approximate the integral, which takes polynomial time. We compute $P(D, NOB)$ using Eq. (2). The SBMA algorithm then derives the posterior probability as $P(OB \mid D) = P(D, OB) / [P(D, OB) + P(D, NOB)]$.

$$P(D, NOB) = P(NOB)\prod_{i=1}^{r} P(D_i \mid NOB_i) = \alpha \prod_{i=1}^{r} P(D_i \mid NOB_i) \quad (2)$$

As a preliminary test, we used the 96 simulated anthrax outbreaks described in [1]. Each outbreak consisted of a simulated time series of patient cases that each presented to an ED with a respiratory chief complaint and a home zip code. The probability that a case was assigned to live in a zip code was proportional to the population of that zip code. We call this a *scattered outbreak*. We then overlaid (injected) these simulated outbreak cases onto real ED cases to create a combined dataset. We ran PANDA [1] and the SBMA algorithm on the combined datasets.

RESULTS AND CONCLUSIONS

The figure below shows the AMOC curves for PANDA and SBMA on the scattered injections when $\alpha = 0.9$. The SBMA algorithm shows a better detection performance than PANDA, as expected due to the injections being widely scattered spatially rather than having a plume-like pattern of an outdoor release of anthrax, which PANDA monitors for. Thus, the two algorithms appear to be complementary.



ACKNOWLEDGEMENTS

This research was funded by a grant from the National Science Foundation (NSF IIS-0325581).

REFERENCES

[1] 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) 94-104.

[2] http://www.dbmi.pitt.edu/panda/papers/Shen/ISDS07.pdf