Modeling Baseline Shifts in Multivariate Disease Outbreak Detection

Jialan Que* and Fu-Chiang Tsui

University of Pittsburgh, Pittsburgh, PA, USA

Objective

Outbreak detection algorithms monitoring only disease-relevant data streams may be prone to false alarms due to baseline shifts. In this paper, we propose a Multinomial-Generalized-Dirichlet (MGD) model to adjust for baseline shifts.

Introduction

Population surges or large events may cause shift of data collected by biosurveillance systems [1]. For example, the Cherry Blossom Festival brings hundreds of thousands of people to DC every year, which results in simultaneous elevations in multiple data streams (Fig. 1). In this paper, we propose an MGD model to accommodate the needs of dealing with baseline shifts.

Methods

Existing multivariate algorithms only model disease-relevant data streams (e.g., anti-fever medication sales or patient visits with constitutional syndrome for detection of flu outbreak). On the contrary, we also incorporate a non-disease-relevant data stream as a control factor.

We assume that the counts from all data streams follow a Multinomial distribution. Given this distribution, the expected value of the distribution parameter is not subject to change during a baseline shift; however, it has to change in order to model an outbreak. Therefore, this distribution inherently adjusts for the baseline shifts. In addition, we use the generalized Dirichlet (GD) distribution to model the parameter, since GD distribution is one of the conjugate prior of Multinomial [2]. We call this model the Multinomial-Generalized-Dirichlet (MGD) model.

Results

We applied MGD model in our previous proposed Rank-Based Spatial Clustering (MRSC) algorithm [3]. We simulated both outbreak cases and baseline shift phenomena. The experiment includes two groups of data sets. The first includes the data sets only injected with outbreak cases, and the second includes the ones with both outbreak cases and baseline shifts. We apply MRSC algorithm and a reference method, the Multivariate Bayesian Scan Statistic (MBSS) algorithm (which only analyzes the disease-relevant data streams) [4], to both data sets. Fig. 2 shows the performance of outbreak detection: the ROC curves and AMOC curves of analyzing the data sets with baseline shifts (solid lines) and without (dashed lines). We can see from Fig. 2 that the performance of MBSS dropped much more significantly than MRSC when analyzing the data sets with baseline shifts.

Conclusions

The MGD model can be a good supplement model used to detect disease outbreaks in order to achieve both better sensitivity and better specificity especially when baseline shifts are present in the data.

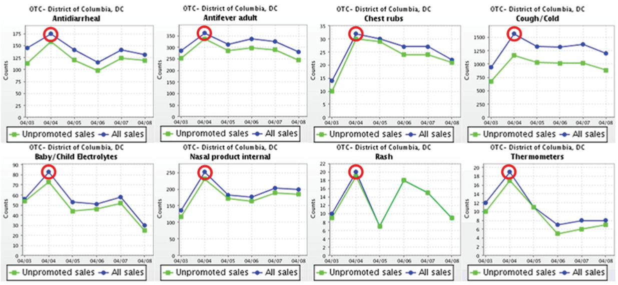

Fig. 1 Eight data streams of NRDM categories collected by RODS system (Anti-Diarrhea, Anti-Fever Adult, Chest Rubs, Cough/Cold, Baby/Child Electrolytes, Nasal Products, Rash and Thermometers) between Apr. 3, 2011 and Apr. 8, 2011 in Washington DC.

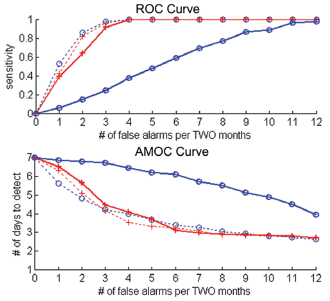

Fig. 2 ROC and AMOC curves of MRSC (red) and MBSS (blue). The solid lines represent the two algorithms applied on the data sets injected with both outbreak cases and baseline shift phenomena. The dashed lines represent the two algorithms applied on the data sets injected with outbreak cases only.

Keywords

Biosurveillance; Disease outbreak detection; Algorithm

Acknowledgments

This research was funded by PA Department of Health syndromic surveillance grant and CDC Center of Excellence grant P01-HK000086.

References

[1] Reis, BY, Kohane, IS and Mandl, KD, An epidemiological network model for disease outbreak detection, PLoS Medicine, vol. 4, p. 210, 2007.

[2] Wong, TT, Generalized Dirichlet distribution in Bayesian analysis, Applied Mathematics and Computation 97, pp. 165-181, 1998.

[3] Que, J and Tsui, FC, Rank-based spatial clustering: an algorithm for rapid outbreak detection, J Am Med Inform Assoc, vol. 18, pp. 218-224, 2011.

[4] Neill, DB and Cooper, GF, A multivariate Bayesian scan statistic for early event detection and characterization, Machine Learning, vol. 29, pp. 261-282, 2010.

*Jialan Que
E-mail: jialan.que@gmail.com