Development and Evaluation of a Data-adaptive Algorithm for Univariate Temporal Biosurveillance Data

Yevgeniy Elbert[1], Howard S. Burkom[1], Sean P. Murphy[1], and Galit Shmueli[2]

[1]*The Johns Hopkins University Applied Physics Laboratory*

[2]*Decision & Information Technology Dept. R.H. Smith School of Business, University of Maryland College Park*

OBJECTIVE

The objective is to develop and evaluate an operational alerting algorithm appropriate for the variety of time series behavior observed in biosurveillance data. The Holt-Winters (H-W) implementation of generalized exponential smoothing, comparable to complex regression models in predictive capability and far easier to specify and adapt, is built into a robust detection method.

BACKGROUND

Numerous recent papers have evaluated algorithms for biosurveillance anomaly detection. Common essential problems in the disparate, evolving data environment include trends, day-of-week effects, and other systematic behavior. Public health monitors have expressed the need for modifiable case definitions, requiring monitoring of time series that cannot be modeled in advance. Thus, automated algorithm selection is required. Recent research showed superior predictive performance of the H-W forecasting method [1] compared to regression based predictors applied to syndromic data [2,3]. This effort discusses extension to a practical monitoring tool, including selection from parametric and initialization settings based on limited data history, selection criteria for routine updating, specification of confidence limits, and validation of the resulting algorithm.

METHODS

To extend H-W forecasting for anomaly detection, we determined a small collection of useful parameter sets with grid search techniques minimizing median absolute deviations. Simple schemes for initial value selection were compared for both rich and sparse data streams. Features were added for special handling of holidays and unexpected outliers for both modified predictions and selective updating.

For comparison and validation, we applied the derived H-W algorithm, an adaptive regression approach, and a popular sliding z-score-based method to a representative collection of authentic outpatient data covering 1402 days of clinic visit counts. Receiver operating characteristic curves were obtained from Monte Carlo runs in which realistic outbreak signals were added to the data. Both predictive capability across multiple data types and detection probabilities at practical background alert rates were used to compare algorithm performance.

.

RESULTS

H-W based algorithm results were comparable to those of the adaptive regression method and superior to those of the simpler methods for the data streams tested.. Table 1 gives sample detection probabilities obtained for practical background alert rates using stochastic signals theoretically peaking at three times the standard deviation of the detrended data.

Syndrome	Data Mean	Holt-Winters	Adaptive Regression	Sliding Z-score
Resp	334.6	0.880	0.748	0.651
Lower GI	54.9	0.862	0.767	0.628
Botulism-like	3.2	0.252	0.332	0.210

CONCLUSIONS

The developed H-W-based algorithm is a versatile tool that provides reliable, sometimes near-optimal alerting capability and requires little data history for appropriate parameter specification and initialization.

REFERENCES

1. Burkom, H., Murphy, S.P., and Shmueli G, Automated Time Series Forecasting for Biosurveillance, 2007, Statistics in Medicine, available at http://www3.interscience.wiley.com/cgi-bin/fulltext/114131913/PDFSTART.
2. Brillman JC, et.al. Modeling emergency department visit patterns for infectious disease complaints: results and application to disease surveillance, BMC Medical Informatics and Decision Making 2005, 5:4, pp 1-14
3. Reis BY, Mandl KD, Time series modeling for syndromic surveillance (2003). BMC Medical Informatics and Decision Making 2003.