

Kalman Filter Based Anomaly Detection Technique for Biosurveillance Applications

Karen Cheng, David Crary

Health Effects and Medical Response Group, Applied Research Associates, Inc., Arlington, VA

Jaideep Ray, Cosmin Safta

Sandia National Laboratories, Livermore, CA

OBJECTIVE

This paper describes a novel approach to anomaly detection in medical data streams. This technique has been developed as part of an outbreak identification and classification system.

BACKGROUND

The idea behind syndromic surveillance systems is to provide early warning of disease outbreaks by detecting abnormally high levels of morbidity in the collected data. There are typically two well known problems in detecting anomalies in syndromic time series data. First, the data typically has daily, weekly, and seasonal variations which complicate anomaly detection schemes. Second, syndromic data sets typically include missing points which may occur at both regular and irregular time intervals.

In traditional time series modeling schemes, such as ARMA or ARIMA, it is not possible to include the known cyclic variations of the data in the formulation of the model itself. Typically, they would require that some sort of ad-hoc pre or post filtering is applied to the data. This is undesirable as the results of anomaly detection are too sensitive to the method of filtering approach applied^[1]. Also, these methods do not have well established techniques for handling missing data points, usually relying on ad-hoc techniques such as interpolation or smoothing of the original data. In this work, we present an approach that includes the dynamics of the process of interest within the anomaly detection model itself, and which naturally handles known cyclic terms and missing data points. This anomaly detection scheme has been used as the front-end trigger to a Bayesian syndromic classification module in a disease identification system, which is described in another paper^[2].

METHODS

Our anomaly detection approach uses a state space (Kalman Filter) technique that fits a temporally varying model to the data, and which can include weekly cycles and seasonal effects. This model was originally developed for analysis of economic time series data and we have adapted it to the requirements of syndromic data. Similar models have been used in analysis of medical data^[3], and epidemiology^[4], but these studies do not include weekly or seasonal effects.

The model calculates a level of stochasticity, in the form of a time-varying covariance for the data, which we use to calculate the deviations of actual data measurements from expected values. The approach is disease-agnostic and can be applied to any time-series data stream, provides one step ahead prediction capabilities with a formal statistical interpretation, and provides a consistent framework for handling missing data values.

RESULTS

In this paper, we present preliminary results showing detection of inhalational anthrax and pneumonic plague outbreaks. These outbreaks have been simulated using a large-scale, network-based model to calculate the disease dynamics, along with a realistic model of the background morbidity. We demonstrate our approach using a scenario where a rapid rise in respiratory distress is detected using our Kalman filter anomaly detector. Simulations are done to show the effect of missing data on the timeliness of the detection, as well as the effect of background morbidity levels.

CONCLUSIONS

We have proposed a new technique for anomaly detection of syndromic time series data. The technique is robust to missing data points, and provides a natural method of incorporating known cyclic variations. We have tested this approach on simulated data representing a wide range of conditions, and are currently using it as the front end to a Bayesian syndromic classification model described in a different paper [2].

REFERENCES

- [1] Bloom, R, Buckeridge D, Cheng KE *J. Am. Med. Inform. Assoc.* 2007, 14:76-85
- [2] Ray, J et al. this conference.
- [3] Smith, AFM, West M. *Biometrics* 1983, 39: 867-878
- [4] Stroup DF, Thacker SB, *Epidemiology* 1993, 4(5): 435-443