Recursive Least Squares Prediction of Syndromic Data for Surveillance

A. H. Najmi, D.Phil.

The Johns Hopkins University Applied Physics Laboratory, Laurel, MD 20723

OBJECTIVE

This paper describes a method to predict syndromic data for surveillance of public health using the method of recursive least squares and a new method of correcting for the day of week effect in order to have a prediction of the background upon which detections of actual events can be computed.

BACKGROUND

Syndromic surveillance of public health has been the subject of intense study in the last few years. Timely detection of an outbreak threat is a major goal of surveillance of public health data, and so special emphasis must be placed on a monitoring system that can reliably forecast the expected background data upon which a useful detection system can be built. Although an ultimate detection system will probably operate in a multi-stream environment, useful techniques must be able to work on single stream data sets, such as syndromic data. In a previous publication [Ref. 1] I addressed the multi-stream data adaptive prediction and detection problem using a novel prediction method based on recursive least squares. In the present paper I am exclusively interested in using the above method for predicting a "background" from a single syndromic data channel. A major issue in any data prediction method is the "equalization" problem (otherwise known as the weekend and day-of-week effects). Detectors based on any such prediction methods will be adversely affected if the equalization problem persists. The present paper addresses the latter issue systematically and integrates it in an adaptive prediction methodology based on the work in Ref. 1

METHODS

When processing syndromic data one must correct the counts for three effects: holidays, weekends and day of week. Without proper equalization, prediction results are practically worthless. My approach to this issue is a two-step method. Assume the weekend days and weekdays to be represented by two Gaussian random variables X and Y, with associated distribution functions $f_X(x)$ and $f_Y(y)$. A first set of corrections are made by considering a linear transformation between the two random variables $Y = aX + b$, whose coefficients are uniquely determined from estimates of the means and the variances of the two random variables.

The second step in the method consists of calculating the sorted values of the counts for every day of the week. The middle portion of these curves is fit with a straight line and the data are transformed to a mean line using linear transformations of the fitted lines. All correction coefficients are computed adaptively and are stored so that data can be transformed to original counts after prediction results are obtained. Adaptively corrected data are used to predict "background" counts.

RESULTS

The new integrated method is applied to syndromic data to predict a reliable background for detection purposes. The effectiveness of the new method is first measured by comparing results reported in Ref. 1 (using an ad-hoc method of equalization) with the present method. Next I apply the method to data consisting of a 3 year period (between 1994 and 1997) of respiratory syndrome patient counts reported by military hospitals in the Southwestern United States. Prediction goodness-of-fit measures will be displayed and discussed. Results will be compared to those of the current CDC C2 and W2 methods the results.

CONCLUSIONS

Novel recursive least squares adaptive prediction methods combined with novel techniques of addressing the day of week effect are shown to produce effective forecasts of background data for syndromic surveillance.

REFERENCES

[1] Najmi A. H. and Magruder, S. F., **An adaptive prediction and detection algorithm for multistream syndromic surveillance**, BMC Medical Informatics and Decision Making, **2005, 5:33**

Further Information: A. H. Najmi, najmi@jhuapl.edu