ISDS 2015 Conference Abstracts

Enhancing Biosurveillance Specificity Using Praedico™, A Next Generation Application

Alireza Vahdatpour[2], Cynthia A. Lucero-Obusan[1], Chris Lee[2], Gina Oda[1], Patricia Schirmer[1], Anosh Mostaghimi[1], Farshid Sedghi[2], Payam Etminani[2] and Mark Holodniy*[1]

[1]Office of Public Health, Department of Veterans Affairs, Palo Alto, CA, USA; [2]Bitscopic, Inc., Redwood City, CA, USA

Objective

To compare syndromic surveillance alerting in VA using Praedico™ and ESSENCE.

Introduction

Many methods to detect outbreaks currently exist, although most are ineffective in the face of real data, resulting in high false positivity. More complicated methods have better precision, but can be difficult to interpret and justify. Praedico™ is a next generation biosurveillance application built on top of a Hadoop High Performance Cluster that incorporates multiple syndromic surveillance methods of alerting, and a machine-learning (ML) model using a decision tree classifier [1] evaluating over 100 different signals simultaneously, within a user friendly interface.

Methods

513 million VA patient records, incorporating over 5 years of syndromic surveillance, were analyzed for identical syndromic surveillance groupings from outpatient ICD-9 diagnosis codes. The Praedico™ ML layer was trained by utilizing hundreds of false positive and true positive syndromic alerts. To guarantee high detection recall, the Praedico™ algorithm leverages many known detection algorithms, including versions of CDC, CUSUM, EWMA, and regression models [2]. The ML model combines these models and uses additional time series features to detect anomalies and user feedback received on previous alerts (high confidence alerts). Low confidence alerts, which many not trigger an alert but show a natural increase or normal divergence from the mean in data distribution are presented for users review. To facilitate alert interpretation, a natural language interface provides human interpretable messages relaying event significance by comparing historical values collected from the same facilities. Praedico™ syndromic alerts were compared to those generated by VA ESSENCE for the period of June 2014 thru May 2015.

Results

Praedico™ alerts were significantly lower compared to ESSENCE generated alerts (Table 1). This was expected as Praedico™ leveraged user feedback to enhance anomaly detection and improve precision of outbreak detection. Both Praedico™ and ESSENCE categorized alerts as high and low confidence groups. In both systems, higher deviation levels from expected values resulted in high confidence alerts. 62% of Praedico™ alerts directly correlated with ESSENCE alerts, suggesting that although the total number of Praedico alerts was smaller, they were not simply a subset of ESSENCE alerts. Praedico™ demonstrated higher seasonal sensitivity, adjusting for seasonality using historical and seasonal information, while ESSENCE alerts were more uniformly distributed over the year (Figure 1). The increased December and January alerts were due to ILI syndrome alerts, likely due to elevated influenza activity. Respiratory, Fever, and ILI syndrome groups had the highest number of alerts, and were significantly higher with ESSENCE (Figure 2).

Conclusions

Praedico™ demonstrated improved precision of surveillance syndrome clusters compared to VA ESSENCE by reducing the number of alerts. By reducing alerting fatigue, users' sensitivity to computer-generated alerts remain high, which in return results in further usage, feedback, and more gradual improvement in the algorithm's output (specificity and sensitivity), adapting to the interest of users.

Table 1. Total number of VA syndromic alerts generated by Praedico Biosurveillance and ESSENCE

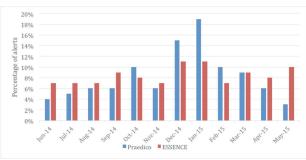| June 2014-May 2015	Praedico	ESSENCE
High Confidence Alerts	771	15,846
Low Confidence Alerts	3,452	21,502

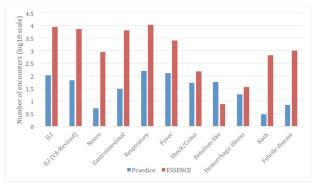Figure 1. Percentage of VA total alerts by month for Praedico and ESSENCE

Figure 2. Syndromic distribution of Praedico and ESSENCE alerts

Keywords

ESSENCE; biosurveillance; Veterans; Big Data

References

1. Classification and Regression Trees. L. Breiman, et al., Wadsworth, Belmont, CA, 1984.
2. Algorithms for rapid outbreak detection: a research synthesis. Buckeridge, David L. et al., Journal of Biomedical Informatics, Volume 38, Issue 2, 99 – 113.

*Mark Holodniy
E-mail: holodniy@stanford.edu