ISDS 2016 Conference Abstracts

Soda Pop: A Time-Series Clustering, Alarming and Disease Forecasting Application

Jeremiah Rounds*, Lauren Charles-Smith and Courtney D. Corley

Applied Statistics and Computational Modeling, Pacific Northwest National Laboratory, Richland, WA, USA

Objective

To introduce Soda Pop, an R/Shiny application designed to be a disease agnostic time-series clustering, alarming, and forecasting tool to assist in disease surveillance "triage, analysis and reporting" workflows within the Biosurveillance Ecosystem (BSVE) [1]. In this poster, we highlight the new capabilities that are brought to the BSVE by Soda Pop with an emphasis on the impact of metholodogical decisions.

Introduction

The Biosurveillance Ecosystem (BSVE) is a biological and chemical threat surveillance system sponsored by the Defense Threat Reduction Agency (DTRA). BSVE is intended to be user-friendly, multi-agency, cooperative, modular and threat agnostic platform for biosurveillance [2]. In BSVE, a web-based workbench presents the analyst with applications (apps) developed by various DTRA-funded researchers, which are deployed on-demand in the cloud (e.g., Amazon Web Services). These apps aim to address emerging needs and refine capabilities to enable early warning of chemical and biological threats for multiple users across local, state, and federal agencies.

Soda Pop is an app developed by Pacific Northwest National Laboratory (PNNL) to meet the current needs of the BSVE for early warning and detection of disease outbreaks. Aimed for use by a diverse set of analysts, the application is agnostic to data source and spatial scale enabling it to be generalizable across many diseases and locations. To achieve this, we placed a particular emphasis on clustering and alerting of disease signals within Soda Pop without strong prior assumptions on the nature of observed diseased counts.

Methods

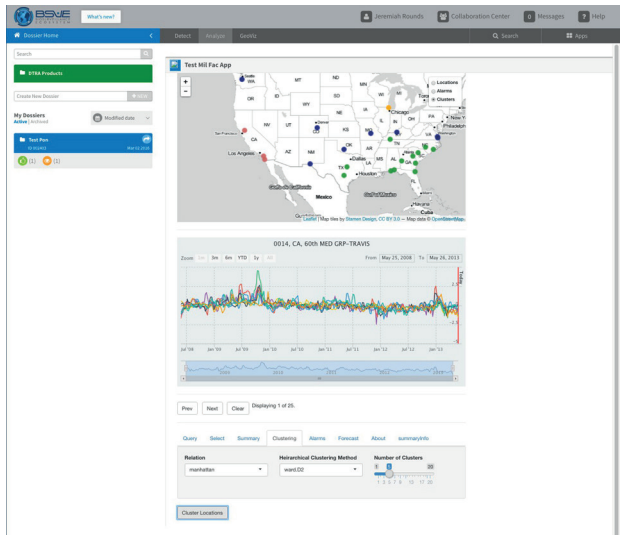
Although designed to be agnostic to the data source, Soda Pop was initially developed and tested on data summarizing Influenza-Like Illness in military hospitals from collaboration with the Armed Forces Health Surveillance Branch. Currently, the data incorporated also includes the CDC's National Notifiable Diseases Surveillance System (NNDSS) tables [3] and the WHO's Influenza A/B Influenza Data (Flunet) [4]. These data sources are now present in BSVE's Postgres data storage for direct access.

Soda Pop is designed to automate time-series tasks of data summarization, exploration, clustering, alarming and forecasting. Built as an R/Shiny application, Soda Pop is founded on the powerful statistical tool R [5]. Where applicable, Soda Pop facilitates non-parametric seasonal decomposition of time-series; hierarchical agglomerative clustering across reporting areas and between diseases within reporting areas; and a variety of alarming techniques including Exponential Weighted Moving Average alarms and Early Aberration Detection [6].

Soda Pop embeds these techniques within a user-interface designed to enhance an analyst's understanding of emerging trends in their data and enables the inclusion of its graphical elements into their dossier for further tracking and reporting. The ultimate goal of this software is to facilitate the discovery of unknown disease signals along with increasing the speed of detection of unusual patterns within these signals.

Conclusions

Soda Pop organizes common statistical disease surveillance tasks in a manner integrated with BSVE data source inputs and outputs. The app analyzes time-series disease data and supports a robust set of clustering and alarming routines that avoid strong assumptions on the nature of observed disease counts. This attribute allows for flexibility in the data source, spatial scale, and disease types making it useful to a wide range of analysts



Soda Pop within the BSVE.

Keywords

BSVE; Biosurveillance; R/Shiny; Clustering; Alarming

Acknowledgments

This work was supported by the Defense Threat Reduction Agency under contract CB10082 with Pacific Northwest National Laboratory

References

1. Dasey, Timothy, et al. "Biosurveillance Ecosystem (BSVE) Workflow Analysis." Online journal of public health informatics 5.1 (2013).
2. http://www.defense.gov/News/Article/Article/681832/dtra-scientists-develop-cloud-based-biosurveillance-ecosystem. Accessed 9/6/2016.
3. Centers for Disease Control and Prevention. "National Notifiable Diseases Surveillance System (NNDSS)."
4. World Health Organization. "FluNet." Global Influenza Surveillance and Response System (GISRS).
5. R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
6. Salmon, Maëlle, et al. "Monitoring Count Time Series in R: Aberration Detection in Public Health Surveillance." Journal of Statistical Software [Online], 70.10 (2016): 1 - 35.

***Jeremiah Rounds**
E-mail: jeremiah.rounds@pnnl.gov