Enabling User-Driven Public Health Analyses through Automated Data Querying

Raj Ashar

Johns Hopkins University Applied Physics Laboratory, Laurel, MD

BACKGROUND

Public health officials are now receiving more data than ever in electronic formats, and also stand to benefit more than ever from ongoing advances in the medical and epidemiological sciences. At the same time, this growing body of knowledge as well as volatile world events present an increasingly complex set of threats to population health. As a consequence, public health officials are finding that they need to ask many more, and more complex, questions of their data in order to keep sight of the state of the public's health. Most current disease surveillance systems enable users to ask many different questions of health data, but are limited in that users can only extract results one question, or query, at a time.

OBJECTIVES

Develop an Automated Data Query (ADQ) tool to allow public health officials to easily extract batches of raw medical encounter data using custom queries that the officials themselves set up. Additionally, the tool shall be capable of running anomaly detection algorithms against the raw data and returning the statistics [1]. Users shall be able to perform their own analyses on the data and/or the statistical results after using the tool to collect the information efficiently. The tool will help them spot trends of interest that may be specific to their own jurisdictions.

METHODS

To satisfy these objectives, an ADQ tool was prototyped as an enhancement to an existing operational surveillance system [2, 3]. This ADQ tool enables users to extract batches of medical records and statistical results to files without needing to repetitively query the disease surveillance system and copy or download the results by hand. In this ADQ implementation, users provide query information and cached results from prior queries to a simple input user interface. The application then retrieves and runs detector algorithms on the requested data and returns the results in a single merged dataset.

RESULTS

Based on a study that used ADQ to extract data the ADQ concept has proven itself to be a useful tool [4]. During this study, the ADQ's ability to re-use cached query results significantly improved the overall query execution time required for multiple retrospective analyses.

CONCLUSIONS

We have made a public health dataset more accessible for users to perform their own analyses by developing an efficient tool that allows users to automatically extract specific batches of medical record and statistical data. This Automated Data Query tool can be applied for several different analytic purposes, such as:

- research studies, including advanced data fusion efforts
- automated surveillance systems
- *ad hoc* querying.

Ultimately, providing flexible access to data and information is paramount in enabling users to fully understand their populations' health.

REFERENCES

1. H. S. Burkom, Y. A. Elbert. "Development, Adaptation, and Assessment of Alerting Algorithms for Biosurveillance." *Johns Hopkins University Applied Physics Laboratory Technical Digest*, 2003; 24:335-42.

2. J. S. Lombardo, D. Buckeridge *et al*. "Disease Surveillance: A Public Health Informatics Approach." Wiley-Interscience, 2007.

3. J. Lombardo, H. Burkom, E. Elbert, S. F. Magruder, S. Lewis, W. Loschen, J. Sari, C. Sniegoski, R. Wojcik, and J. Pavlin. "A Systems Overview of the Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE II)." *Journal of Urban Health*, Volume 80(2)S1: i32-i42, 2003.

4. Zaruhi R. Mnatsakanyan, PhD, Sean Patrick Murphy, Raj Ashar, Howard Burkom, PhD. "Hybrid Probabilistic Modeling and Automated Data Fusion for Biosurveillance Applications". *Advances in Disease Surveillance*, Volume 2: 161, 2006.