International System for Total Early Disease Detection (InSTEDD) Platform

Taha A. Kass-Hout, M.D., M.S., Nicolas di Tada
InSTEDD, Palo Alto, California

OBJECTIVE

This paper describes a hybrid (event-based and indicator-based) surveillance platform designed to streamline the collaboration between domain experts and machine learning algorithms for detection, prediction and response to health-related events (such as disease outbreaks).

BACKGROUND

Over the last decade, the majority of the designs, analyses and evaluations of early detection [1] (or biosurveillance) systems have been geared towards specific data sources and detection algorithms. Much less effort has been focused on how these systems will "interact" with humans [2]. For example, consider multiple domain experts working at different levels across different organizations in an environment where numerous biosurveillance algorithms may provide contradictory interpretations of ongoing events. This paper discusses the anticipated contribution of social networking, machine learning and collaboration techniques to address these emerging issues by drawing upon methods, models and technologies that have been proven to work in other challenging domains.

METHODS

The platform consists of several high-level modules, including: 1) Data gathering, 2) Automatic feature extraction, data classification and tagging, 3) Human input, hypotheses generation and review, 4) Predictions and alerts output, and 5) Field confirmation and feedback. The **data gathering** module allows users to collect information from several sources (SMS messages, RSS feeds, email list (e.g., ProMed), documents, web pages, electronic medical records, animal disease data, environmental feed, remote sensing, etc.). The **automatic feature extraction, data classification and tagging** module is an extensible architecture that allows the introduction of machine learning algorithms (e.g., Bayesian). These components extract and augment the features (or metadata) from multiple data streams; such as: source and target geo-location, time, route of transmission (e.g., person-to-person, waterborne), etc. In addition, these components help detect relationships between these extracted features within a collaborative space or across different collaborative spaces. Furthermore, with human input, these components can suggest possible events or event types (e.g., at the earliest stages of a disease outbreak: "there is an unknown respiratory event, transmitted person-to-person, detected in location X, and with a certain spatio-temporal pattern"). The **human input and review** module is exposed as a set of functionalities that allows users to comment, tag, and rank the elements (positive, neutral, or negative). Additionally, users can generate and test multiple hypotheses in parallel, further collect and rank sets of related items (evidence), and model against baseline information (for cyclical or known events). The platform maintains a list of ongoing possible threats allowing domain experts to focus their field information and either **confirm or reject** the hypotheses created. That feedback is then fed into the system to update (increase or decrease) the reliability of the sources and credibility of the users in light of their inferences or decisions.

RESULTS

The platform synthesizes health-related event indicators from a wide variety of information sources (structured and unstructured) into a consolidated picture for analysis, maintenance of "community-wide coherence" [3], and collaboration processes. This helps detect anomalies, visualize clusters of potential events, predict the rate and spread of a disease outbreak and provide decision makers with tools, methodologies and processes to investigate the event. Presently, the platform and associated modules are being piloted for the Mekong Basin region in SE Asia.

CONCLUSIONS

In this paper we describe a platform that enables detection, prediction and response to health-related events through a collaborative approach that combines data exploration, integration, search and inferencing – providing more complex analysis and deeper insight. We believe that such a platform represents the next generation of early detection and response systems.

REFERENCES

[1] Ping Yan, Daniel Dajun Zeng, Hsinchun Chen: A Review of Public Health Syndromic Surveillance Systems. ISI 2006; 249-260.

[2] Thagard, P. Coherence, truth, and the development of scientific knowledge. Philosophy of Science 2007;74, 28-47.

[3] Billman, D., Convertino, G. Shrager, J. Massar, JP. and Pirolli P. The CACHE Study: Supporting Collaborative Intelligence. Paper presented at the HCIC 2006 Winter Workshop on Collaboration, Cooperation, Coordination. Feb. 2006.

Further Information:
Taha Kass-Hout, kasshout@instedd.org
http://www.instedd.org