Hybrid Probabilistic Modeling and Automated Data Fusion for Biosurveillance Applications

Zaruhi R. Mnatsakanyan, PhD, Sean P. Murphy, Raj J. Ashar, Howard Burkom, PhD
The Johns Hopkins University Applied Physics Laboratory

Objective

An intelligent information fusion approach is proposed to identify and provide early alerting of naturally-occurring disease outbreaks, as well as bioterrorist attacks, while reducing false positives. The proposed system statistically preprocesses information from multiple sources [1] and fuses it in a manner comparable with the domain expert's decision-making process. Currently, system users lower the false alarm rate by "explaining away" the statistical data anomalies with alternative hypotheses derived from external, non-syndromic knowledge. We seek to incorporate this heuristic decision-making into a probabilistic network that accepts the outputs of statistical algorithms in a hybrid model of domain knowledge and data inference.

Background

The increased threat of bioterrorism and naturally occurring diseases, such as pandemic influenza, continually forces public health authorities to review methods for evaluating data and reports. The objective of bio-surveillance is to automatically process large amounts of information in order to rapidly provide the user with a situational awareness. Most systems currently deployed in health departments use only statistical algorithms to filter data for decision-making. These algorithms are capable of high sensitivity, but this sensitivity comes at the cost of excessive false positives [2], especially when multiple syndrome groups and data types are processed.

Methods

A hybrid probabilistic model was built to process statistical anomalies with a Bayesian belief network (BBN) to estimate public health status in a fashion similar to the decision-making processes of the epidemiologist. A BBN, a type of probabilistic graphical model typically visualized as a directed acyclic graph with nodes representing random variables and directed edges representing probabilistic dependencies, offers a compact representation of the relationships among all variables in the graph. Expert knowledge is embedded into the BBN at two levels: First, the gross structure of the network captures an expert's view of the interdependencies of the causal factors in the network. Second, the conditional probability table attached to each node quantifies the probabilistic relationships between connected variables.

The data inputs used by the system were collected by the Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE) system. These data include records of classified and processed chief complaints, ICD-9 diagnosis codes, physician office visits, and pharmacy data. In addition, information such as weather, seasonal patterns, calendar events, traffic patterns, etc. is collected. All of this information is potentially available to the user via the ESSENCE interface or sources available via the internet.

Our system is a decision support system and not a decision making system. It is not intended to replace the domain experts; instead, the system analyzes available information and generates alarms that require the attention of a domain expert to decide if additional investigation is necessary. As biosurveillance is a strongly region dependent domain, the system must learn from data and automatically reconfigure to adapt to a new location. However, the design of the inference process is fixed and represents validated domain knowledge and decision-making protocols.

Conclusions

With an increasing concern over emerging infectious diseases, both natural and manmade, efficient and reliable health monitoring is critical and the issue of reducing the high false positive rate becomes important. The prototype system in development, described by this abstract, offers the potential to address the false alarm issue while increasing sensitivity levels and reducing the workload of health monitors. This comprehensive capability is needed to bolster public health acceptance of biosurveillance systems by making the complex information environment more manageable and by achieving more robust performance.

References

[1] H. Burkom: Development, Adaptation, and Assessment of Alerting Algorithms for Biosurveillance
[2] D. Siegrist, J. Pavlin: Bio-ALIRT Biosurveillance Detection Algorithm Evaluation.