

A BIO-SURVEILLANCE SYSTEM USING EFFICIENT ALGORITHM FOR INFORMATION SHARING APPLICATIONS

Shobhanjaly P. Nair, K. Silambarasan, N. Bindhu and R. Srinivasan

Department of Computer Science and Engineering,
Vel Tech Dr. RR and Dr. SR Technical University, Chennai, Tamil Nadu, India

Received 2014-01-23; Revised 2014-02-04; Accepted 2014-02-27

ABSTRACT

Bio-Surveillance represents a health information system for public that stores and integrates health related information and data across a variety of data. Bio-surveillance strategies are projected for detection of disease outbreaks from existing diagnostic data, but their performance is typically not understandable. The focus of the paper is to develop a bio-surveillance system which can serve the aim of biosurveillance by collecting health connected data from medical experts or individual person and check for relative health modifications to observe any silent change in health condition by employing a anomaly detection algorithm and time series analysis. Simulation is employed to check practical performance with speculative results. This data is analyzed daily by the system, permitting epidemiologists to envision potential epidemics via mapping tools. An alert message is send through message gateway or electronic mail gateway. Reports are generated in graphs and maps that show the outbreaks probability at specific location and by informing through electronic mail can help in alerting the user an early outbreak of any changes in specific surroundings.

Keywords: Bio-Surveillance, Bayesian Rule, Anomaly Detection Algorithm, Time Series Analysis

1. INTRODUCTION

Bio-Surveillance is defined as a scientific method that monitors the environment for biological agents that cause diseases; detects the disease and characterizes outbreaks of such disease. The ability to recognize quickly, the signs of an intentional bio-attack or naturally occurring outbreak is crucial to safeguard the general public. The term bio-surveillance is a combination of the terms illness surveillance and public health surveillance, each centered on the methods for data assortment and analysis for disease detection. The principle behind bio-surveillance is to provide early notification of disease outbreaks, this helps to reduce the severity of the impact caused by the disease. Bio-surveillance could be a continuous method, that take from continuous data assortment, aggregating, conducting analysis (Koteeswaran *et al.*, 2012), confirms the outbreaks and responses, with a feedback circuit back to the aggregation method.

Primary attention of the sector of bio-surveillance focuses on outbreaks and investigations.

Bio-surveillance is the monitoring of a population for changes against a standard norm that might indicate a silent biological condition. Intentionally created pathogens or naturally occurring infectious diseases will act as a bio-weapon and therefore the traits/changes it brings in us may be too slow to identify initially. In the event of a silent bioterrorism attack, the present system is unable to identify the changes at the initial stages and therefore the risk it possesses is threatening.

As in **Fig. 1** a general bio-surveillance system will have administrator at one end and health worker or any user who provides details of specific location and health conditions prevailing. bio-surveillance includes information about human health that can be preliminary and indicative of health issues, like groups of disease symptoms (syndromes) and laboratory check orders. This information is required to prevent and control the disease.

Correspondent Author: Shobhanjaly, P. Nair, Department of Computer Science and Engineering,
Vel. Tech Dr. RR and Dr. SR Technical University, Chennai, Tamil Nadu, India

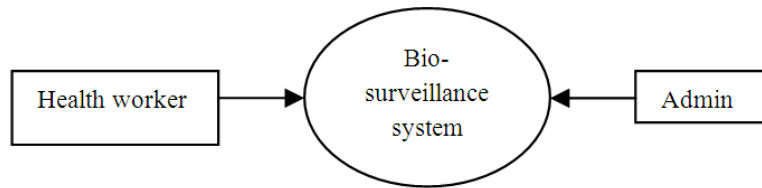


Fig. 1. General representation of bio-surveillance system

2. LITERATURE REVIEW

Lotze and Shmueli (2008) have proposed a system for detecting disease outbreaks from already diagnosed data and illustrate the association involving forecast accuracy and the detection outbreaks. The existing detection methods for bio-surveillance are forecasting component coupled with a monitoring or detection component, which helps in better forecasting and detection. To overcome difficulties like regular, dynamic and auto correlated character of the health data being monitored, algorithm does forecasting and then monitors the errors using control charts. The current bio-surveillance data are usually monitored daily at regular interval.

Dubrawski *et al.* (2009) has explained the concept of t-cube web interface which is a basic device to envision and influence massive scale variable statistic datasets. The interface allows the client to execute composite queries promptly and to run various types of statistical tests on the loaded information. It will show its effectiveness in a significant application scenario: Real-time bio-surveillance system designed to support quick detection and reduce bio-medical fear in developing countries. The benefits of this technology are leveraged in RTBP. The capabilities of the device comprises, idea of disease data, navigation through different levels of data aggregation and selected statistical analyses including spatial scan for early detection disease outbreaks.

Parikh and Javid (2009) have discussed a work based on record keeping which can be efficient and flexible like current information tools. A mobile phone is the perfect bridging device. The cam mobile document is the processing system, in which a mobile phone is used for capturing the image. The system will process paper forms containing cam-shell programs-embedded commands that are decoded from the captured electronic image. Cam unites a user interface, a programming language and a delivery mechanism for accessing networked information services.

Woodall *et al.* (2006) has proposed control charts applications such as Cumulative Sum Method (CUSUM)

and Exponentially Weighted Moving Average (EWMA) for industrial practitioners that are applicable in health-care monitoring and public- health surveillance. Industrial practitioners are made to consider the usefulness of risk-adjustment and regression adjusted variables control charts, which proves to be very helpful in health-related applications. The need for better health concern is well established and the significance of surveillance in public health care is growing. Practitioners and researchers in industrial statistics can make some further contributions to the theory and purpose of health-related surveillance.

Bradley *et al.* (2005) has briefed about Bio-Sense, which helps in avoiding any kind of biologic terrorism attack and support early detection other health issues. Bio-sense does early detection of outbreak by providing the information similar to real-time reporting. Bio-sense collects data from all quarters. The application summarizes and presents analytical results and data visualizations.

Brownstein *et al.* (2009) discusses about, internet being critical medium for every individual to get information about health issues and diseases outbreaks. Also helps in reducing the time to detect an outbreak. Excess data, fake reports, lack of signal specificity and interference from external forces will limit the utilization of internet technology. The Internet serves as effective channel for interaction and it is health care professionals and the public who must determine how to use this channel for surveillance and control the emerging diseases outbreak.

Burkom *et al.* (2004) discussed a system with function of spatial and temporal data-aggregation strategies for multiple data streams using alerting algorithms to identify the health threat. Such a strategy was applied and evaluated for a composite, valid, multi-source, multiregional data records from a system-analysis exercise for the Defense Advanced Research Project Agency (DARPA). The achieved result indicates that a bio-surveillance system can provide a early alert than customary public health monitoring systems.

Gow *et al.* (2010) have explained Real-Time Bio-surveillance Program (RTBP), serves as a multi-partner

research initiative and helps to improve early detection and warning of disease outbreaks and discuss the technology used among health workers. Since internet access at remote area is unstable, mobile phones are used for frontline data reporting. The mobile phone offers digitizing patient records and helps to create an end-to-end bio-surveillance system to support RTBP in developing countries. This study achieves formative assessment of the digitization process using analysis framework based latency, quality, reliability and portability.

Zarui *et al.* (2009), introduces fusion algorithms to enhance disease surveillance systems with Bayesian decision support capabilities. A detection system is introduced and analyzed using records of outpatients. Data oddity was identified and time offsets were distributed between various data streams events. The Bayesian Network was used to fuse data from multiple sources and identify epidemiological diseases, which exhibited high specificity and sensitivity. Further more implementation and validation are required to standardize data fusion and make it robust all types of data and health threat.

3. BIO-SURVEILLANCE SYSTEM

In the existing bio-surveillance system, diagnosed health data are monitored for the purpose of analyzing and detecting health related issues. The bio-surveillance system database typically consist of weekly or daily collected data regarding illness reported at health centers, daily sales of medicines at pharmacy and other relevant data that helps in detecting early mark of a disease occurrence. Outbreaks of disease may be due to terrorist driven attack, or naturally occurring epidemics. In both the type of outbreak the main aim is to alert public about the situation prevailing. The study aims at developing a system which will serve the purpose of bio-surveillance by collecting health related data from population (people of a particular area) and check for relative health changes to detect any silent change in health condition and inform public through graphical and other alerts or triggers. The data is analyzed daily in a system, by epidemiologists. The system detects the outbreak of a disease and sends information to user stored in the central database through the global system for mobile cellular network over the GPRS or through electronic mails. Also a website is developed where users can login and get the information regarding the outbreak of disease, either through the maps or tables or graphs displayed in the website.

The proposed work will take the relative data of health condition of people either through mobile

interface or user can update their details in the website by filling the form displayed. This information about the user is stored in the system repository and can be retrieved when required. The information is also updated at regular intervals. The system analyzes the data collected from user by comparing it with existing data (already stored information) like individual's age, location, gender, profession, symptoms in the database collected from user and healthcare centers. Also monitors the environmental change every second by continuously analyzing the updated details from the user. This helps in detecting any outbreaks at early stage itself. The report of outbreaks is displayed on the web interface by highlighting the particular location in the maps through alert pins, where disease outbreak is expected. The reports of outbreak are also displayed in charts and graphs form, so that the percentage and severity of the outbreak can be easily understood by the user. Based upon the details collected from the user, alert message is send to respective user of particular age, gender, profession, location, previous health history. The alerts are sent through short message service or electronic-mail to individual user.

The proposed algorithm is used for analyzing the different conditions and criteria based on which detection of any outbreak is predicted. Also a Time Series Analysis used to analyze the data collected with respect to data trends and patterns.

3.1. Proposed System

The key aspect of proposed algorithm is using Bayesian rule. The proposed algorithm is an unambiguous method where each individual's details in the population are considered as a model. The models of these individual nodes are used to evaluate the information about any disease outbreaks. Each and every detail of an individual is taken as a separate data to be analyzed, so that the algorithm can analyze by comparing each data with the existing data stored in the database and arrive at an accurate outcome (outbreak detection). By taking each individual as a sub-network, it is possible to coherently represent different types of background data in the model. Using anomaly detection algorithm the data collected from population and disease related details collected from health care centers are compared and analyzed based on, individual's input (age, gender, profession, location, symptoms), climatic conditions and existing disease outbreak. By analyzing the entire aspects, disease outbreak is predicted. As for objective flexibility, the individual

modeling allows new forms of data and evidence to be included into the model readily.

Anomaly Detection Algorithm represents the probabilistic relationships between population of particular area and symptoms or disease reported. The algorithm will detect the disease outbreak and its location more specifically based on the individual's details and disease symptoms collected.

Permits specification of previous information and belief:

- Knowledge about outbreak diseases
- Belief regarding whether, when and how an outbreak can occur, based on inputs collected from individual and disease details from healthcare centers

This yields inference as:

- P (hypothesis | data), which can be used directly in a decision analysis about what to do
- Other statistics of interest, like the expected variety of people infected in a probable outbreak scenario based on age, gender and profession are all calculated
- A value is set as threshold value, if the resulting value is less than threshold value then there is no disease outbreak at the present time
- If it exceeds the threshold value then there are possibilities of various types of outbreaks, such as anthrax, small pox, plague and many others

3.2. Anomaly Detection Algorithm

Input: Population_{aggregation}, disease_{changes}

Output: Outbreak_{detection}

Population ← initialize population

(Population_{aggregation})

Analyze population, disease

Outbreak ← GetBestSolution (population)

While (StopCondition ())

Selected ← select fit solution (population, disease)

Model ← anomaly detection algorithm

(Population, disease)

For (i to Population_{aggregation})

Outbreak ← probabilistically construct solution (model)

End

Evaluate population (outbreak)

Population ← combine (population, outbreak)

End

Return (outbreak).

3.3. Time Series Analysis

Time series analysis is the method of using statistical techniques to mold and explain a time-dependent series of data points. Time series prediction is the method of employing a model to predict disease outbreak using annual, monthly, weekly or daily seasonal variations. Time series data features sequential ordering - this differs from typical data mining or machine learning applications, in which each data is a self-sufficient sample of a concept to be learned and the order of information points in a data set is unimportant. Major time-series-related tasks embrace query by content, anomaly detection, motif discovery, prediction, clustering, classification and segmentation.

Time-series data mining unveils various aspects of complication. The important issues occur from the high spatial property of time-series data and the problem of defining similarity measure based on human perception. With the growth of digital sources of data, time-series mining algorithms have to match progressively huge datasets.

4. SYSTEM ARCHITECTURE

The working process of a bio-surveillance system is represented in **Fig. 2** the system has a client or user, who updates their details and these details are stored in the information process. The details can be retrieved when required from the repository and the system will analyze the changes and represents the changes either through maps or charts to the user. An alert message is also sent to users either through mobile service or through e-mail interface. **Figure 3** represents the flow of a bio-surveillance system and its step by step process. It will give a clear and perspective view of system process and describes each module of the system.

4.1. Data Collection Subsystem

This subsystem collects the medical inputs from the public (individual person) or from the health workers on daily basis through web interface and mobile application.

4.1.1. Web Interface

In this user interface the individual person or the health workers enters the details pertaining to the patient, such as their location, gender, date of birth (age), profession, contact details (mobile number, mail id) and symptoms (if they have any) by logging on to the system and filling the online form. These details are stored in the database of admin and updated at regular intervals.

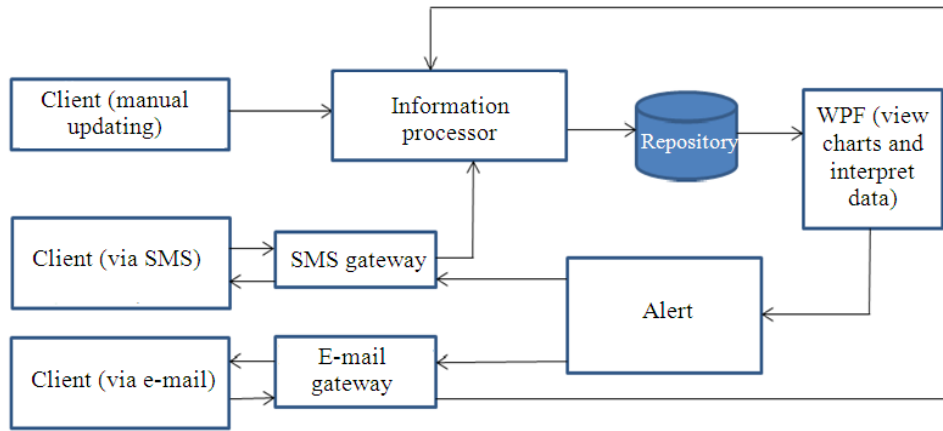


Fig. 2. System architecture of bio-surveillance system

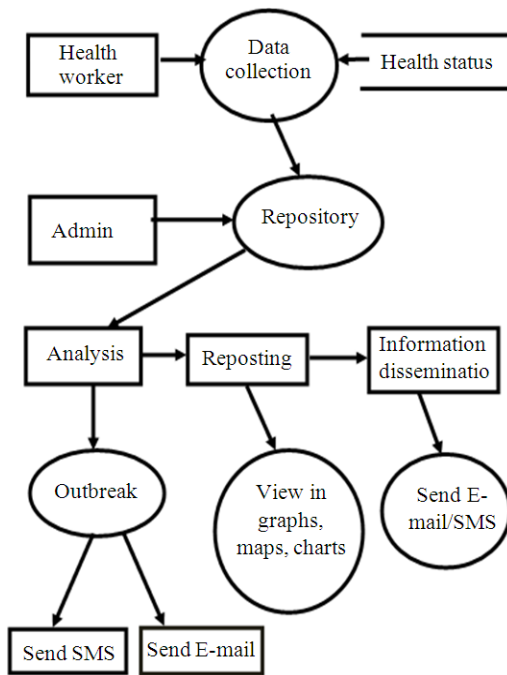


Fig. 3. Flow of bio-surveillance system

4.1.2. Mobile Application

The health worker or individual person enters the details pertaining to the person, such as their location, gender, date of birth (age), profession, contact details (mobile number, mail id) and symptoms (if they have any) by means of short messaging service. These details are stored in the database of admin and updated at regular intervals.

4.2. Analysis Subsystem

The analysis subsystem provides data analysis services such as time series analysis, outbreak detection and aggregation analysis using Anomaly Detection algorithm.

4.2.1. Time Series Analysis

Time series analysis is important to identify the trends and patterns of disease occurrences. Analysis the data stored in the database according to Time Series Algorithm and interprets the pathological outbreaks.

4.2.2. Outbreak Detection

Outbreak detection sub-module will alert the users in case of disease outbreak. This is done by analyzing the data gathered and will help the user to take active measures to prevent the spread of disease.

4.2.3. Aggregation Analysis

The aggregation analysis aids to identify patterns of outbreak based on gender, age, location and profession.

4.3. Reporting

The reporting module generates reports and displays it as tables, graphs and maps on the web interface.

4.3.1. Tables

This module presents data in repository in a tabular form. The tables are used when details about several attributes are involved. This table is displayed in the website indicating the percentage of outbreak of specific disease with respect to age, gender, location etc.

4.3.2. Graphs

This module represents data visually in the form of graphs and charts. The bar graphs helps user to compare and identify the disease outbreak at specific region.

4.3.3. Mapping Module

This module represents data spatially on a map. The heat maps will display the spatio-temporal information and is useful to spot a particular pathological outbreak is spreading. Also alert pins are displayed in maps. The alert pins will indicate the onset of outbreak at any particular location.

4.4. Information Dissemination

This module is used to raise alerts in case of any pathological outbreak. The alert is raised through the following subsystems. Analysis the given information (data of individual user) with Bayesian algorithm and checks the severity level of the disease and alerts user using.

4.4.1. Electronic-Mail Engine

Detailed information about the disease outbreak excepted and also preventive measures will be sent to user's electronic-mail id.

4.4.2. SMS Engine

The information about any disease outbreak excepted, are communicated to the user via Short Messaging Service (SMS).

5. RESULT ANALYSIS

The information of the user is collected through web interface or mobile interface. Users can provide their details by filling the online form and submitting it to the system. This information is stored in the database repository. The database also contains the list of excepted disease outbreaks which is analyzed based on details collected from users. Analysis is done based on many aspects which include the climatic conditions, environmental changes and users inputs.

Using Anomaly Detection algorithm, analysis can be done accurately. A threshold value is taken and then by using the algorithm, the probability of occurrences of outbreak is evaluated.

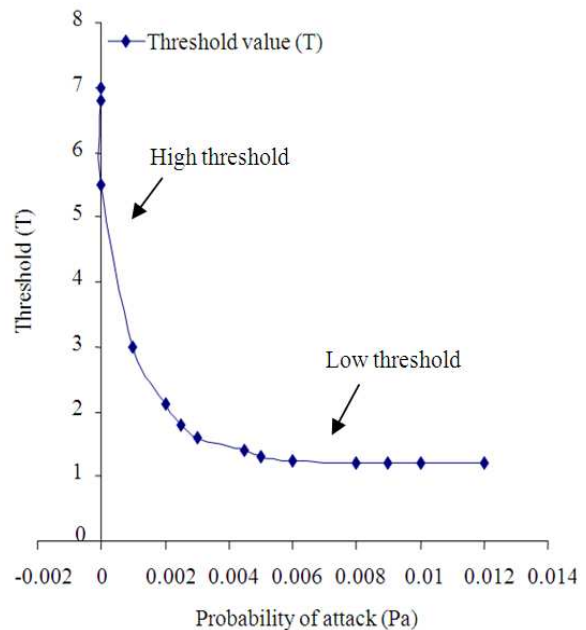


Fig. 4. Graph indicating threshold values versus probability of attack

The threshold values versus probability of attack graph is represented in **Fig. 4**. The threshold values are calculated by considering previous history of outbreak occurrences, based on the value the probability of outbreak is predicted. The probabilities of outbreaks are high at locations with low threshold values. The indication of outbreaks is given by displaying alert pins at that particular location.

The bio-surveillance system analysis the data and evaluates them, to detect the outbreak and these are reported to user by displaying the location of outbreak and also alerting the users by sending mails and messages.

6. CONCLUSION

The application is simple and cost effective biosurveillance system. The system provides improved results in detection, based on data records. The system could easily adapt to different changes in the environment and provide an efficient result in detecting the changes quickly. The application will alert people, in case of any disease outbreak through sending messages or e-mail system or maps indicating disease and its location displayed on the web interface. The system can be combined with other health surveillance system and develop an enhanced version of proposed system.

7. REFERENCES

- Bradley, C.A., H. Rolka, D. Walker and J. Loonsk, 2005. Biosense: Implementation of a national early event detection and situational awareness system. *Morb. Mortal Wkly. Rep.*, 54: 11-19. PMID: 16177687
- Brownstein, J.S., C.C. Freifeld and L.C. Madoff, 2009. Digital disease detection-harnessing the web for public health surveillance. *New Eng. J. Med.*, 360: 2153-2157. DOI: 10.1056/NEJMp0900702
- Burkom, H.S., Y. Elbert, A. Feldman and J. Lin 2004. Role of data aggregation in biosurveillance detection strategies with applications from ESSENCE. *MMWR Morb. Mortal. Wkly. Rep.*, 53: 67-73. PMID: 15714632
- Dubrawski, A., M. Sabhnani, M. Knight, M. Baysek and D. Neill *et al.*, 2009. T-cube web interface in support of real-time bio-surveillance program. *Proceedings of the 3rd International Conference Information Communication Technologies Development*, Apr. 17-19, IEEE Xplore Press, Doha, pp: 64-67. DOI: 10.1109/ICTD.2009.5426730
- Gow, G.A., V.P. Mary and N. Waidyanatha, 2010. Using mobile phones in a real-time biosurveillance program: Lessons from the frontlines in Sri Lanka and India. *Proceedings of the IEEE International Symposium Technology Society*, Jun. 7-9, IEEE Xplore Press, Wollongong, NSW, pp: 366-374. DOI: 10.1109/ISTAS.2010.5514617
- Koteeswaran, S., P. Visu and J. Janet, 2012. A review on clustering and outlier analysis techniques in datamining. *Am. J. Applied Sci.*, 9: 254-258. DOI: 10.3844/ajassp.2012.254.258
- Lotze, T. and G. Shmueli, 2008. On the relationship between forecast accuracy and detection performance: An application to biosurveillance. *Proceedings of the IEEE Conference Technologies Homeland Security*, May. 12-13, IEEE Xplore Press, Waltham, MA, pp: 100-105. DOI: 10.1109/THS.2008.4534431
- Parikh, T. and P. Javid, 2009. CAM: A mobile interaction framework for digitizing paper processes in the developing world. *Proceedings of the ACM Symposium User Interface Software and Technology*, (UIS' 09), pp: 119-122.
- Woodall, W.H., B.M. Adams and J.C. Benneyan, 2006. The use of control charts in healthcare and public-health surveillance. *J. Quality Technol.*, 38: 89-104. DOI: 10.1002/9781119940012.ch12
- Zaruhi, R.M., H.S. Burkom, J.S. Coberly and J.S. Lombardo, 2009. Bayesian information fusion networks for biosurveillance applications. *J. Am. Med. Inform. Assoc.*, 16: 855-863. DOI: 10.1197/jamia.M2647