G Model
ASOC 2462 1–11

Applied Soft Computing xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

Outbreak detection model based on danger theory

Q1 Mohamad Farhan Mohamad Mohsin*, Azuraliza Abu Bakar, Abdul Razak Hamdan

Data Mining and Optimization Research Group, Centre for Artificial Intelligence Technology, Faculty of Science & Information Technology, Universiti Kebangsaan Malaysia, Selangor, Malaysia

ARTICLE INFO

Article history:
Received 28 October 2013
Received in revised form 17 June 2014
Accepted 12 August 2014
Available online xxx

Keywords:
Outbreak detection
Artificial immune system
Danger theory
Dendritic cell algorithm

ABSTRACT

In outbreak detection, one of the key issues is the need to deal with the weakness of early outbreak signals because this causes the detection model to have has less capability in terms of robustness when unseen outbreak patterns vary from those in the trained model. As a result, an imbalance between high detection rate and low false alarm rate occurs. To solve this problem, this study proposes a novel outbreak detection model based on danger theory; a bio-inspired method that replicates how the human body fights pathogens. We propose a signal formalization approach based on cumulative sum and a cumulative mature antigen contact value to suit the outbreak characteristic and danger theory. Two outbreak diseases, dengue and SARS, are subjected to a danger theory algorithm; namely the dendritic cell algorithm. To evaluate the model, four measurement metrics are applied: detection rate, specificity, false alarm rate, and accuracy. From the experiment, the proposed model outperforms the other detection approaches and shows a significant improvement for both diseases outbreak detection. The findings reveal that the robustness of the proposed immune model increases when dealing with inconsistent outbreak signals. The model is able to detect new unknown outbreak patterns and can discriminate between outbreak and non-outbreak cases with a consistent high detection rate, high sensitivity, and lower false alarm rate even without a training phase.

1. Introduction

The aim of an outbreak detection system is to assist epidemiologists monitor the progress of a disease and raise an alert when there is an impending outbreak. An accurate and fast alarm notification system allows healthcare professionals to set up an early prevention plan before an outbreak spreads to a wider geographical area. If an outbreak is uncontrollable, it may lead to high death tolls and worse still; it kills without warning [1]. For example, the spread of severe acute respiratory syndrome (SARS) [2], influenza [3,4], and the avian virus [5] led to high mortality rates shortly after the first case was detected. Beside those viruses, there are seasonal types of diseases such as malaria, dengue, cholera and many more that can be classified as having the potential to escalate to outbreaks according to the World Health Organization and Centers for Disease Control and Prevention portal. One of the indirect effects of an outbreak is that it has a detrimental economic impact on a country. This is exemplified by the SARS and avian virus outbreaks in 2003, which led to a huge loss of around US$20 million for the tourism industry in most affected Asian countries such China, Hong Kong, Singapore and Vietnam because of an ineffective surveillance system [6]. For this reason, we need to be able to rely on an accurate detection system to monitor for signs of an emerging outbreak. The importance of having such a system has been emphasized in much of the literature, not least because it can help authorities to control the spread of an outbreak and reduce the mortality rate [4,7].

An outbreak is defined as a sudden spread of disease with a much greater number cases reported than expected over time, where the point of initial occurrence is a small area but quickly spreads to a wider geographic area [8]. A health surveillance system can monitor the progress of incidences of a disease over time: daily, weekly, or monthly. If there is an extraordinary pattern, this might indicate that an outbreak has started [9]. A sudden increase in the number of suspected cases from previous day is an indicator of a potential outbreak and it changes are viewed as abnormal characteristics; therefore an outbreak detection system is actually looking for anomalies in health data [10,11]. In outbreak detection studies, an outbreak is a collective anomaly where it requires more than one case to be detected before it can be labelled as an actual outbreak [12,13]. For example, if a single patient has been detected as having a certain disease, the alert system will not be activated until the number of similar cases has reached a certain default number.

Q2 * Corresponding author. +60 124773435.
E-mail addresses: farhan@uum.edu.my, mfarhan.mmohsin@gmail.com, mfarhan.mmohisn@gmail.com (M.F. Mohamad Mohsin), aab@ftsm.ukm.my (A. Abu Bakar), arh@ftsm.ukm.my (A.R. Hamdan).

http://dx.doi.org/10.1016/j.asoc.2014.08.030
1568-4946/© 2014 Published by Elsevier B.V.

Each disease has a different default number which is determined by epidemiologists. For instance, in the case of a disease such as dengue just more than a suspected cases difference from previous day will raise an alarm while cholera requires up to 20 registered cases before it can be classified as an outbreak [14]. However, when the spread of an outbreak is slow, the number of cases is not the key indicator and further epidemiological investigation is required [15]. Asthe notification of outbreak is based on collective anomalies that need to be tailored with same locality and time, this makes outbreak detection different from other detection tasks such as intrusion, fraud and fault.

In recent years, many researchers have worked on various outbreak detection methods. Those methods can be classified into three main approaches: statistical [1,9,16–18], artificial intelligence [15,19–25], and a combination of these two approaches into hybrid methods [26–28]. The statistical approach was applied in early outbreak detection models and to date it remains the preferred approach in many health surveillance systems [29,30]. One of the key issues that needs to be addressed in existing models is how to obtain a high detection rate while at the same time reducing the false alarm rate. However, due to the weakness of the outbreak signal which always behaves under uncertainties, existing systems produce an inconsistent false alarm rate during detection. Here, uncertainty refers to inconsistent outbreak signals, where the outbreak pattern frequently changes and differs between years. This means that a trained model has less capability in terms of robustness particularly when unseen outbreak patterns vary from the trained model. Robustness is also lost when detection algorithms require a sample from both an outbreak and non-outbreak session for model development. However, in practice, it is a challenging task to define a sample for an outbreak period because most of the data available relate only to non-outbreak periods.

Previously, outbreak detection models were mostly based on the univariate surveillance approach [18,31]. However, due to the weakness of the outbreak signal, the possibility of these models generating an imbalanced result between detection rate and false alarm rate is high because they rely on a single attribute [28]. As a possible solution to this problem of a weak signal, researchers have been investigating the use of multivariate surveillance by injecting the weak signal with a stronger signal, for instance by combining spatial and temporal data [1,16,32]. In addition, efforts have been made to combine multiple syndromic data to boost detection, for instance by combining emergency visit data with weather information, or clinical diagnosis results [33,34] and by investigating social network status and internet tracking search [22,35,36]. Since an outbreak is observed over time, distraction factors such as the seasonal event effect always disturb the detection results [9]. In relation to the data issue, each disease has a different outbreak definition which is determined by its the environment, government policies, outbreak rate and the medium for spread. Information on aspects such as outbreak duration (outbreak period), outbreak magnitude (rate of outbreak cases) and onset date are required and they are different depending on the disease [7]. The above-mentioned factors indirectly influence the effectiveness of an outbreak algorithm. Fig. 1 depicts the general outbreak detection concept.

In this conceptual model, an outbreak is defined based on three parameters: similar time, similar place, and huge number of reported cases. The aim of the outbreak detection model is to generate an accurate alarm which has high sensitivity and fast detection time. Since an outbreak is viewed as an abnormal period, anomaly detection analysis is used to detect the point at which data start to abnormally change. The quality of outbreak data and disease characteristic are the two factors that influence detection performance.

To improve the outbreak detection model's robustness when dealing with uncertain outbreak signals, we sought to identify an appropriate model from various disciplines, but especially from the field of biology, to detect early outbreak signals with good detection results. In this paper, an outbreak detection model based on danger theory is proposed. Danger theory is a bio-inspired method that replicates how the human body fights pathogens and the literature has shown that it can solve problems mainly related to the detection process. While not many works that apply danger theory have addressed the outbreak detection problem, danger theory has been successfully applied in intrusion [37,38], fraud [39,40], and fault detection problems [41] with good detection performance. Since the capability of danger theory as a good detector is proven in other areas, it motivates us to adapt the artificial immune system (AIS) as an outbreak detection model. The robustness of the immune system mostly lies in the ability of the dendrite cell to sense an early death cell (viewed as an outbreak signal), which can be replicated in an outbreak detection model to reduce the high false alarm rate. Moreover, danger theory offers a multivariate detection approach without relying on a training phase, which can improve a model's robustness.

In the proposed model, a signal formalization approach based on cumulative sum (*CUSUM*) is formalized and a cumulative mature antigen contact value (cMCAV) is proposed to suit the outbreak characteristic and danger theory. In experiments, the model is applied to two outbreak datasets; a real-world dengue outbreak and a synthetic SARS outbreak. To evaluate the model, four evaluation criteria are used: detection rate, specificity, false alarm rate, and accuracy. Then, the result is compared with three statistical control chart approaches: the CUSUM, Exponentially-weighted

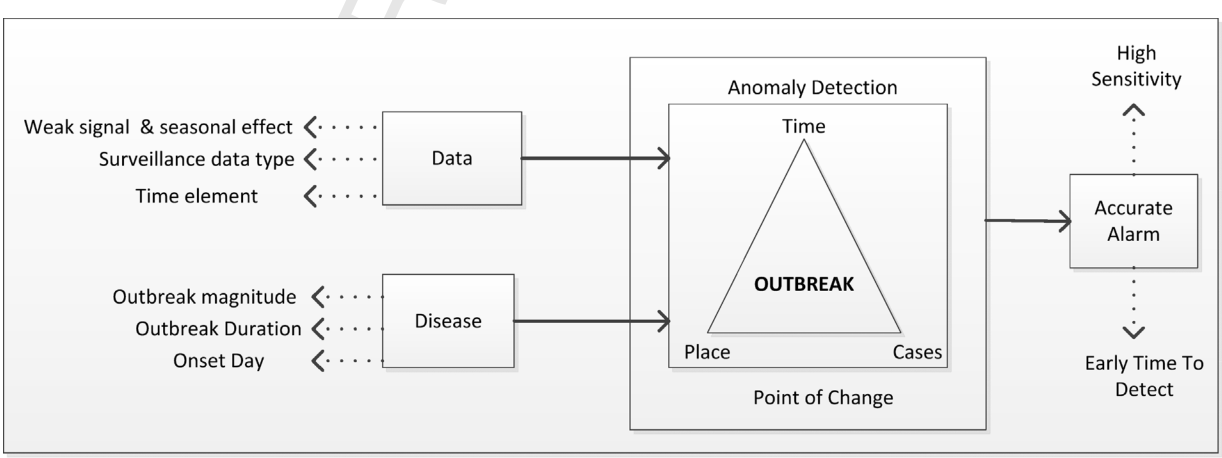

Fig. 1. General concept of outbreak detection.