Chapter 105
Pattern Mining for Outbreak Discovery Preparedness

Zalizah Awang Long
Malaysia Institute Information Technology, Universiti Kuala Lumpur, Malaysia

Abdul Razak Hamdan
Universiti Kebengsaan Malaysia, Malaysia

Azuraliza Abu Bakar
Universiti Kebengsaan Malaysia, Malaysia

Mazrura Sahani
Universiti Kebengsaan Malaysia, Malaysia

ABSTRACT

Today, the objective of public health surveillance system is to reduce the impact of outbreaks by enabling appropriate intervention. Commonly used techniques are based on the changes or aberration in health events when compared with normal history to detect an outbreak. The main problem encountered in outbreaks is high rates of false alarm. High false alarm rates can lead to unnecessary interventions, and falsely detected outbreaks will lead to costly investigation. In this chapter, the authors review data mining techniques focusing on frequent and outlier mining to develop generic outbreak detection process model, named as "Frequent-outlier" model. The process model was tested against the real dengue dataset obtained from FSK, UKM, and also tested on the synthetic respiratory dataset obtained from AUTON LAB. The ROC was run to analyze the overall performance of "frequent-outlier" with CUSUM and Moving Average (MA). The results were promising and were evaluated using detection rate, false positive rate, and overall performance. An important outcome of this study is the knowledge rules derived from the notification of the outbreak cases to be used in counter measure assessment for outbreak preparedness.

DOI: 10.4018/978-1-4666-2455-9.ch105

INTRODUCTION

The public health surveillance system caught the researcher's attention into developing the detection algorithm after 21 September for terrorist detection and also for H1N1 pandemic. WHO reported that earlier case reported for H1N1 was first noticed in North America in April 2009 and then the spread to the other part of the world. In June 2009 there were 74 countries affected by H1N1. Various data sources had been considered to detect the outbreak including data from emergency department, over the counter sales, medical image and also test-order for early detection of an outbreak case.

In detecting outbreak, the detection system identified the anomaly pattern when compared to previous pattern to detect an outbreak. In statistical analysis, it is known as aberration (Wong, 2004). The main problem indicated in most of the studies shows that high in false alarms will lead to unnecessary intervention resulting high in operation cost.

To solve the indicated problem there are few techniques that have been identified such as statistical-based, Bayesian-based, and also knowledge-based techniques. Those discussed techniques showed the improvement in terms of detection rate and also false-positive-rate. The techniques proposed are more toward the complement observation and not to replace any available techniques (G. F. Cooper, Dowling, Levander, & Sutovsky, 2007; Shen & Cooper, 2007). Statistical-based methods face problems in requiring long training time (Guthrie, Stacey, & Calvert, 2005) and validation for the error rate is not significant since the value is too small while Bayesian and knowledge-based suffered to provide more accurate results due to limited numbers of actual cases of true outbreaks to be tested on the approaches.

To reduce the false positive rate and high in detection rate are the objectives of most outbreak detection techniques. Our approach is to provide a new process model in detecting outbreak and provide rules for the preparedness awareness. These will lead towards prevention of the recurrent outbreaks.

BACKGROUND

There is no specific definition in defining outbreak. According to Lai & Kwong, (2010), from the epidemiology view, an outbreak occurs if individuals develop similar symptoms one after another and the disease incidence is higher than usual. The general definition from Center Disease Control (CDC) defined an outbreak as the occurrence of more cases of disease than what is expected in a given area over a particular period of time. Different diseases possess their own outbreak definition. Dengue outbreak, for example, is defined as increase of cases per week persisting for at least 3 successive weeks to a level at least three times above the mean of previous 3 weeks (Runge-Ranzinger, Horstick, Marx, & Kroeger, 2008). Another definition obtained for dengue outbreaks is an increase in 2 SD above the mean ((Carme, et al., 2003; Oum, Chandramohan, & Cairncross, 2005; Rigau-Pérez, et al., 1998; Talarmin, et al., 2000)). In this study, the focus concentrates on the Malaysia dengue environment. Based on the study conducted by Seng, Chong, & Moore, (2005) the dengue outbreak was defined according to Johor Health State Department, which is an occurrence of more than one case in the same locality, where the date of onset between the cases are greater than 14 days. The outbreak is clear when there is no new case reported within 14 days.

There are numbers of outbreak detection algorithm based on statistical analysis such as CuSUM focusing on detecting the outbreak with the same motivation to reduce numbers of false detection and at the same time to increase the detection rate. Outbreak detection required combination of clinical and non-clinical data source for better detecting outbreaks and also depending on the automatic analysis on the combination of data sources(Blind & Das, 2007; Brown & Gray, 2005; Buehler, et al., 2004b; D. L. Cooper, et al., 2005; German, Armstrong, Birkhead, Horan, & Herrera, 2001; Stoto, Schonlau, & Mariano, 2004). Surveillance system required definition

Related Content