Chapter 28
Disease Surveillance System for Big Climate Data Processing and Dengue Transmission

Gunasekaran Manogaran
VIT University, India

Daphne Lopez
VIT University, India

ABSTRACT

Ambient intelligence is an emerging platform that provides advances in sensors and sensor networks, pervasive computing, and artificial intelligence to capture the real time climate data. This result continuously generates several exabytes of unstructured sensor data and so it is often called big climate data. Nowadays, researchers are trying to use big climate data to monitor and predict the climate change and possible diseases. Traditional data processing techniques and tools are not capable of handling such huge amount of climate data. Hence, there is a need to develop advanced big data architecture for processing the real time climate data. The purpose of this paper is to propose a big data based surveillance system that analyzes spatial climate big data and performs continuous monitoring of correlation between climate change and Dengue. Proposed disease surveillance system has been implemented with the help of Apache Hadoop MapReduce and its supporting tools.

1. INTRODUCTION

Big data describes any voluminous amount of structured, semi-structured and unstructured data that has the potential to be mined for useful information. The basic characteristics of big data can be defined by 3Vs: the large volume of data, the wide variety of types of data and the velocity the data in and out. Much of the data acquired today is not in the structured format; hence heterogeneity, complexity, scale, timeliness and privacy with big data pose challenges in every stage of processing (Thilagavathi et al., 2014; Victor et al., 2016). It is not possible to process big data by traditional data processing frameworks; it requires new techniques and architectures to extract useful information from large amount of data. For

DOI: 10.4018/978-1-5225-7501-6.ch028

example, Azar et al (2015) have used neural-fuzzy classifier to reduce the dimensionality of medical big data. Kamal et al (2016) have used MapReduce framework to process huge DNA dataset. Traditionally, data mining and soft computing techniques are most often used in disease diagnosis (Dey et al., 2014; Dey et al., 2016; Kausar et al., 2016; Zemmal et al., 2016). Nowadays, big data plays a vital role in many applications such as healthcare, social networking, education and training, security platforms, industries, sensor networks, climate simulation and others (Hassanien et al., 2015).

Especially, climate simulation plays a vital role in day-to-day human health. Climatologists use the term Climate "normals" to compare current climatological trends. In order to calculate the normal climate, climatologists use World Weather Records (WWR) that was established in 1923, generates a huge volume of daily precipitation, temperature, rainfall, wind speed, and pressure data from thousands of weather stations around the world. In addition, World Metrological Organization (WMO) is maintaining climate data by digital publication collected from thousands of weather stations around the worldwide. The goal of this big climate database is to store and validate extreme climate changes, such as the highest/lowest temperature and strongest wind speed, and greatest precipitation on Earth.

Many researchers suggest that global climate change has become one of our era's best challenges in 21st century. For example, Lee & Kang (2015) and Nativi et al (2015) have identified the climate big data challenges and opportunities in global earth observation system. The climate change researchers now deal with big data for identifying possible future climate change scenarios. For example, EnviroAtlas is a project that provides impact of climate change on healthcare, society and ecosystems.

The major public health organizations of the world have said that climate change is an important public health issue. Nowadays, number of early warning systems is available to monitor the climate change. Traditional early warning systems detect the outbreak mainly from virology and clinical data collected through health departments, randomized telephone, and public survey and so on. These systems suffer with the limitation of slow reporting time, missing out real time climate data and rapidly emerging diseases. The immediate need is to develop intelligent models for efficient monitoring of real time climate change.

Ambient intelligence is an emerging platform that provides advances in sensors and sensor networks to capture the real-time climate data (Dingli et al., 2012; Ribeiro et al., 2013). As a result of advancement in climate processing framework, the size of climate data has increased to several exabytes and so it is often called big climate data. Hence, there is a need to develop advance big data architecture for processing the real-time climate data. Recently, Faghmous & Kumar (2014) have identified challenges and opportunities to mine large climate datasets using big data technologies. Hay et al. (2013) have identified big data opportunities for global infectious disease surveillance. The intention of this paper is to propose a big data based surveillance system that analyzes spatial climate big data, and performs continuous monitoring of correlation between climate change and Dengue. The proposed architecture can continuously monitor the climatic change and public health as they unfold. Effective surveillance systems are needed for developing countries like India for taking corrective actions to improve health conditions during climate change scenarios.

The remainder of this article is structured as follows: Section 2 surveys related work. Section 3 presents an overview and implementation of proposed disease surveillance system for climate change and dengue. Performance evaluation, measurement and optimization technique of the proposed system is explained in Section 4 and 5 respectively. Result and discussion are described in Section 6. Finally, section 7 concludes this article by summarizing our work.