Australian Journal of Basic and Applied Sciences, 9(11) May 2015, Pages: 611-616

ISSN:1991-8178

Australian Journal of Basic and Applied Sciences

Journal home page: www.ajbasweb.com

Classification and Prediction of Earthquake and Tsunami using Big Data Analytics

Dhinakaran K, Silviya Nancy J, Saranya R

Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha University, Chennai, India

ARTICLE INFO

Article history:
Received 12 March 2015
Accepted 28 April 2015
Available online 1 June 2015

Keywords:
Geographical Information Systems (GIS), Satellite communication, Big data analytics, Apache Hadoop, Tsunami, Earthquake.

ABSTRACT

The proposal of this paper work is to analyze the geo-spatial data whose production exceeds exa-bytes of data for every second, which is due to the advent of sensor and satellite communication. This is massive growth requires a special modeling that supports the immediate expectation of the clients and the society. To mark a revolution in Geographical Information System (GIS), cloud computing has extended its hand to warehouse the vast amount of data. So, the chief aim of the work is to predict the occurrences of Tsunami and earthquake to avoid natural and public disaster. The analysis is done with classification and prediction concepts. Obviously, the geo-spatial data is claimed as bigger data and requires analytics to sort out the predicted results. In this view, Apache Hadoop offers the platform to process the data and satellite images faster providing a better solution.

© 2015 AENSI Publisher All rights reserved.

To Cite This Article: Dhinakaran K, Silviya Nancy J, Saranya R , Classification and Prediction of Earthquake and Tsunami using Big Data Analytics. ***Aust. J. Basic & Appl. Sci.,*** *9(11): 611-616, 2015*

INTRODUCTION

Today's world is stacked with huge creation of data that is producing critical hassle in traditional database infrastructures. There are collected from various sources obviously including IT organizations, research and academics, but not excluding new data sectors like sensors, geo-spatial data, to the mobile devices that are used generates massive amount of data for every click. These are referred as "Big Data", which is crawling around each and every industry. Big data has become the most popular technology that washes away all the industries and organizations. In today's world, increase in volume, velocity and different variety of data from numerous springs have reached the edge with huge collection of data which is not compatible to handle by the traditional databases. A survey states that, the business markets have raised its performance up to 30 percent, with increasing its productivity in the backing of big data analytics. It does not stop here and every year it increases by 20 percent.

Big data incorporates everything starting from digital data to health data, meteorological and geological data. Big data houses datum from various back theories in structured like databases, sensor data, location-aware data and unstructured formats videos, pictures, and explicitly the social and media data. The big data analytics provides an efficient analytic platform for handing the big data with several tools like Hadoop, Hive, mango DB, NoSQL, etc.

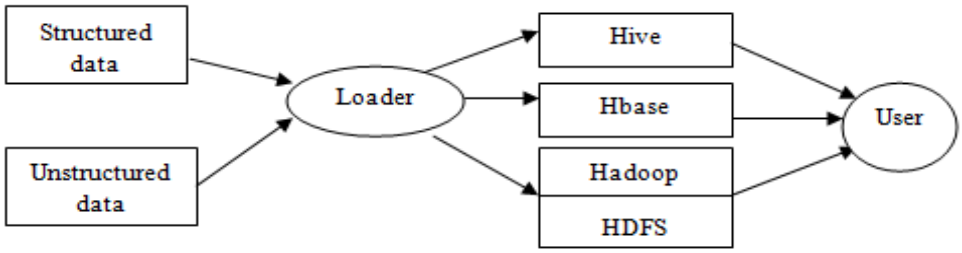

Fig. 1: Processing of Big Data.

The flooding of data is not only in the field of IT, academia and health care; it also combines the geospatial data. With the advancement in satellite and communication technology, the geospatial data were collected, are being fetched now and gathered even today with variety of techniques like satellite

Corresponding Author: Silviya Nancy, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha University, Chennai, India.
Tel: +91 9487752343; E-mail: jsilviyanancy@gmail.com

remote sensing, global positioning system (GPS), global navigation satellite system (GNSS), photographic surveys, digital cameras, radar etc. The sheer size of this geological data have exceeded exponentially which was not able accommodate by the traditional databases in 2007 itself (AgusSuharyanto, 2012). The UN-GGIM stated that, 2.5 quintillion geo-spatial data is been produced everyday with the location-aware details specifically. This current trend is the interconnection of stipulated technology, which eventual give rise to Internet of Things (IOT). These resources will exceed more than 50 billion by 2020.

The Big data management also has become a crucial task, because data can be duplicated. There is also a need to generate and reserve the metadata has turned out to be a biggest mission. Storage cost is cheaper but the storage space is not able to manage with the growing size of data. The solution to the adhering problem is to identify the conceptual data sets, compress it without data loss with profound big data tools and techniques mentioned above. This is solely applicable for geo-spatial data also. The geo-spatial big data analytics is serving around for many years, but it has not been identified popularly due to its much storage occupying problem. The big data analytics is recognized as an efficient methodological concept for handling any type of data.

One of the well-known big data analytic tools is Hadoop, which is the Apache's open source framework. With the help of Hadoop Distributed File System (HDFS), the geo-spatial master data set is processed to analyze the outcome. The mainstay of this proposal work is to analyze the big data sets of tsunami and earthquake for the future prediction using machine learning techniques and big data analytics. The work is to compress the data sets of Tsunami and earthquake and to predict the future occurrences to avoid disasters.

Related survey:

This chapter concentrates on the literature review of the Tsunami and earthquake prediction techniques. And the way in which these huge sets of data are processed efficiently in previously the existing systems and the proposed view. The survey on recent geo-spatial systems reveals that 86 percent of Tsunami is occurred due to the earthquake. Since the earthquake prompted Tsunami is becoming more frequent, forecasting of those occurrences must be done preservative. It cannot be identified with the raw data, but with the aid of machine learning tool called mahout techniques like clustering and classification can easily be done to identify the future statistics.

Earthquake:

The earth is subjected to various natural disasters like flood, earthquake, Tsunami, volcanic eruptions, hurricanes, etc. The furious among these is earthquake, which stimulates severe tectonic disruptions and shakes the earth surface causing environmental disaster and human life loss. The earthquakes are caused by the discharge of seismic waves that causes vibration in the earth crust causing severe damage AgusSuharyanto *et al*. (2012). The different types of earthquakes are defined as follows; tectonic earthquake is one that occurs usually due to the changes in the tectonic plates. The volcanic earthquake erupts with the onset of volcanic eruption, collapse earthquake targets the different mines, and the last is the explosion earthquake which emulates due to the chemical explosives. The form and movement of waves also varies based on the intensity of the attack.

Tsunami:

Tsunami is a very strong wavelength of waves which flows about hundred kilometers from the ocean. This can be erected by dangerous earthquake and volcanic eruption in the ocean usually above the magnitude level of 6.5 and above. The origin of the word Tsunami was given by the Japanese and here we call it as "harbor wave". The manifestation degree of Tsunami is directly proportional to the vertical displacement of the slip faults. It can also be stimulated by the landslides in the sea bed, volcanic and meteorite eruptions. On an average, two times the Tsunami erupts in a year somewhere in the world. A review contemplates that around every 15 years, a destructive Tsunami takes place. Specifically there are two types of Tsunami, namely local and distance. The former one is less intensive which causes disaster with 100kilometers. The one which is denoted later is the very damaging one causing disruptions up to 1000kilometers. There are quite a few present technologies are there to detect it like seismographs. Ordering to that a new technique is been discussed in this paper with Hadoop by scrutinizing the previous occurrence data to predict the future A. Hoechner *et al*. (2013).

Agus Suharyanto *et al* (2012) in this article speaks about the Tsunami disaster took place in Dec 2004. They have also explained about the different disastrous occurrences of Tsunami in various places of the world. The chief aim of the article exploits about the case history of all the Tsunami records and with the help of GIS, they proposed to predict the earthquake based on the area.

G.V. Otari and R. V. Kulkarni (2012) enlightens about the various data mining models that are used to predict the earthquake. They have included logistic models, neural networks, and Bayesian belief network for predicting the earthquake earlier. This article provides social awarenesswith ample amount of knowledge on techniques of prediction and classification for earthquake prediction.

The authors Fumihiko IMAMURAand Suppasr ANAWAT (2012) presented a very powerful work on earthquake and Tsunami took place in Japan in

March 2011 at the pacific coast. The paper elucidates the entire history of the disaster in a very proficient way. They have spoken about the early prediction of natural calamities in the evacuated areas with the available previous history.

The issue from the special Earth Science authored by Kusala Rajendran (2013) states and proceeds with the Sumatra – Andaman and Japan devastation took place in 2004 and 2011. With the collection of history and observations, the presenter tries to identify the future falls of subduction zones. This article also enlightens about the advances in space technology to predict the occurrences with the seismic graphs.

The volume in Natural Disaster written by Boris W.Levin describes and mitigates about the mostly occurred natural disaster Tsunami with its causes and prediction methodology.

The research work compares the amplitudes of far-field Tsunami and local Tsunami with the previously available information. Parameters from different sources include local and far-field waves whose adversity is based on the height and magnitude for identifying it at earlier range were discussed by Eric L. Geist (2013).

Roopa.V (2014) engages the importance of satellite remote sensing techniques in finding the disaster. This research work inculcates Geographic Information System (GIS) and Web Technology to predict the earthquake and Tsunamis with decision making systems.

A. Hoechner *et al*. (2013) presents his work with GPS real-time processing for issuing warning on on-set of Tsunami. This framework insists on setting GPS system on coastal areas for monitoring to intimate Tsunami quick cautioning within minutes.

The author Shongkour Roy (2014) in his paper discussed about the probability and statistical method for predicting the earthquakes using weibull distribution. In the data collection phase, the researcher has collected the data from previous history for normalization.

Proposed work:

With the related survey on earthquake and Tsunami, it is understood that they are unimaginable natural disasters. There are many evidences provided in this study explaining about the environmental and physical risks and challenges.

The Tsunami triggered by the underwater and ground earthquake is considered to be the most dangerous of all. So, the mainstay of the proposed research work is to build a predictive analysis system for earlier identification of the disaster which would save .millions of lives and property.

The figure 2 describes the overview of the predictive system using Hadoop, which involves the collection of in sequence from previous history and research. Next is processing the huge data set in the Hadoop that is taken care of Hadoop Distributed File System, where the resultant is classification of the data set with MapReduce framework. As a result of these proceedings, predicted results of next occurrences' can be obtained.

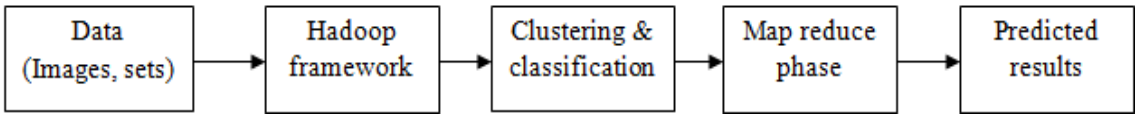

Fig. 2: Overview of predictive analysis system using Hadoop MapReduce.

Design of system for predictive data analysis for Tsunami:

The analysis and steps are depicted as follows;

Data Collection:

To build a predictive system, data collection is the most important criterion for the analysis. This involves the collection of information from various sources and institutes like National Tsunami warning centers in the form of data sets and images from the satellites.

Big Data analytics:

With the collected data from various sources, the resource manager allocates the data set to the available node manager with the resources. The map reduce framework analyzes and maps the data with the similar set and produces the reduced and predicted results.

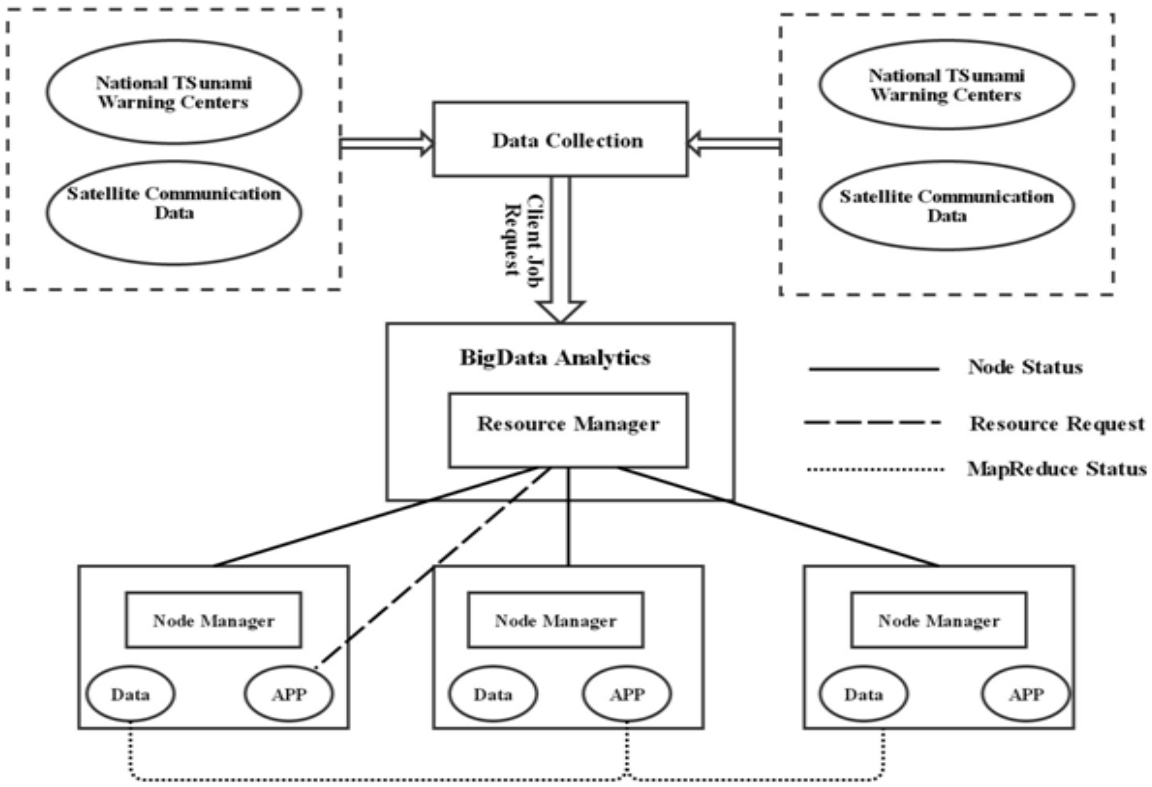

Fig. 3: Proposed system model for predictive data analysis for Tsunami.

Methods and Discussions:

This chapter explains about the classification method used for identifying the near occurrence.

Prediction rule for earthquake:

Table 1: Sample collection of magnitude and causes.

Magnitude	Causes
2.5	Not felt can be recorded by seismograph
3 – 3.9	Minor
4 – 4.9	Light
5 – 5.9	Moderate
6 – 6.9	Strong
7 – 7.9	Major
8 – more	Disaster

Table 2: Sample survey rate on recent earthquake in year 2015

S. No.	Magnitude	Region
1	2.3	Hawaii
2	2.2	Western Turkey
3	2.0	Turkey-Syria Border
4	2.4	Island of Hawaii, Hawaii
5	2.4	Sicily, Italy
6	2.4	Island of Hawaii, Hawaii
7	2.5	Western Turkey
8	2.7	Near the coast of Western Turkey
9	5.3	South Africa
10	1.5	Germany

Courtesy – EMSC (Dated - 15-05-2015 – 2 hours data from 12.30PM to 2.00PM)

The above Table 2 shows the frequency of the earthquake which has to be taken care of for preventing the Tsunami.

Classification Algorithm:

```
Begin
If ((magnitude < 8) && (sharp waves in the sea))
then
Tsunami occurrence ^ earthquake  = true
Else if
Server earthquake = true
End
```

Fig. 4: Classification rule.

Alternate decision tree:

ADT automatically generates the lower and higher proximity of earthquake dataset. It inculcates both the positive and negative aspects of the data set clearly. All the given attributes are classified and interlinked to give expected outcome.

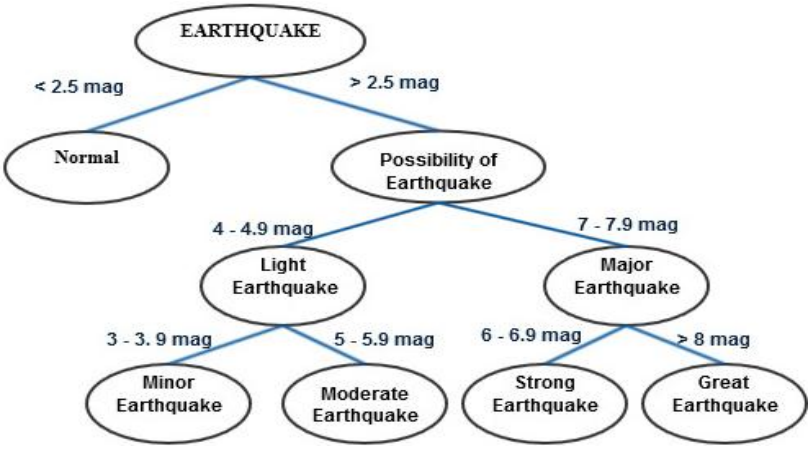

Fig. 5: Alternate decision tree.

Predictive analysis:

The graph below shows the scale of disaster at various ranges of magnitude.

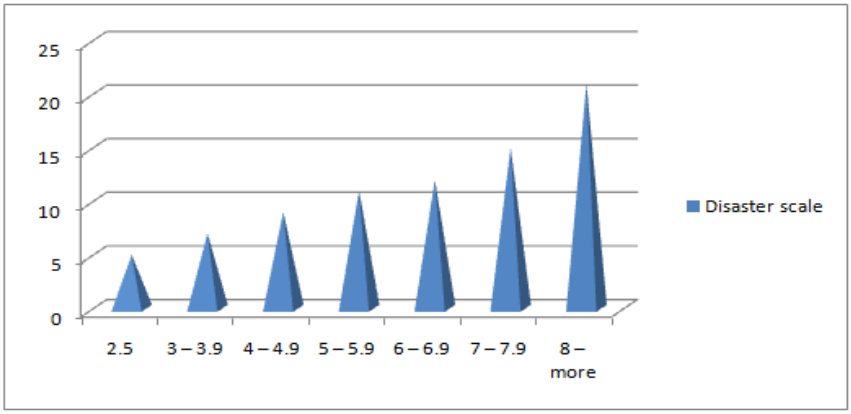

Fig. 6: Simulation of predictive results for Tsunami with earthquake data set.

Conclusion:

The research work focused mainly related survey of Tsunami and earthquake. The different strategies on analyzing the geo-spatial data were discussed. A clear explanation of big data and geo-data incorporation was efficiently displayed. With this analysis, a classification rule is defined for forecasting the Tsunami which is triggered by the severe earthquakes with alternate decision tree which takes the chance to explicitly display the probabilities of occurrences'.

REFERENCES

Agus Suharyanto, Alwafi Pujiraharjo, Fadly Usman, Keisuke Murakam,Chikashi Deguchi, 2012. Predicting Tsunami Inundated Area and Evacuation Road Basedon Local Condition Using GIS, IOSR Journal of Environmental Science, Toxicology and Food Technology (IOSR-JESTFT)ISSN: 2319-2402, ISBN: 2319-2399, 1(4): 05-11.

Boris Levin, W , "Tsunamis: causes, consequences, predictions and response", NATURAL DISASTER – Vol I.

Eric Geist, L., 2005. Rapid Tsunami Models and Earthquake Source Parameters: Far-Field and Local Applications, ISET Journal of Earthquake Technology, Paper No. 460, 42(4): 127-136.

Fumihiko IMAMURA and Suppasr ANAWAT, 2012. Damage Due to the 2011 Tohoku Earthquake Tsunami and its Lessonsfor Future Mitigation, Proceedings of the International Symposium on Engineering Lessons Learned from the 2011 Great East Japan Earthquake, March 1-4, 2012, Tokyo, Japan.

Hoechner, A., M. Ge, A.Y. Babeyko and S.V. Sobolev, 2013. Instant tsunami early warning based on real-time GPS – Tohoku 2011 case study, Published by Copernicus Publications on behalf of the European Geosciences Union, Nat. Hazards Earth Syst. Sci., 13: 1285–1292.

http://study.com/academy/lesson/how-plate-movement-affects-earthquakes-tsunamis-volcanic-eruptions.html

http://www.livescience.com/21486-earthquakes-causes.html

Kusala Rajendran, 2013. On the recurrence of great subduction zone earthquakes, CURRENT SCIENCE, VOL. 104, NO. 7.

Otari, G.V. and Dr. R.V. Kulkarni, 2012. A Review of Application of Data Mining in Earthquake Prediction, (IJCSIT) International Journal of Computer Science and Information Technologies, 3(2):3570-3574.

Roopa, V., 2014. Remote Sensing& its applications in Disaster Management like Earthquake and Tsunamis, International Journal of Science and Research (IJSR) ISSN (Online), pp: 2319-7064.

Shongkour Roy, 2014. Probabilistic Prediction for Earthquake in Bangladesh: Just How Big Does the Earthquake Have to Be Next Years? Open Journal of Earthquake Research, 3: 108-114.

Stefano Tinti, Alberto Armigliato, Roberto Tonini, Alessandra Maramai and Laura Graziani, 2004. Assessing the Hazard Related to Tsunamis of Tectonic Origin: A Hybrid Statistical-Deterministic Method Applied to Southern Italy Coasts, ISET Journal of Earthquake Technology, Paper No. 465, 42(4): 189-201.