IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 3, March 2016

Outliers Analysis of Real-time Metereological data using Elastic Search repositories for a proposed Natural Disaster Prediction System (NDPS)

Manoj Sakhala[1], Yogesh Dasgaonkar[2], Ajay Kumar[3], Jignasa Patil [4]

Department of Computer Engineering, Veermata Jijabai Technological Institute VJTI, Mumbai, Maharashtra, India[1,2,3,4]

Abstract: Weather forecasts are made by collecting quantitative data about the current state of the atmosphere at a given place and using scientific understanding of atmospheric processes to project how the atmosphere will change. But in order to make quick and exact forecasting decisions we will need to have a critical observation points to give alerts about the fluctuations in the weather conditions. Quick prediction of natural calamities or other environment fluctuations on a real time data can be implemented with the help of using and analyzing Outliers within the data inside elastic search repositories. Our proposed system will consist of five major components:

1) Weather data collecting agent: Weather live feed/Historic data.
2) Processing agent: Coverts the data in JSON format for storing in elastic search repository.
3) Repository: Using Elastic Search technology.
4) Outlier Analyzer: Contains the logic to perform analysis of the disaster prediction. To improve the accuracy of prediction the outlier rules have been mapped to a Bayesian network model.
5) Display agent: Provides a display in the form of alerts, alarms or graphical for warnings based on what outlier was activated.

Keywords: Weather Forcasting,Outlier,ElasticSearch,JSON,Bayesian Network,Natural Calamity.

I. INTRODUCTION

Every day we have natural disasters occurring in various locations of the world. This causes damages of life , land, crops and structures. Even though it is not possible to avert the losses occurring due to land, crops and structure damages, we can certainly prevent the loss of life or at least minimize this loss. This can be done if we are able to predict the calamity well in advance and alert the government departments and surrounding people so that they can move to another safe location.

This could also allow the business organization to shift all their processing to the Disaster recovery site as quickly as possible.

II. NATURAL DISASTER PREDICTION CHALLENGES

1.Earthquakes

No system has been able to predict earthquakes reliably enough and quickly to allow the evacuation of threatened cities. Scientists and research studies that so many factors decide whether a fault will rupture that earthquakes could be unpredictable[6].

One basic idea behind earthquake prediction is that faults send out subtle but detectable warnings before they slip. Scientists and systems have looked at a host of potential warning signals. In the next section we have accumulated such precursors that would help us predict this calamity with some reliability.

2.Droughts

Drought prediction depends on the ability to forecast precipitation and temperature. Scientists aren't able to predict drought a month or more in advance for most parts of the world.

NASA has launched a satellite SMAP(Soil Moisture Active Passive) on Dec 31 2015, that could be used for Monitoring droughts and predicting floods.[10]

3.Floods

Several types of data can be collected to assist hydrologists predict when and where floods might occur. The first and most important is monitoring rainfall received in a region on a real-time (actual) basis. Second, the rate of change in river stage is monitored on a real-time basis can help indicate the severity and immediacy of the threat. Thirdly, having knowledge about the river's drainage system, such as soil-moisture conditions, ground temperature, no of open areas and their coverage, can help to predict how extensive and damaging an impending flood might be and cause a lot of damage.[6]

In December , a flood hit Chennai due to a heavy rain and subsequent overflow of the Chembarambakkam dam cause loss of life and property.(ref news: indianexpress.com/ dated December 14,2015)

Yet the local government failed to alert the public, police or power utility. If we had a system like the proposed

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 3, March 2016

NDPS this would lead to proactive alerting system for becoming better prepared for such calamities rather than depend on human intervention.

Even though we have so much data (including Big data) and so much money put in to the weather forecasting systems by way of satellites , we are still not able to predict or at least be near predictions for the onset of a calamity .Many fields of science, geography and space are capable to collecting data , but to infer proper predictions about the occurrence of natural calamity remains a challenge even for the best scientists in the world.

III. ARCHITECTURE OF THE PROPOSED SYSTEM

We have proposed a system that would be able to predict the occurrence of the natural calamity by using the five major components as shown in the diagrammatic representation:

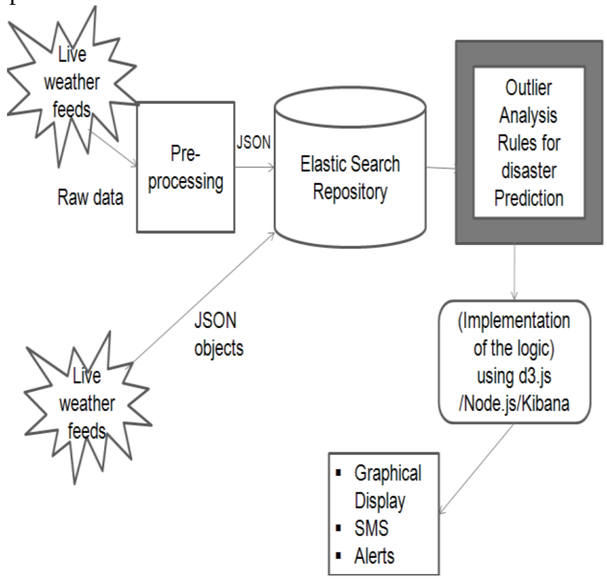

Fig 1: Working of the proposed National Disaster Prediction system(NDPS)

We can describe the five components of the described system as follows:

1) Weather data collecting agent:
Weather live feed/Historic data will be collected from the internet via API calls. This data can be made available faster and easily if we get recommendation from the meteorological and climate conditions measuring departments

2) Processing agent:
Not all data at our disposal may be in the required format for our elastic search repository . Hence we need to parse the data in whatever format we receive then make alteration into it so that data is transformed as JSON elements for inputs into our system,

3) Repository:
Using Elastic search repository over traditional database offers advantages like faster searches as it is required to maintain only one form of record structure and we can easily parse whatever particular parameters we want for our outlier analysis.

4) Outlier Analyzer:
This is the main area of our research as we have to use the various concepts of probability theory and artificial intelligence. We can see from Fig 2. on how to use the Bayesian network model to provide a good estimate of an natural calamity to occur.

We make use of a Bayes Theorem to calculate the probabilities of an natural calamity to be predicted correctly

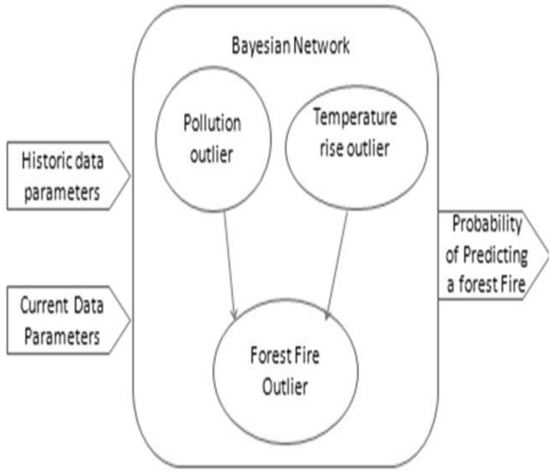

Fig 2: A sample Bayesian network for predicting forest fire

Historical data: It is the meteorological data of the past events . It helps this system to seed the probabilities tables for the events in the network based on previously occurred events. Using this data we build the initial conditional pprobability table for the individual events .Based on the current data received the CPTs can be updated to reflect more accurate marginal probabilities of each event w.r.t other events

Current Data: Is in the form of live feeds helps the outlier to detect if the events have occurred based on which probabilities can be calculated using conditional probabilities from Bayes Theorem. In the implementation of this system this data needs to be pre-processed and converted into JSON format as input to our elastic search repository.

5) Display agent:
Provides a display to read alerts . Kibana ,Node JS or d3.js could be used for this purpose . A feature to send SMS alerts will be incorporated in the system.

IV.OUTLIER RULES

Outlier Points for Earthquake prediction:

1. Temperature Change:
Outlier point : A 10 to 15 degree increase in temperature

2. Water Level:
3 to 10 days before an earthquake, the water level begins to fall by 3-15 cms. After a short period, it starts rising when the earthquake strikes

3. Oil Wells:
Large fluctuation in oil Outflow (case study Gulf of Suez and Eilat oil wells)[6]

 DOI 10.17148/IJARCCE.2016.53224

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 3, March 2016

4. Chemical composition of underground water :

(i) Concentration levels of dissolved minerals and gaseous components remained almost constant during seismically inactive period.

(ii) An appreciable increase in concentration of dissolved minerals was noticed 2 to 8 days before an earthquake.

(iii) After the earthquake, anomalies in concentrations of the gaseous and mineral components disappear.

(iv) We assume that a rise in sea level has a direct influence on the change in chemical composition of underground water.

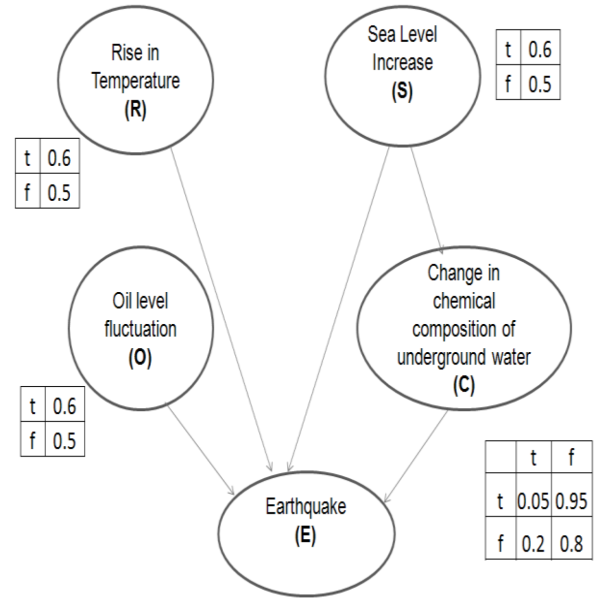

Fig 3: A sample Bayesian network for Earthquake

Rise in Temp	Sea level rise	Oil lvl fluc.	Change chemical comp.	Earthquake t	t
t	t	t	t	0.8	0.2
t	t	t	f	0.8	0.2
t	t	f	t	0.8	0.2
t	T	f	f	0.7	0.3
t	f	t	t	0.8	0.2
t	f	t	f	0.7	0.3
t	f	f	t	0.7	0.3
t	f	f	f	0.5	0.5
f	t	t	t	0.6	0.4
f	t	t	f	0.6	0.4
f	t	f	t	0.3	0.7
f	t	f	f	0.3	0.7
f	f	t	t	0.5	0.5
f	f	t	f	0.4	0.6
f	f	f	t	0.1	0.9
F	f	f	f	0.01	0.99

Table1: Conditional Probability Table (CPT) with sample initial seeds for marginal probabilities of individual events w.r.t earthquake event

Based on historic data we have to seed the initial value of Probability of Earthquake → P (E=t) =0.01 and P (E=f) =1- P (F=t) =0.99 and t, f indicates values true and false respectively.

"What is the probability of predicting a Earthquake correctly given that outlier detected events Temperature Rise Sea Level Rise, Oil Level Fluctuation and Change in chemical composition of underground water?"

Using conditional probability formula from notations use in Fig 3, we get:

$$P\{\frac{\mathbf{E=t}}{\mathbf{R=t,S=t,O=t,C=t}}\}$$

$$= \frac{\mathbf{P(R=t,S=t,O=t,C=t\,,E=t)}}{\mathbf{\Sigma\ (E\ \varepsilon\ (t,f))\ \ P(R=t,S=t,O=t,C=t\,,E)}}$$

$$= \frac{\mathbf{P}\left\{\frac{\mathbf{R=t,S=t,O=t,C=t}}{\mathbf{E=t}}\right\} * \mathbf{P(E=t)}}{\mathbf{P}\left\{\frac{\mathbf{R=t,S=t,O=t,C=t}}{\mathbf{E=t}}\right\} * \mathbf{P(E=t)} + \mathbf{P}\left\{\frac{\mathbf{R=t,S=t,O=t,C=t}}{\mathbf{E=t}}\right\} * \mathbf{P(E=f)}}$$

$$= \frac{0.9 * 0.01}{(0.9 * 0.01\ +\ 0.1\ * 0.99)} = \mathbf{8.33\%}$$

Outlier Points for forest fire prediction:

1.Forest data :
Can be collected from the forest department in advance

2. Temperature rise:
The average temperature vs. rising temperature

3.Pollution in air (smoke detection:
Pollution sensing devices installed in the forest vicinity

We construct a Bayesian net for the outliers as shown below:

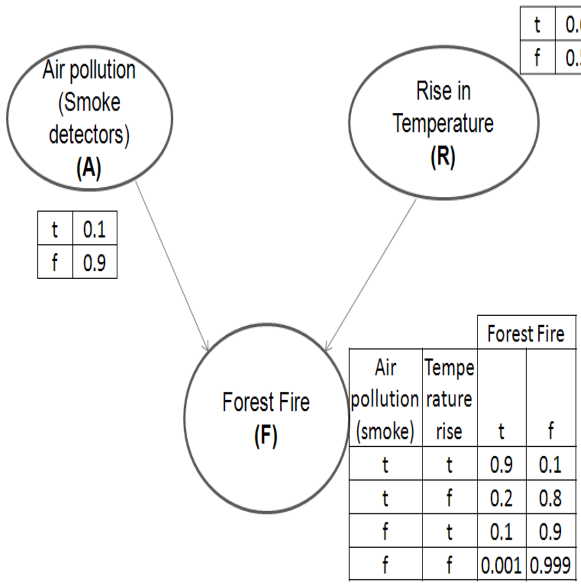

Fig 4. A sample Bayesian network to predict forest Fire

Calculating the probability for Forest Fire.

"What is the probability of predicting a forest fire correctly given that outlier detected events Air Quality and Temperature rise?"

Based on historic data we have to seed the initial value of Probability of Forest Fire → P (F=t) =0.05

and P(F-f)=1- P(F=t)=0.95 and t ,f indicates values true and false respectively.

Using conditional probability formula from notations use in Fig 4, we get:

$$P\{\frac{\mathbf{F=t}}{\mathbf{A=t\,,R=t}}\}$$

$$= P\{\frac{\mathbf{A=t,R=t,F=t}}{\mathbf{\Sigma\,(F\,\varepsilon\,(t,f))\;P\,(A=t,R=\;t,F)}}\}$$

$$= \frac{\mathbf{P}\left\{\frac{\mathbf{A=t,R=t}}{\mathbf{F=t}}\right\} * \mathbf{P(F=t)}}{\mathbf{P}\left\{\frac{\mathbf{A=t,R=t}}{\mathbf{F=t}}\right\} * \mathbf{P(F=t)} + \mathbf{P}\left\{\frac{\mathbf{A=t,R=t}}{\mathbf{F=t}}\right\} * \mathbf{P(F=f)}}\}$$

$$= \frac{0.9 * 0.05}{(0.9 * 0.05\; + \;0.1\; * 0.95)} = \mathbf{32.14}\%$$

Bayes network can be modeled for the following :-

Outlier Points for flood prediction:

1.Rainfall in the region:
Collected from meteorological department on live feed

2.Dams water level data :
Can be collected from the water department in advance
Rise in dam level or increased Dam's outflow.

3. Sea level data:
Rise in the sea level
Other factors that contribute to sea level : Storm Surge + Tide + Waves + Rivers + ground temperature + soil moisture conditions

4.City drainage cleaning data :
City drains not cleaned could be a outlier point

V.CONCLUSION

Using this proposed system we can provide an estimation to the occurrence of a natural calamity. This system provides the ability to detect natural calamities like earthquake, forest fires, droughts etc with a some accuracy. As the system improves its probability values by comparing with continuous inputs of historic and current data and the probability with which the calamity was detected correctly to the incorrect predictions, the system will refined even further. This refinement would help the system infer dependencies between the various meteorological events and make it robust. In future, this prototype could help researchers to make use of their own prediction models and data mining method and help us predict the calamity efficiently, accurately and most importantly quickly.

REFERENCES

1. Mandal, S. ; Sch. of Manage., XLRI Jamshedpur, India, India ; Saha, D. ; Banerjee, T.A neural network based prediction model for flood in a disaster management system with sensor networks, Jan. 2005, IEEE
2. Lianying Sun ; Coll. of Inf., Beijing Union Univ., Beijing, China ; Tao Peng ; Bao Hong Notice of Retraction Key technology of disaster prediction and intelligent decision-making system, July 2010,IEEE
3. Refonaa, J. ; Dept. of Comput. Sci. & Engineeering, Sathyabama Univ., Chennai, India ; Lakshmi, M. ; Vivek, V.Analysis and prediction of natural disaster using spatial data mining technique, March 2015 ,IEEE
4. Dazhong Xia ; Coll. of Hydrol. & Water Resources, Hohai Univ., Nanjing, China ; Xingnan Zhang ; Yuanhao Fang,Study and application of storm surge early-warning and prediction and disaster evaluation system, Sept. 2011,IEEE
5. http://www.historiasiglo20.org/../ /Can_Natural_Disasters_Be_Predicted_1_.do
6. http://www.nature.com/nature /journal/ v247 /n5442/abs/ 247 534a0.html
7. A survey for different approaches of Outlier Detection in data mining, Chandarana, D.R.; Dhamecha, M.V.,2015,IEEE
8. Outlier Detection Algorithms in Data Mining, Jingke Xi ; Sch. of Comput. Sci. & Technol., China Univ. of Min. & Technol., Xuzhou, Dec. 2008,IEEE
9. JMX Based Metrics Collection and Monitoring of Java application ,Manoj Sakhala , Supriya Jahagirdar , Shradha Baranwal , Pratik Magar , Prof S. Karande
10. http://smap.jpl.nasa.gov/ -NASA's SMAP satellite website [For the data sets and live feeds]
11. http://www.imd.gov.in/Welcome%20To%20IMD/Welcome.php – Indian MeterologicalDepartment
12. http://openweathermap.org/api -APIs for weather live data
13. https://data.gov.in/ -Open Government Data latform,India