RISK ASSESMENT OF NATURAL DISASTER USING WEB MINING

Anagha .S. Chandran,
Department Of Information Technology
Anna University Regional Campus
Coimbatore,India.

Mannar Mannan.J,
Department Of Information Technology
Anna University Regional Campus
Coimbatore,India.

Abstract: Internet of Things (IoTs) based prototype is used for disaster risk analysis and effective pre and post management of disaster stricken areas. Data mining techniques along with IoT are used to facilitate information sharing as early as possible and to ensure participation of as many as private and public service providers for major disaster prediction and recovery processes. The main challenge faced during prediction about the chances of occurrences of disasters is the analysis of the available dynamic data and information sharing at an early stage. In this paper, integrated data from different meteorological sources have to be collected and analysed using clustering and mining algorithms to predict natural calamities. Semantic analysis using ontology has been taken to ensure the prediction. On detecting this trust worthy abnormal conditions, cyclone can be predicted and convey this information to the service providers via internet and to the public as early as possible.

Keywords: Cyclone, Data mining, Ontology, Internet of Things

I.INTRODUCTION

In Indian peninsula almost 85% (according to NDMA) of the crisis occurred is due to multiple hazards. Among them the most devastating phenomena is based on tropical cyclones (TC). The main challenge faced during prediction about the chances of occurrences of cyclones is the analysis of the available dynamic data and information sharing at an early stage. There are many computer based methods to predict about the occurrences of cyclone. Data mining techniques along with IoT are the major methods used for environmental disaster risk and effect management. Data mining is a powerful new technology with great potential to the extraction of hidden predictive information from large data base. Data mining is used to predict future trends and conduct, allowing the work to make proactive, knowledge-driven decisions. The automated, future analysis offered by data mining move beyond the analysis of past events provided by backdated models of decision support systems. They clean databases for hidden patterns, identifying predictive information that experts may miss because it may lie beyond their expectations.

A. Data Mining in Disaster Management

Weather forecasting and to make an accurate prediction is one of the major challenges faced by the meteorologist all over the world. Weather forecasting entails predicting how the present state of the atmosphere will change. Adam Zagorecki et al. explains the prediction was carried out based on Artificial Neural Network and Decision Tree algorithms [10]. A data model for the meteorological data was developed and was used to train the classifier algorithms. These algorithms were balanced using standard performance metrics and gave the best outcomes used to generate classification rules for the mean weather variables.Shen Bin et al., suggested [14] the data mining techniques is used to effectively discover, collect, organize, search and disseminate real time disaster information. Mainly two parallel systems are found developed and operative, which are a web-based prototype and mobile applications which aims to reduce the vulnerability of people.

B. Disaster Management Ontology

The domain of data mining (DM) deals with analyzing different types of data. A major problem is to treat and represent the mining of different types of structured data such as graphs, sequences, networks, text, image, multimedia and relational data in a uniform fashion and also support of the Knowledge Discovery in Databases (KDD) process.A formalization of the domain of data mining is a solution that addresses these challenges which is motivated by a reference modular ontology called OntoDM.Ontologyplays a major role in the format of Semanti c Web integration.Hereinformation is given well-defined meaning and theSearch engines deploy ontology to find pages with words that are syntactically different but semantically similar suggested by St.Werder et al.,[12].

C. Data Mining and IoT

The Internet of Things (IoT) introduce to the next generation of Internet. It is the next uprising technology after the revolution of computer and Internet. It incorporates the new technologies of computing and communications. The prediction of occurrence of tropical cyclone is a complication. Even though many researches were undertaken about the most accurate prediction it couldn't be able to forecast the correct intensity path. In the work of HajoRijgersberg[11] a supervised learning methodology is carried out which is an alternative method to the hurricane intensity prediction. The main objective is to produce accurate early warnings by analysing changes in wind speed of a particular storm. The two methods used are: particle swarm optimization and classification and class association rules which areused to discover patterns and convey this knowledge into sets of rules.

D. The Formation Of Cyclone

The most devastating happenings among all natural disasters are tropical cyclones (TC). Cyclones are to occur by very strong winds, torrential rains and storm surges. The destruction caused by cyclones to shipping in the high seas and coastal surroundings along the Indian coasts due to above mentioned adverse weather. The tropical warm North Indian

Ocean (NIO), the tropical North Atlantic, the South Pacific and the North West Pacific, is a breeding ground for the disastrous TC phenomenon.

The forecasting and prediction of cyclone is a challenging and difficult task. The parameters to be analysed are changing dynamically. Tropical Cyclone formation is divided into three stages.

1. Formation and initial development
2. Mature tropical cyclones
3. Modification and decay

The main objective is to produce error free early warnings with respect to changes in the wind speed and pressure difference.

Table 1.1 Classification Cyclonic disturbances Over the NIO (since 1999)

Low pressure system	Maximum Sustained Surface Winds
Low pressure area (L)	< 17 knots
Depression (D)	17 – 27 knots
Deep Depression (DD)	28 – 33 knots
Cyclonic storm (CS)	34 – 47 knots
Severe Cyclonic storm (SCS)	48 – 63 knots
Very Severe Cyclonic storm(VSCS)	64 – 119 knots
Extreme Super Cyclonic storm(ESCS)	120 knots & above

II. RELATED WORKS

Early warning is a major component of disaster risk reduction. Early Warning System (EWS) with semantic Internet of Thing (IoT) recommended by Stefan Posland et al., [1] is used for environmental disaster risk management. Data collected from sensors are transmitted through Information Communication Technology (ICT)to semantic web. The EWS include easier sensor data source and more dynamic metadata driven data analysis. Heavy weight semantic is selectively used to collect data from multi sensor IoT. Multiple domain ontologies may need to be combined for multiple knowledge representation. AnEWS is fundamental type of IoT information system used for environment disaster risk and effect management. It assist prevent loss of life and reduces the economic and material impact of disasters. An EWS differs from other types of environment ICT monitoring systems via four main functions: Risk analysis; Timely monitoring and warning; Dissemination and communication information; Response capacity.The continually generatedresult sets are updated by online data-mining and data-fusion techniques, each creating data at a variety of semantic levels.

EWS aims to minimize the vulnerability of people to calamities by disseminating warnings. A fail-proof EWS and conveyance system and network would be crucial in dealing with post-calamity situations. A reference model and methodology for building a new EWS as framework is meant to solve problems common to all EWS, so that creation of a new EWS is reduced as much as possible said by BartoszBalis et al., [2]. A scientific challenge is to find the best methods and tools for adjustable and efficient integration and analysis of these computer and data intensive applications in a way that fulfils the requirements specific for the EWSs as a class of systems.

Since atmospheric values are dynamic in nature it cannot be predicted accurately. Weather forecasting is an important approach in meteorology and is also the most challenging problem. Folorunsho and Adesesan suggested [3] the data mining techniques is used for forecasting the parameters like maximum temperature, rainfall, evaporation and wind speed. The prediction was carried out based on Artificial Neural Network and Decision Tree algorithms. A data model for the meteorological data was developed and was used to train the classifier algorithms. The accomplishment of these algorithms was balanced using classification rules and gave the best outcomes. A predictive Neural Network model developed for the weather prediction and the output is generated based on the comparison between input data with actual weather data for the predicted periods and the observed patterns.

Knowledge about occurrences of disaster is conveyed mainly based on data mining methods. The data mining method is well organised to collect, organize, identify and disseminate real-time disaster information. This is useful foe efficient crisis management and disaster recovery tasks. Li Zheng et al., proposed [4] a unified framework prototype was developed to integrates the different techniques with applications separately which includes two parallel system a web-based prototype of a Business Continuity Information Network system and an All-Hazard Disaster Situation Browser system that run on mobile devices. They are used to share information and collaborative platforms for preparedness and disaster recovery to better understand the current disaster situation is and how getting back of community.

A monsoon conduit is a significant origin region for tropical cyclone. The S shaped tracks of tropical cyclone tend to be related with reverse-oriented monsoon troughs. The arrival of monsoon is eagerly awaited in the Indian sub-continent as it has deep down impact in the economic and social domain. The emergence of satellite imagery is possible to monitor the different parameters which influence or gets affected by the monsoon in a more global scale. According to Dinu John and Dr. B BMesh ram [5] the onset of monsoon is predicted based on the features bring out from satellite images using data mining approach.

Reliable forecasting of Tropical Cyclone (TC) track and intensity are crucial for protecting the lives and property of communities in areas that are subject to TC impact. As per the thesis of Michale L Jankulak [6] associative classifier called Apriori Grad based on both association rule mining technique and apriori algorithm for frequent item set is used. It includes made to order for detecting rare events and for categorizing a series of interrelated classification the occurrences as yes/no thresholds on an underlying continuous measurement of 24*7 TC intensity change. The performance on this domain is compared to a various classification techniques, and suggestions for possible development as an operational

forecasting device or for future meteorological study are examined.

A Knowledge Management System (KMS) are essential for disaster detection, response planning and management. The KMS incorporates right set of real time information governed by an effective information system which helps in early warning ,disaster response and recovery management through consolidation and collaboration from different organizations and agencies. Besides during Humanitarian Assistance and Disaster Relief (HA/DR) also require collaboration of various agencies in a concerted and timely manner said by Nuha Abdulla Hassan et al., [7] . Users find difficulties to put up and convert the vast diversity of data into real-time, multi-dimensional information that could be comprised in to common HA/DR situational awareness pictures. Additionally, managing knowledge across various entities involved in such effort is critical. This includes right set of information that is timely, relevant, and is governed by an effective communication process.

The forecasting and prediction of natural calamities can be done by analysing multiple learning algorithms such as supervised learning, clustering, classification etc. These methods together called ensemble methods which are suitable to obtain better prediction achievement. Sim D Aberson et al., [8] proposed GFDL hurricane model based ensemble forecasting is used to predict tropical cyclone intensity. The ensembles are able to quantify the potential errors in each forecast but does not show forecast skill until 36 hrs. The model itself is not effective to predict while sudden changes occur. Improvements to the vertex initialization in the GFDL model and to the NCEP ensembles are under progress which will help to improve the forecasts in the future.

The data collected and analysed is to be give out to public by the medium of Internet of Things (IoT). IoTs is a web of multiple devices which includes vehicles, buildings, electronic devices etc fix with sensors, internet connection and electronic softwares. The prediction of a problem domain using data mining in IoT is a challenging task. Kyriakos and Andrew suggested [9] four different data mining models for the Internet of Things are proposed, which are Multi-layer data mining model which used to collect, pre-process, process and servicing the data., Distributed data mining model is used to manage problems by accumulating data from various sites., Grid based data mining model achieves the function of data mining by grid framework and data mining model from multi-technology integration perspective reports the corresponding framework for the future Internet.

The advancement in information technology have had an ardent impact on disaster management by making unvalued volumes of data available to the decision makers. Therefore new challenges are faced related to the effective management of vast amount of data. Adam Zagorecki et al. [10] explains the application of data mining and machine learning techniques to support the decision-making processes for the disaster and crisis management. The problem-specific applications of data mining and machine learning that addresses an extensive scope of problems, such as situational awareness and real-time threat assessment using diverse streams of data. They apex and deny the challenges and indicate future trends and also providing situational awareness and real-time threat assessment. One feasible direction would be applying social network data mining to real-time analysis of data supplied from mobile applications used during the response phase of disaster management. During this assessment, the problem of the use of replicate data has been discussed. The lack of actual data is often addressed by using duplicate data. Such artificial data can lead to performance evaluations of the proposed algorithms.

Ontology supports making quantitative research data more explicit, so that data can be integrated, verified and reproduced. Ontology is a formal naming and definition of types, properties and interrelationships of entities that really or fundamentally exist for a particular domain or discourse said by HajoRijgersberg[11]. These include mathematical applications. OWL based ontology editor or manager that would allow specifying the data type of meta-knowledge and generating the ontology from it. Such issues might also play a role in other ontologies, so that this is a lacking component in ontology management life cycle. Similarly the data procurement for visualization of geo – related information in situation maps is based on hundreds of written messages containing daily information from damage sites. St.Werder et al., [12] explains historic military domain data model is scrutinized and an ontology support knowledge base is created. The automated geo-spatial data acquisition and mapping provides the conversion of verbally given geo-information in to a situation map. The said model is based on the Command and Control Information Exchange Data Model.

The data collected from different sensors and devices have various types of differences like temperature, light, sound quality, validity etc. Developing solutions that can effectively analyse and interpret physical world data is a real challenge. Integrating the real world data to the web and providing web based interaction with IoT is described by PayamBarnaghi et al., [13]. Web of Things (WoT) is a driving force to generate access and integrate data from physical cyber and social sources. WoT data is combination of voluminous, continuous, streaming, real time, dynamic and volatile. The distribution and efficient scalable managing of data in WoT, in addition to enhanced data publication and dissemination, will be dependent on powerful mechanisms for in-network processing, aggregation and summarization. The ability to analyse and interpret this data to create meaningful understanding, extract knowledge and create situational awareness is crucial to fulfilling the potential of big WoT data.

The prediction of occurrence of tropical cyclone is a complication. Although many researches were undertaken about the most accurate prediction it couldn't be able to forecast the correct intensity path. Shen Bin et al., [14] introduced a supervised learning methodology is carried out which is an alternative method to the hurricane intensity prediction. The main objective is to produce accurate early warnings by analysing changes in wind speed of a particular storm. The two methods used are: particle swarm optimization which employs on a population based search method to optimize a rule quality function and pattern identification, classification and class association rules bear trends in the data

and transfers this knowledge into sets of rules. Other methods used are entropy based pre-processing techniques, feature selection methodologies and classifier combination schemes. Even though there are free availability of online weather forecasting channel, the spot accuracy and conveyance of the regional weather information is not executed properly. Yi Zhang and Vic I. Hanby [15] discuss the possibilities of short term prediction of temperature and solar radiation based on the comparison of meteorological forecast data and local observation data.

III. RISK ASSESMENT OF NATURAL DISASTER USING WEB MINING DESIGN AND IMPLEMENTATION

In order to achieve the immediate and accurate prediction, the analysis of meteorological data sets is done along with following three distinct categories.

1. Data mining techniques such as classification and clustering.
2. Semantic Analysis using Ontology.
3. Tomcat server is used to connect with web application

The following is the model of the proposed work. The collected Meteorological datasets are taken as input and are taken for Data cleaning in order to remove noise, irrelevant data. Then the data can be taken for classification depends upon the parameters such as temperature, wind, rainfall and environmental factors. The classified data are performed with Semantic relationships between given input data and analyzed data. Finally, the output is determined as whether the possibilities of occurrence of cyclone is there or not.

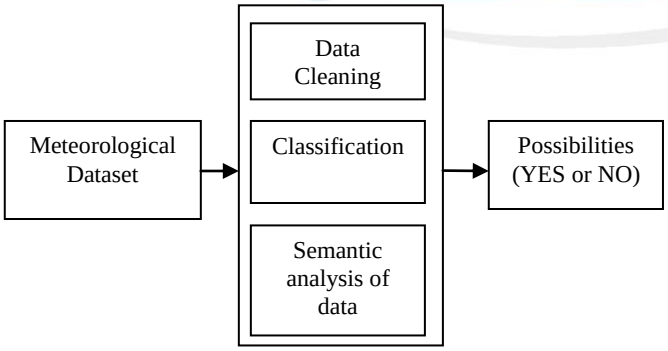

Figure 1: Proposed Works

A. PATTERN MINING AND CLASSIFICATION

a) Decision Tree algorithm J48:

This method of classification is used to construct decision tree to perform classification model. Once the tree is built, it is applied to each element in row in the database and results in classification for that row's element. While building a tree, J48 neglect the missing values i.e. the value by which prediction can be perform has taken as training sample. J48 performs classification using Decision trees is incorporated in Weka tool.

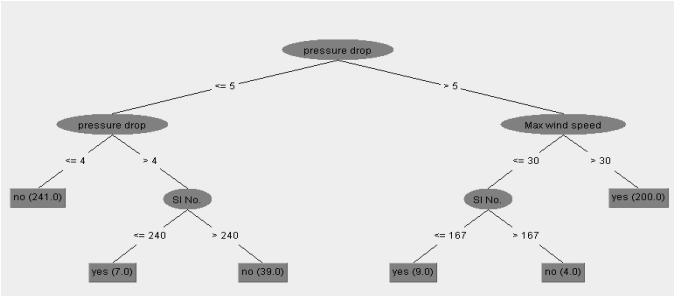

Figure 2: Decision Tree

b) Naïve Bayes classifier:

The Naïve Bayes algorithm is a probability function based classifier which performs calculation on a set of probabilities by calculating the frequency and combinations of values in a given dataset.

Naïve Bayes classifier is based on Bayes theorem and the total probability theorem. The probability that a document d with vector x=<x1, x2…xn>

$$P(h1 \mid xi) = P(xi \mid hi).\frac{P(h1)}{P(xi \mid h1)..P(h1) + P(xi \mid h2).P(h2)} \quad (1)$$

Here, P(h1|xi) is posterior probability, while P(h1) is the prior probability associated with hypothesis h1.

$$P(xi) = \sum_{k=0}^{n} P(xi \mid hi)P(hi) \quad (2)$$

Thus, we have

$$P(h1|xi) = P(xi|h1)P(h1)/P(xi) \quad (3)$$

The frequent pattern mined which satisfies the below condition are taken as significant Frequent Pattern

$$SW(i) = \sum_{i-0}^{n} m * t * n \quad (4)$$

Where SW(i) is the significant feature of pattern, m is no. of data;t is no. of attributes;n is no. of times occurences of value.

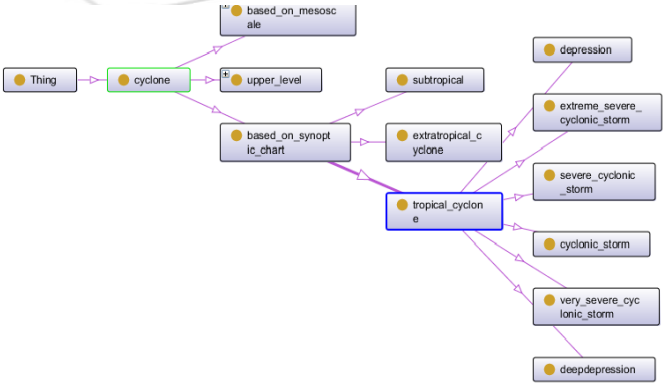

Figure 3: OntoGraf of cyclone occurrence

The above figure explains the classification of cyclone based on weather map and climatic scales.

IV. CONCLUSION

An effective EWS which can reduce damages caused is the main objective of the works undertaken here.Early detection of the abnormal conditions and communication of the accurate situation to the mass and service providers to initiate action at an early stage.

1. The Data mining algorithms and IoTs are used together foran early, trust worthy prediction of occurrences of cyclone in a particular area.
2. Data collected from Indian Meteorological Department (IMD) is stored in a database and is compared with the current input to predict the occurrences of cyclone as early as possible so as to initiate remedial actions at an early stage.
3. Semantic analysis using ontology is used for the prediction of cyclone.
4. Tomcat server is used to convey the above time critical information to the publicas well as remedial / rescue operators and various government departments, media etc.

V.REFERENCES

[1] Stefan Posland, Stuart .E. Middleton,FernandoChaves,RanTao,OcalNecmioglu and Ulrich Bugel,``A Semantic IoT Early Warning System for Natural Environmental Crisis Management" *IEEE Transactions on Emerging Topics In Computing., vol. 3*,no. 2, June 2015.

[2] B. Balis*et al., ``The urban flood common information space for early*warning systems," in *Proc. Int. Conf. Comput. Sci. (ICCS), vol. 4. 2011*,pp. 96105.

[3] FolorunshoOlaiya and Adesesan Barnabas Adeyemo" Application of Data Mining Techniques in Weather Prediction and Climate Change Studies"*IJEEB.,vol.1, July 2012, pp. 51-59.*

[4] Li Zheng, Chao Shen, Liang Tang, Chunqiu Zeng, Tao Li, Steve Luis, and Shu-Ching Chen,Stefan Posland, ``Data Mining Meets the Needs of Disaster Information Management" *IEEE Transactions on Human-Machine System.., vol. 43*,no. 5, September 2013.

[5] Dinu John and Dr. B. B. Meshram" A Data mining approach for Monsoon prediction using Satellite image data"*International Journal of Computer Science & Communication Networks., vol. 2(3) ,pp. 421-424.*

[6] Jankulak, Michael L., "Prediction of Rapid Intensity Changes in Tropical Cyclones Using Associative Classification" (2012). *Open Access Theses.*Paper 364.

[7] Nuha Abdullah Hassan,Nur-AdibHayiyusuh and Rasha K. Nouri, "The Implementation of Knowledge Management System (KMS) for the Support of Humanitarian Assistance/Disaster Relief (HA/DR) in Malaysia"*IJHSS ., vol. 1*,no. 4, April 2011

[8] Sim D. Aberson, Morris A. Bender, and Robert E. Tuleya,``Ensemble Forecasting Of Tropical Cyclone Intensity". P 1.31.

[9] KyriakosChatzidimitriouandAndrew Sutton ,``Alternative Data Mining Techniques for Predicting Tropical Cyclone Intensification" *vol. 3*,no. 2, June 2015.

[10] Adam Zagorecki, David Johnson and JozefRistvej,"Data mining and machine learning in the context of disaster and crisis management" Article in IJEM ., January 2013.

[11] H. Rijgersberg, M. van Assem, and J. Top, ``Ontology of units of measureand related concepts," *Semantic Web, vol. 4, no. 1, pp. 313, 2013*

[12] Ch. Lucas, St. Werder, H.-P. Bähr "Information Mining For Disaster Management"International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences , 36(3/W49A)., *September* 2015.

[13] P. Barnaghi, A. Sheth, and C. Henson, ``From data to actionable knowledge:Big data challenges in theWeb of Things" *IEEE Intell. Syst., vol. 28*,no. 6, pp. 611, Nov./Dec. 2013.

[14] Shen Bin, Liu Yuan, Wang Xiaoyi,''Research on Data Mining Models for the Internet of Things" *IEEE,2010*

[15] Yi Zhang, Vic I.Hanby,``Short-Term Prediction Of Weather Parameters Using Online Weather Forecasts"' Building Simulation 2007