Earthquake Prediction based on Spatio - Temporal Data Mining Approach

Vijayasankari S[$] and Indhuja P
Department of Computer Application, E M G Yadava Women's College, Madurai 625014
[$]Corresponding Author, E-Mail - svsankari@hotmail.com

ABSTRACT

Disaster Management is an organization and management of resources and responsibilities for dealing with all human iteration aspects of emergencies, in particular preparedness, response and recovery in order to minimize the impact of disasters. Apriori is one of the most crucial algorithm for frequent item set finding. In this present work also deals with analyzing the earthquakes, that can range in size from those that are so weak and that cannot be felt to those violent enough to toss people around and destroy whole cities. Here a deep learning technique called Long Short-Term Memory (LSTM) networks to learn the spatio - temporal relationship among earthquakes in different locations and make predictions by taking advantage of that relationship was employed for the analysis.

Keywords - Disaster Management, Earthquake, Destroy, Data mining, Spatio - Temporal.

1. Introduction

Data Mining is the non-trivial extraction of implicit, previously unknown and potential useful information from the data and the process of sorting through large data sets to identify patterns and establish relationships to solve problems through data analysis. Data mining tools allow enterprises to predict future trends. Data mining is the computing process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics, and database systems. It is an interdisciplinary subfield of computer application. The overall goal of this data mining process is to extract information from a data set and transform it into an understandable structure for further use. Aside from the raw analysis step, it involves database and data management aspects, data pre-processing, model and inference considerations, interestingness metrics, complexity considerations, post-processing of discovered structures, visualization, and online updating. Data mining is the analysis step of the knowledge discovery in databases (KDD). Data mining information on historical sales patterns and customer behaviors can be used to build prediction models for future sales, new products and services. Companies in the financial industry use data mining tools to build risk models and detect fraud. The manufacturing industry uses data mining tools to improve product safety, identify quality issues, manage the supply chain and improve operations.

Earthquakes are one of the most destructive natural disasters. They usually occur without warning and do not allow much time for people to react. Therefore, earthquakes can cause serious injuries and loss of life and destroy tremendous buildings and infrastructure, leading to great economy loss. The prediction of earthquakes is obviously critical to the safety of our society, but it has been proven to be a very challenging issue in seismology.

According to the employed methodologies, existing works on earthquake prediction can be mainly classified into four categories, like mathematical analysis, precursor signal investigation, machine learning algorithms like decision trees and support vector machines (SVM) and deep learning. In this the first type of work tries to formulate the earthquake prediction problem by using different mathematical tools, like the FDL (Fibonacci, Dual and Lucas) method, kinds of probability distribution or other mathematics proving and spatial connection theory. In the second type of work, researchers study earthquake precursor signals

to help with earthquake prediction. For example, electromagnetic signals, aerosol optical depth (AOD), lithosphere-atmosphere-ionosphere and cloud image have been explored. Even animals abnormal behavior has been taken into account in this kind of study. The third type of work mainly explores data mining and time series analysis methods, such as J48, multi-objective info-fuzzy network (M-IFN), k-nearest neighbors (kNN), support vector machine (SVM) and artificial neural networks (ANNs) to predict the magnitude of the largest earthquake in the next year based on the previously recorded seismic events in the same region. In the fourth type of work, deep learning algorithms are utilized to predict both the magnitude and the time of major seismic events. Various kinds of neural networks have been adopted, such as multi-layer perceptron (MLP), backward propagation (BP) neural network, feed forward neural network (FFNN), recurrent neural network (RNN), which can work under certain particular circumstances. Although there have been a lot of works on earthquake prediction, very few of them can predict future seismic events accurately. The reason is that the occurrence of earthquakes involves processes of very high complexity and depends on a large number of factors that are difficult to analyze.

The reason is that the occurrence of earthquakes involves processes of very high complexity and depends on a large number of factors that are difficult to analyze. There are obviously complex nonlinear correlations among earthquake occurrences, because of which traditional mathematical, statistical, and machine learning methods cannot analyze well in this process. Recently, deep learning methods like RNNs are shown to be able to capture the nonlinear correlations among data. Particularly, they are mostly used to analyze time-series data so as to make predictions. As a result, when previous works use deep learning to make predictions, they predict earthquakes in a particular location only based on the history time-series data in that location, and still not possible to get good results. In contrast, we contend that the spatio-temporal correlations among history earthquake data have to be investigated in order to make more accurate predictions. To this end, in this paper we investigate earthquake prediction from a spatio-temporal perspective. Specifically, we devise an earthquake prediction scheme by adjusting a long short-term memory (LSTM) network, which is an advanced RNN and has strong nonlinear learning capability even on the data containing long-term interval correlations that the RNN is not able to achieve.

2. Literature Review

Earthquakes are sudden disturbances in the earth's crust that cause huge loss to life and property. It focus on the importance of Information and Communication Technology (ICT) for disaster management and how ICT can be helpful in disseminating earthquakes alerts also with application of wireless sensor network have proposed by Munib Ur. Rehman et al (2016). Earthquake is an unpredictable natural phenomenon that create a vast amount of damage, affecting communities and their environment. To reduce the effects of such hazards, frameworks like building resilience have emerged. These frameworks target on increasing recovery after such disaster, by introducing new designs, technologies, and components to the building have proposed by Kahandawa K.A.R.V.D et al (2017). Chien-Pang-Lee and Yungho Leu, (2011) have used a novel hybrid method for feature selection in microarray data analysis. This method uses a genetic algorithm with dynamic parameters (GADP) setting to generate a number of subsets of genes and to rank the genes according to their occurrence frequencies in three gene subsets. David Meyer et al (2002), have done a benchmark study on comparison of SVMs with sixteen classification methods based on their performance on twenty one dataset from UCI machine learning databases. Hanchuan Peng et al (2005) have proposed many filter based feature or genes selection methods for microarray dataset.

Hua J. Tembe .W and Dougherty. E (2009) have proposed that feature selection is an unavoidable part of classifier design. The objective of feature selection is to select a subset of

features from the possible high dimensional features spaces to low dimensional spaces, which are better suited for retrieval or learning purposes. Hettiarachchi. S (2018) have proposed that the Indian Ocean Tsunami of Boxing Day, December 2004, caused loss of lives and widespread damage to infrastructure and ecosystems across Indian Ocean States. IOTWMS-Indian Ocean Tsunami Warning and Mitigation System as a highly successful end to end warning system under UNESCO / intergovernmental oceanographic commission (IOC). Liyuv Huang et al (2017) developed a novel precursory wave detector by using liquid-suspension principle and super- low frequency (SLF) sound signal detection technology in submarine is presented, then the device of earthquake precursory wave detection and the corresponding software are developed.

Takumi Yoshii et al (2018) have developed the process of tsunami induction, sediment transport and deposition on costal low lands are investigated experimentally. Wang Honghui et al (2017) analized the sudden landslide, collapse, debris flow is one of the most serious geological disaster type in our country because of the short time, strong concealment and the destructive, easy to cause significant casualties and the huge economic losses. For the need of improving the capacity of the geohazards monitoring a professional monitoring system has been developed based on the Internet of things, WSN (Wireless Sensor Network) and network communication technology.

3. Proposed Work

A deep learning technique called long short-term memory (LSTM) networks to learn the spatio-temporal relationship among earthquakes in different locations and make predictions by taking advantage of that relationship. Simulation results show that the LSTM network with two-dimensional input developed in this paper is able to discover and exploit the spatio-temporal correlations among earthquakes to make better predictions than before Clustering represents the algorithm's offline component which uses the data captured by the online component. For simplicity we discuss two-dimensional data first and later discuss implications for higher-dimensional data

Discovering frequent item sets is time consuming on its own, yet mining patterns with temporal information makes it even harder to find a solution in a reasonable time. Therefore, it seems necessary to optimize the performance of the algorithm. We propose a special implementation of a linked list data structure. As discussed earlier, to consider time hierarchies for mining patterns, TCs are proposed. Via linked list data structure, each entity is referenced to its upper and lower level and also to its previous and following entity. Furthermore, there is need to re-trieve transactions with their time stamp information easily. To this end, hash-based data structure is proposed. The obvious outcomes of such disasters are increased number of casualties or individuals requiring treatment, which may overwhelm the available resources in the traditional medical care system. This will lead to a sudden increase or surge in demand for health professionals trained in disaster management. he primary role played by dentists in disasters has been in the form of forensic odontology and the assistance of dental members in Disaster Mortuary Operational Response Teams. Dentists have also played an important role by rendering their services in the armed forces sector. Dental professionals are adept at various skills that can contribute to efficient disaster management. They can be of immense help in obtaining history, taking and interpreting radiographs, record maintenance and data management, infection control procedures, wound management and suturing, managing infections and prescription of medications, immunization procedures, distributing medical sup-plies, providing information to patients and public, man-aging victim triage and patient management, and so forth.

3.1 Load Training Dataset

In this Process to load diabetes dataset to process. And then to insert the dataset on database dynamically. After that insert the new diabetes report on database. Dataset should be loaded after preprocessing automatically and also inserted into database newely whenever it run the process. Here also two types of dataset will be loaded into the given database. Both dataset is loaded and after data is partitioned and then finding results.

3.2 Pre-processing

A preprocessor is a program that processes its input data to produce output that is used as input to another program. Pre-processing as suppose dataset loading with after any data placement is to be null or unstructured data or unwanted data is to be removed. The output is said to be a preprocessed form of the input data, which is often used by some subsequent programs like compilers. Preprocess is removing null values or un structured data from loading into the certain datasets. Here after checking attributes is to be null, then only we check the unwanted data or null data.

3.3 Data Partitioning

Data Partition is also similar to data clustering. Cluster is a group of objects that belongs to the same class. In other words, similar objects are grouped in one cluster and dissimilar objects are grouped in another cluster. Here different set of attributes based load the certain dataset using clustering our data based on some valid attributes. Cluster analysis itself is not one specific algorithm, but the general task to be solved. Partition has our data set is splited into some of small files. So data set is partitioned for data classification method.

3.4 Event based Data Classification

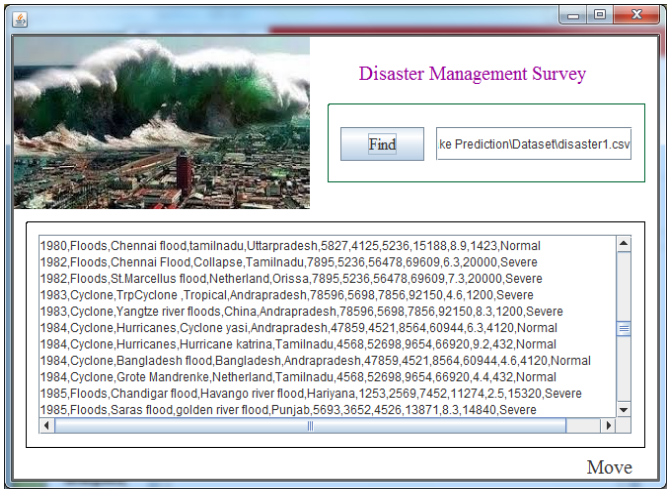
In machine learning algorithms based data is going to be classified. Here different analyzing algorithms as apriori and recommendation algorithms using finding frequent data through given method 1 dataset. Another Method vector machine based analyze feature extraction and specified by the sub types and event types based disasters. Given a set of training examples, each marked as belonging to one or the other of two categories, training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier.

3.5 Finding Frequent Item

Frequent Item set finding is one of most important methods for data mining process. Frequent finding with classification data is only applicable for frequent data selection list. First data is clustering and after classification result based analyzing algorithms. Because every algorithms has some general method for classification process. So here some of algorithms as using for apriori and Cantree and Gtree process.

4. Result and Discussion

A dataset are retrieved from Mockaroo application with total number of 803 datasets. All dataset records are loaded successfully in the database table as shown in the figure 1.



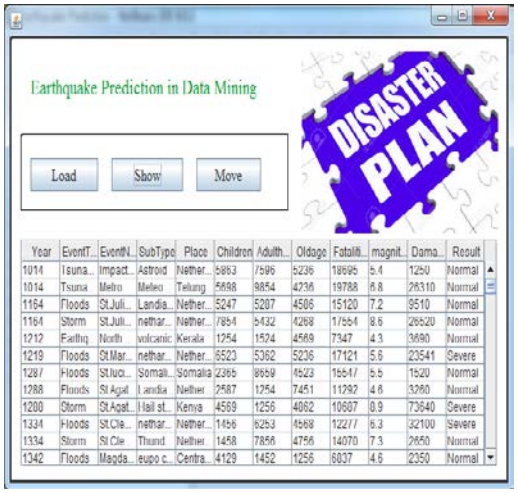
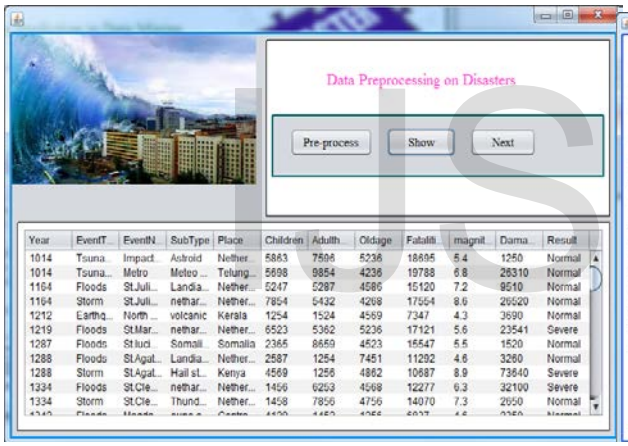


Fig.1 – Earthquake Dataset Table **Fig.2 – Earthquake prediction**

Earthquake prediction in data mining is done with load, show, move task are shown in the figure 2. Figure 3 represents the data preprocessing on disasters which contains the various tasks like preprocess, show, next. Figure 4 represents disaster management based data clustering which contains datasets for earthquake is 61, flood is 56, cyclone is 26, Tsunami is 13, storms is 2. In figure 5 represent earthquake classification and figure 6 represents the SVM



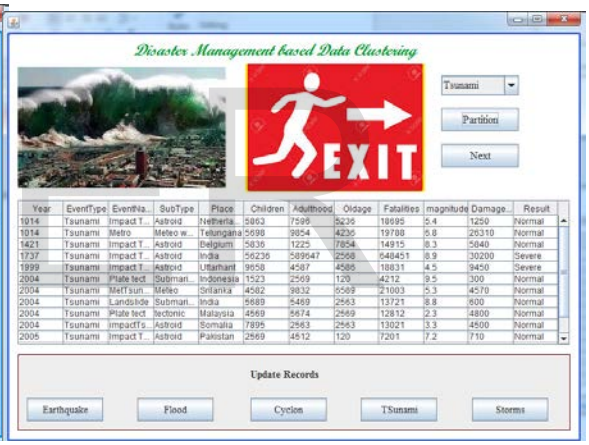
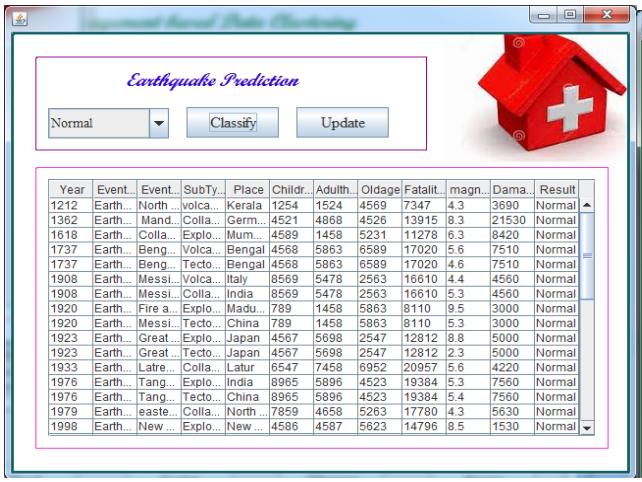


Fig. 3 Data preprocessing on disasters **Fig. 4 Disaster management based data clustering**



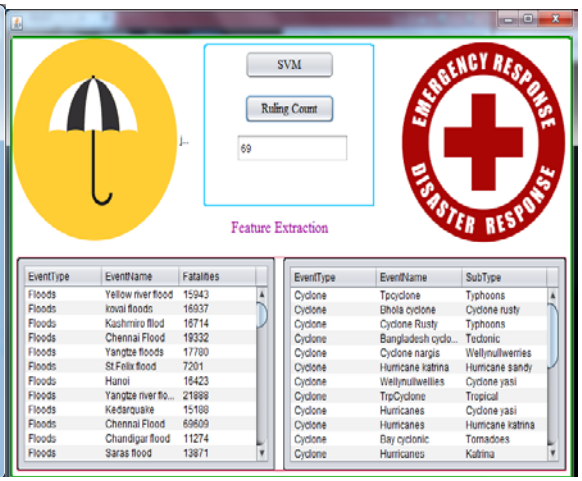


Fig.5 Earthquake classification **Fig. 6 SVM- feature extraction**

feature extraction. From the figure 7 shows the data efficiency and security is compared both for existing and proposed systems. In figure 8 shows the prediction level with average performance for existing and proposed system

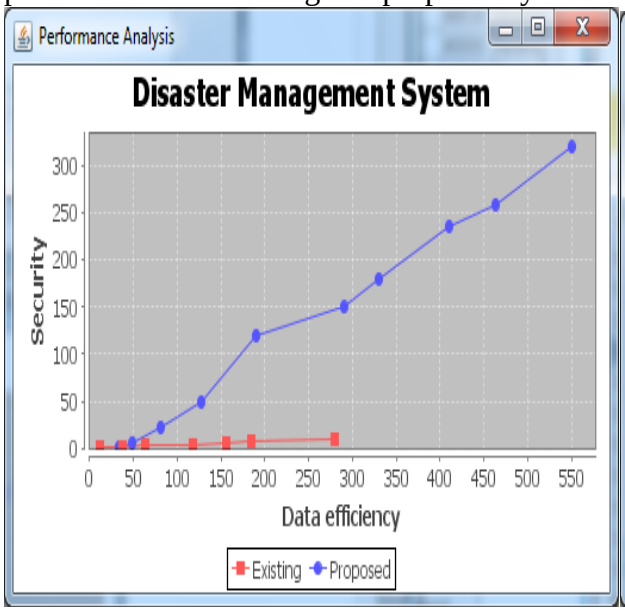


Fig.7 Disaster management system

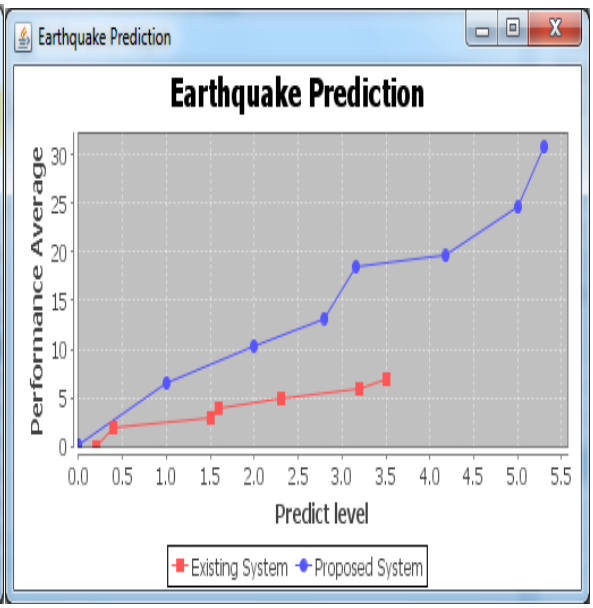


Fig.8 Earthquake prediction

Conclusion

In this present proposed work a new earthquake prediction system from the spatio-temporal perspective is used. Specifically, design an LSTM network with two-dimensional input, which can discover the spatio-temporal correlations among earthquake occurrences and take advantage of the correlations to make accurate earthquake predictions. The proposed decomposition method for improving the effectiveness and efficiency of our LSTM network has been shown to be able to significantly improve the system performance. Simulation results also demonstrate that our system can make accurate predictions with different temporal and spatial prediction granularities. Here also discuss with finding frequent item sets and streaming data mining based methodologies for accessing data controls mechanism. Also an efficient procedure to search in a solution s pace was proposed. Experiments on synthetic datasets showed that the proposed Algorithm is quite effective.

Future Enhancements

In this work the same process with the usage of different algorithm is continued. Experiments and analysis confirm the effectiveness of our schemes and design. In the future, it shall extend the disaster managements system result predictions and safety mechanisms included.

References

1. Munib ur Rahman, Soliha Rahman, Seema Mansoor, Vikas Deep, and Aashkaar M, (2016) 'Implementation of ICT and wireless sensor networks for earthquake alert and disaster management in earthquake prone areas, Procedia computer science, 85, pp.92-99.
2. Kahandawa K.A.R.V.D, Domingo N.D, Park K.S, Uma S.R, (2018), 'Earthquake damage estimation systems: Literature review,' Procedia Engineering, 212, pp.622-628.
3. Chien-Pang-Lee and Yungho Leu, (2011), 'A novel hybrid feature selection method for micro array data analysis', Applied soft computing, 11, pp.208-213.
4. David Meyer, Friedrich Leisch and Kurt Hornik (2002), 'Bench marking support vector machines, SFB adaptive information systems and modelling in economic and management

science' Vienna University of Economics and Business Administration, Augasse 2–6, 1090 Wien, Austria.
5. Hanchuan Peng, Fuhui Long and Chris Ding (2005), 'Feature selection based on mutual information: criteria of Max_Dependency, Max_ Relevance and Min_Redundancy', IEEE Transactions Pattern Analysis and Machine Intelligence, 27, pp.1226-1238.
6. Hua J. Tembe. W and Dougherty, E, (2009), 'Performance of feature selection methods in the classification of high dimensional data' Pattern Recognition, 42, pp.409-424.
7. Hettiarachchi, S (2018), 'Establishing the Indian ocean Tsunami Warning and Mitigation System for human and environmental security' Procedia Engineering, 212, pp.1339-1346.
8. Liyuv Huang, Yingju Du, Jianing Zheng, Yuxiang Zhang, (2017)' Detection of Precursory Wave Using a Novel Sensor and Its Application to Earthquake Prediction' AASRI Procedia, 3, pp. 613-618.
9. Takumi Yoshii, Shiro Tanaka, Masafumi Matsuyama (2018), 'Tsunami inundation, sediment transport and deposition process of tsunami deposits on coastal lowland inferred from the Tsunami sand transport laboratory experiment, Marine Geology, 400, pp.107-118.
10. Wang Honghui, Tuo Xianguo, Li Yan, Liu Qi, Nie Donglin, Meng Lingyu, Yang Jiaxin, (2017), Research of the Hardware Architecture of the Geohazards Monitoring and early warning system based on IOT' Procedia computer science, 107, pp.111-116.