|| Volume 2 ||Issue 2 ||Sept 2017||ISSN (Online) 2456-0774
INTERNATIONAL JOURNAL OF ADVANCE SCIENTIFIC RESEARCH
AND ENGINEERING TRENDS

Amalgamation of Forecasting Environment Based upon Natural Disaster (Earth Quake) Analysis using DM Techniques

M. Karppakameenal[1], K. Sangeetha[2]
M. Phil Department of Computer Science, PRIST University, Thanjavur, India
Assistant Professor, Department of PG Computer Science, P. R. Engineering College, Vallam, Thanjavur

***Abstract*— Data mining is the computer supported process of determining and studying enormous sets of data and then take out the meaningful data. Data mining tools predicts performances and future trends, allowing businesses to make practical decisions. It can answer questions that conventionally were very time consuming to resolve. An event called forecast in a time series is more important for geophysics and economy difficulties. The time series data mining is an amalgamation (association) field of time series and data mining techniques. The ancient data are collected which has follow the time series organization, combine the data mining for pre-processing and finally apply the rules to predict the impact of earthquake which spoils the environment. Environment prediction has done by ancient earthquake time series to inspecting the method at first step ago. Enormous data sets are pre-processed using data mining techniques. Based on this process data prediction is possible. It focused on data mining techniques to analyze the environment through earthquake data. It describes data mining algorithms namely Random Forest, Classification Technique, Support Vector Machine (SVM) and Particle Swarm Optimization (PSO).**

Keywords: *Data Mining, Time Series Analysis, Earthquake Data, Environment Prediction.*

I INTRODUCTION

Earthquake - A natural disaster is the result of a natural hazard (e.g., Flood, tornado earthquake, heat wave, or landslide). Quakes, landslides, tidal wave and volcanoes are complex physical singularity that leads to, environmental or human losses. Prediction of such environmental disasters is the need of the day. Also, forecast of these disasters is a very complex process that be depending on many physical and conservational parameters. Many approaches exits based on analysis for investigating earthquake data for the prediction of environment status.

An observed time series are branded into three types as follows:

1. The Seasonal component
2. The Trend component and
3. The Irregular component

The seasonal data are methodical or regular movements of data. The Trend data mean by Long-term variations and the asymmetrical data mean by unsystematic or short-term variations. The major utility area of the time series model is statistical forecasting. The available calculation approaches are regression, time series and chaotic approaches. Each and every method has its own benefit and disadvantage. The historical sequences of data are to be used for forecasting purposes, because of this aim the time series models are used to predict the upcoming values. The prediction will show what will happen but won't why it transpires.

Time series values are transformed to segment space by using a nonlinear method and then apply the fuzzy logic to predict optimal value. The time series data are resultant from the time interval of any system. Traditional stationary time series representations are Autoregressive Integrated Moving Average (ARIMA) and Minimum Mean Square Error predicting methods. Data mining is used to excerpt useful and more applicable information from the huge database. The author Han describes pattern acknowledgement methods for prediction [2]. The time series data mining is used for calculation of environment [1]. The similarities of decorations are selected as set; these sets are specified in membership function. It is accurately forecast the prediction event. The concepts of time series data mining had been used for huddling and natural event forecast. [16]

Climate is the long-term effect of the sun's radioactivity on the rotating earth's varied surface and thermosphere. The Day to daytime dissimilarities in a given area establishes the weather, whereas climate is the long-term synthesis of such variations. Weather is measured by thermometers, rain gauges, weatherglasses, and other instruments, but the study of temperature relies on statistics. Nowadays, such statistics are fingered efficiently by computers. A simple, long-term immediate of weather changes, however, is still not a true photograph of climate. To obtain this requires the examination of daily, monthly, and yearly decorations [4].

II RELATED WORK

The authors G. V. Otari, Dr. R. V. Kulkarni [5] has deliberated in their paper "A Review of Application of Data Mining in Earthquake Prediction", it is a prediction of environmental disaster. The main data mining techniques used

for earthquake forecast are logistic models, neural networks, and decision trees, all of which provide primary solutions to the problems inherent in the prediction of shakings, tsunamis, and other micro seismic happenings. In their paper they also aim to inspire additional research on topics, and accomplish with several suggestions for further research.

The authors Gunther H. Weber, Marco Schneider, Daniel W. Wilson, Hans Hagen, Bernd Hamann and Bruce L. Kutter has discussed in their paper "Visualization of Experimental Earthquake Data", a conception tool that starts by reading the data which designates experiment set-up and displays this data along with icons for the devices used during data attainment. Different sensor types (measuring acceleration, displacement and strain) were indicated by different icons. One general research set-up was used in an arrangement of replicated earthquake events. Once a user has selected a particular event, restrained data can be viewed as a two-dimensional graph or plot by clicking the corresponding sensors. Multiple instruments can be animated to obtain a three-dimensional picturing of measured data.

Morales-Esteban [9] discusses the deportment of seismic temporal data and that was used to predict the medium and large shakings. The known patterns are determined when the intermediate and large earthquake happen. Chris gray [8] discuss the five geophysical precursors of earthquake. The predecessors are used to predict the earthquake. Here five phases are used to predict the earthquake and give progressive warning system of the earthquake. The major earthquake killed the several mountain people during the period from 2000 to 2010 in the area. Most of the scientists are currently looking to find a way to forecast the earthquake so that the life and possessions can be saved. The statistical analysis will be described by studying the case study of region. Earthquakes are considered by the depth and slip of the earthquake. The assessment of magnitude and years are possible using geometric data.

The major objective of the paper is to review the methods of forecasting earthquake and predicting the future using the time series model. The statistical analysis is used for long-term forecasts of the earthquake. The reviews of the geophysical precursors are dependence on the specific region. The precursors are not suitable for the other region. The impacts of the earthquake are identified through the ground water- level, ground quickening, time interval and P-wave changes.

III STATISTICAL ANALYSIS OF EARTHQUAKES

3.1 Review of Magnitude and number of earthquake occurrence using time series:

An earthquake is well-defined as an event in the seismic time series. An event is counting raise and fall of a gold price. One of the predicting approaches is statistical analysis. In the analysis of case study of the specific region, a persistent or cyclic pattern of the shakings is identified. The sample dataset of the region comprises depth and magnitude details. The prediction is conceivable by identifying the probability of incidence, time interval, magnitude, longitude, and autonomy range of the earthquakes.

3.2 The review of magnitude and depth of earthquakes:

The reviews of the depths are used to identify the responsibility of the earthquake. There are three types of faults are identified through the scale and depth of the earthquake. The faults are namely usual (or detachment) fault, reverse (or thrust) fault and strike-slip fault. Also the plate gestures are identified by historical data.

IV DATA AND METHODOLOGY

4.1 Data Collection

The data used for this work was composed from weather department through the Climatologically Agency. The case data covered for preceding five years, that is, January 2010 to January 2014. It has tropical wet and dry climate with dry conditions predominant for most of the year. It receives an annual rainwater of 1,205 mm (47.44 inches) from rainy season rains during June to September. The highest logged daily rainfall was 304 mm on 14 July 1994.The highest recorded infection in the city was 47.9 °C on May 22, 2013, while the lowest was 3.9 °C. The following procedures were adopted at this stage of the research: Data Cleaning, Data Selection, Data Alteration and Data Mining.

4.1.1 Data Cleaning

In this stage, a consistent format for the data faultless was developed which took care of missing data, finding duplicated data, and preparing out of bad data. Finally, the cleaned data were converted into a format suitable for data mining.

If the data set contains any location of focus, origin time and magnitude of an earthquake, depth and impact. These parameters could be analyzed that there is a perfect data about earthquake. If there is any irrelevant data that is to be avoided.

4.1.2 Data Selection

At this stage, data relevant to the examination was decided on and recovered from the dataset. While an analysis of the numeric values are obtainable. Due to the nature of the Form data where all the values are the same and the high percentage of missing values in the brightness data both were not used in the examination. There is indispensable difference between intensity scales and magnitude scales. Magnitude resembles to the energy released at the source of the earthquake.

4.1.3 Data Alteration

This is also known as data association. It is the stage in which the selected data is converted into forms appropriate for data mining. The data file was saved in Commas Separated Value (CSV) file format and the datasets were normalized to

reduce the effect of climbing on the data. All the data used for prediction of environment could be raw data format in excel. With these raw data one cannot get an effective solution in prediction. So that the raw data from earth quake details are converted into CSV excel sheet (formatted data).

4.1.4 Learning Algorithm

As already explained, different data mining procedures were used in the proposed system which their conditions are as follows. It contains of the optimal values of the momentum term and the knowledge rate, which are 0.077 and 1, respectively. The efficiency of Support Vector Machine (SVM) is monitored by parameter C and parameter γ. Kernel type restriction specifies the type of kernel purpose. Different types of kernels such as dot, radial, polynomial, Gaussian combination and multi-quadric can be designated.

Gaussian combination types with set of examples were unsuccessful in calculations and the conclusion time was more than the anticipated time. In the neural kernel, the process was not caused in a final answer after 25 minutes yet. Two kernel types of dot and polynomial had an output for a dangerous state. The rest of the cases did not have any forecast or record return for a hazardous state.

Polynomial kernels are very good for the problems which all training data have been standardized. The constant C defines the complexity of SVM. It is important to control the correct constant. If the constant is too large, it can lead to a lot of inappropriate connection points, while very small quantities may lead to spread the points more. Max iteration parameter is an optimization stricture which does the stopping process after a specified number of calculations.

So, we can see how small changes have a significant impact on results. Therefore, having a good understanding of used kernel parameters is essential. As well as having a good understanding of the different types of kernel, selecting the most appropriate type of kernel for sample set is also equally important. In this paper, a polynomial kernel function was used instead of some other kernel functions such as Gaussian kernel for SVM algorithm, because Gaussian kernel function had almost low accuracy. Optimized algorithms such as PSO can be used to accelerate the performance of SVM algorithm and enhance its accuracy. The PSO algorithm used to optimize SVM parameters such as C, penalty factor and g or γ, kernel function parameter. In Naive Bayes, the results obtained classes listed in the situation "hazardous" and "nonhazardous" have been reported. Decision tree consists of 95 nodes with 96 terminal leaves.[17]

V RESULT AND DISCUSSION

For these evaluation 2500 data features has been used. In all steps, the SVM-PSO algorithm was almost the fastest algorithm. In other words, dissimilar data which have been used in each step of cross authentication do not change the algorithm accurateness. In order to evaluate the impact of different topographies on the performance, we have divided features into groups. As can be seen in the table, each algorithm using each feature category has dissimilar accuracy. The SVM-PSO algorithms have the uppermost accuracy for the first category.

The accuracy of each step was computed and final accuracy was obtained using averaging of all accuracy. This causes all datasets have participated in training and testing phases. In addition to the accuracy, the computational time of each algorithm was measured. Although the SVM algorithm is the fastest one, it has the smallest accuracy. While SVM-PSO algorithm has the highest accuracy and its computational time is smaller than MLP and NN algorithms. In order to clearly show the accuracy results, we show the overall accuracy of all tested data mining algorithms.

These results show that some features are more significant than others in terms of impact on the accuracy. Comparison of results of this paper and other related work has been presented in Table 1.

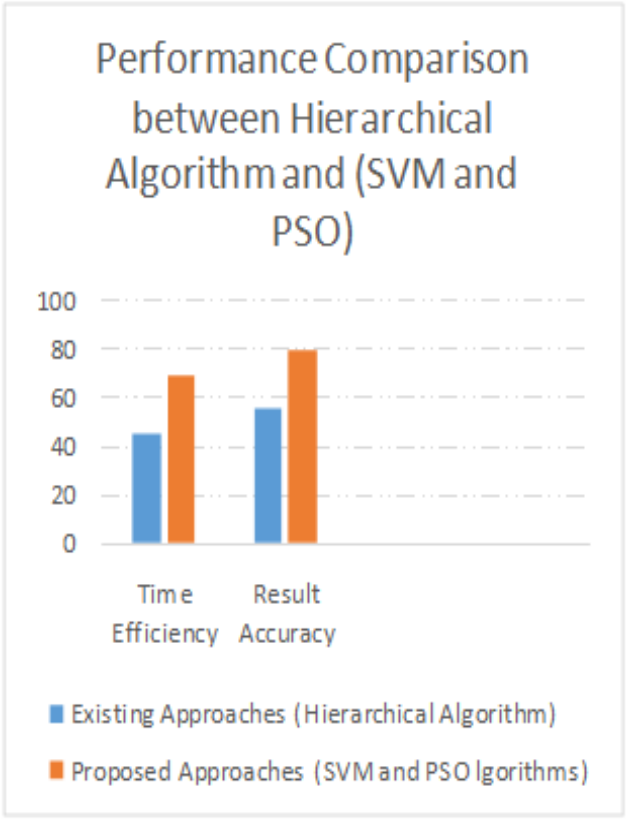

Figure 1 Algorithm Comparison for Existing and Proposed Algorithms with time and accuracy

The results show the ability of SVM-PSO algorithms to predict environment through earthquakes using seismic hazard data is higher than other procedures such as rule-based regression procedure and rule-based classification algorithm in [10]. As can be seen the accuracy of the SVM-PSO algorithm is advanced than [10].

Table 1 Performance of SVM-PSO algorithm in environment prediction through earth quake

Training and Testing Data	**Parameter**	**SVM-PSO**
	False Positive (FP)	0
(Training Data=517-2500)	Accuracy (%)	84.11
(Testing Data=1-516)	Computational Time(Second)	137
	FP	0
(Training Data=1-516, 1034-2584)	Accuracy (%)	93.05
(Testing Data)=517-1033	Computational Time(Second)	134
	FP	0
(Training Data=1-1033, 1551-2584)	Accuracy (%)	95.36
(Testing Data)=1551-2067	Computational Time(Second)	136
	FP	0
(Training Data)=1-1551,2500	Accuracy (%)	96.52
(Testing Data)=2068-2500	Computational Time(Second)	136

VI CONCLUSION

The results using real earth quake data demonstrations that support vector machine algorithm and particle swarm optimization (SVM-PSO) is the wildest and also has the highest accuracy. It provides better accuracy in the computation of environment prediction based upon earth quake data. The benefits of particle swarm optimization algorithm comprise memory utilization, sharing material, and high-speed conjunction. Support vector machine optimized using particle swarm optimization algorithm has the highest accuracy associated to other.

Future work in the area could focus on multiple research directions. The study will be improved with more complex models on crustal dynamic charting analysis that guarantees the possibility of a large earthquake being activated by another noticeable strain announcement event.

REFERENCE

[1]. I.Aydin, M. Karakose and E.Akin, The Prediction Algorithm Based on Fuzzy Logic Using Time Series Data Mining Method.

[2]. J. Han and M. Kamber, Data Mining: Concepts and Techniques, San Francisco: Academic Press, 800 p., 2005

[3]. R. J. Povinelli, "Time Series Data Mining: Identifying Temporal Patterns for Characterization and Prediction of Time Series Events", Ph.D. Dissertation, Marquette University, 180 p., 1999.

[4] Fairbridge R. W., 2007, "Climate" Microsoft® Student 2008 [DVD], Redmond, WA: Microsoft Corporation, 2007.

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

[6]. Dave A.Yuen1, Benjamin J Kadlec1, Evan F Bollig1, Witold Dzwinel2, Zachary A.Garbow1, Cesar da Silva3, "Clustering and Visualization of Earthquake data in a Grid Environment".

[7]. Kyle Springer, Atsushi Inoue, "Novel Visualization Scheme for Reasoning With data mining", the 28th North American Fuzzy Information ,Processing Society Annual Conference(NAFIPS2009) Cincinnati, Ohio, USA - June – 2009.

[8]. Chris gray, A Review of Two Methods of Predicting Earthquakes.

[9]. A. Morales- Esteban, F. Martinez-Alvarez, A. Troncoso, J.L.Justo, C. Rubio-Escudero, Pattern recognition to forecast seismic time series, Elsevier, Experts System with Application 37(2010) 8333 – 8342

[10]. Kabiesz, J., B. Sikora, M. Sikora, and L. Wrobell, 2013. Application of rule-based models for seismic hazard prediction in coal mines, Acta Montanistica Slovaca, 18 (4): 262-277.

[11]. Ahumada, A., A. Altunkaynak and A. Ayoub, 2015. Fuzzy logic-based attenuation relationships of strong motion earthquake records. Expert Systems with Applications, 42 (3): 1287-1297.

[12]. Siler, W. and J.J. Buckley, 2005. Fuzzy expert system and fuzzy reasoning, John Wiley & Sons, Inc.

[13]. Ikram, A. and U. Qamar, 2014. A rule-based expert system for earthquake prediction, J Intel Inf Syst, 43 (2): 205-230.

[14]. Ikram, A. and U. Qamar, 2015. Developing and expert system based on association rules and predicate logic for earthquake prediction, Knowledge-Based Systems, 75, 87-103.

[15]. Abrahamson, N. and W. Silva, 2008. Summary of the Abrahamson and Silva NGA ground-motion relations. Earthquake Spectra, 24 (1): 67-97.

[16] G.Preethi, B.Santhi Study on Techniques of Earthquake Prediction International Journal of Computer Applications (0975 – 8887) Volume 29– No.4, September 2011 page 55-59

[17]Asadollah Shahbahrami and Zinat Mehdidoust Jalali Evaluation of Different Data Mining Algorithms to Predict Earthquakes Using Seismic Hazard Data J. Appl. Environ. Biol. Sci., 7(2)142-150, 2017 © 2017, TextRoad Publication ISSN: 2090-4274 Journal of Applied Environmental and Biological Sciences