Earthquake Prediction and Seismic Analysis Using Hadoop

Faraz Ahmad(16BCE0920), Om Ashish Mishra(16BCE0789), Shivam Bhagwani (16BCE0717)

Abstract: Characteristic perils like earthquakes are for the most part the consequence of spreading seismic waves underneath the surface of the earth. Tremors are dangerous absolutely in light of the fact that they're erratic, striking without warning, triggering fires and tsunamis and leading to deaths of countless individuals. If researchers could caution people in weeks or months ahead of time about seismographic disturbances, clearing and different arrangements could bee made to spare incalculable lives. An early detection and prediction system using machine learning classification models can prove to be very useful for disaster management teams. The earthquake stations continuously collect data even when there is no event. From this data, we need to distinguish earthquake and non-earthquake. Machine learning techniques can be used to analyse continuous time series data to detect earthquakes effectively. The proposed work uses Pig-Hive optimisation in Hadoop for processing, mining and analysing earthquake data.

Keywords: earthquake, pig-hive, prediction, classification, machine learning

I. Literature Survey

Title of the paper with Year	Journal Name	Algorithm & Application used	Performance	Future Scope
Why is earthquake prediction research not progressing faster?	*Tectonophysics* 338.3-4 (2015): 217-223.	Analysing Crystal deformations with the help of GPS and InSAR.	Theoretical enumeration	The strong critical attack lead opened up new avenues and acted as the foundations for research on algorithmic processing.
BIRCH: an efficient data clustering method for very large databases.	*ACM Sigmod Record*. Vol. 25. No. 2. ACM, 1996.	Clustering Algorithm called BIRCH for large dataset.	Finds useful pattern in large datasets. Works really well with them.	If given proper attention the algorithm can perform even better in improved and resourceful environment.
Efficient big data processing in Hadoop MapReduce.	*VLDB Endowment* 5.12 (2012): 2014-2015.	MapReduce	This journal has highlighted the similarities and differences between Hadoop MapReduce and Parallel DBMS. MapReduce algorithm is a slow but mostly	MapReduce is being given various improvements daily and its variations are being released with many different names. It is expected to

			an accurate way of decomposing tasks and carrying out there execution parallelly.	bloom in future as the tasks are becoming more complex in the world surrounded by computers.
Mammoth: Gearing Hadoop towards Memory-Intensive MapReduce Applications	*IEEE Transactions on Parallel and Distributed Systems* 26.8 (2015): 2300-2315.	Mammoth	The algorithm used is called Mammoth which has been recently developed for processing highly related datasets which can be reduced into even finer clusters with greater efficiency. Results show that Mammoth decreases execution time by **40%** as compared to normal MapReduce algorithm.	Developments are being done. The researchers are trying to upgrade algorithm's functionality and roll out updates almost every 6 months giving some more abilities every time.

Artificial neural networks for earthquake prediction using time series magnitude data or Seismic Electric Signals.	*Expert Systems with Applications 38 (2011) 15032–15039*	Neural Networks	A very impressive approach and probably the most defining move towards earthquake prediction was the use of neural networks. This paper presents a new method for earthquake prediction. In the proposed method the variation of geomagnetic field declination, horizontal component and hourly relative humidity, temperature ground, rainy rate per day such as hum, rrr (the average of rainy hours' time per day), and tgtg (temperature ground) are used to predict magnitude of earthquake 2 days before the occurrence of earthquake occurrence by using a neural network. As a case study, earth geomagnetic field measured data is used. Earthquake prediction methods have vastly improved, however there is no method of maximised accuracy using big data intensive	As a general observation, we can argue that the neural network models used in this study are able to predict the magnitude and the time lag of a major seismic event relatively accurately. This fact confirms that once the neural networks are trained by using the appropriate data, they are able to generalise and predict unknown seismic events accurately. The accuracy rates presented in the current paper are all based on the out-of-sample performance for each model. In other words, the data used for testing the networks are different from data used for training. According to Tashman (2000), for a good evaluation of a forecasting study, the method should be assessed based on the out-of-sample performance. In this way the predictive capability of the model will match the

			approaches.	conditions of the real-world.

Earthquake prediction in California using regression algorithms and cloud-based big data infrastructure	*Computers and Geosciences, Elsevier* *ISSN : 0098-3004*	Linear regression and Random Forest	Combines several linear regression with ensemble learning in the context of big data. These include generalized linear models, gradient boosting machine, deep learning, random forests and stacking ensembles. A total of 1GB of data was analysed divided into 27 datasets and processed by means of cloud based infrastructure. The stacking-based ensemble learning had been applied, reporting relative errors verging on **10% and absolute errors verging on 0.5**.The methods based on trees yielded better results and lower regression errors.	The algorithms lead to accurate earthquake prediction on a week based metric for the hottest zones of California. The results yielded were promising with optimisation to 5% relative errors in predicted magnitudes. Laid the benchmark for a cloud based big data architecture for all earthquake prediction pipeline.

PERFORMANCE INVESTIGATION OF NEURO-FUZZY SYSTEM FOR EARTHQUAKE PREDICTION	*ASIAN JOURNAL OF CIVIL ENGINEERING (BHRC) VOL. 17, NO. 2 (2016) PAGES 213-223*	Adaptive Neuro-Fuzzy Inference System (ANFIS)	Hybrid neuro-fuzzy systems for estimating a function, a Sugeno-type fuzzy system a special five-layer network. **This algorithm increases efficiency and optimisation of predicting latitudes and longitudes of earthquake, depth and intensity**. It can also predict the seismic time between tremors thus acting as a benchmark for preparation and responsiveness of earthquakes. Lack of uniformity in seismic catalogue made the authors use available regression equations. These equations which are used to determine moment magnitude of earthquake records followed by calculating seismic moment lead to error.	In the non-logarithmic mode, amount of error is more than logarithmic mode, but the correlation factor for the predicted amount which is calculate from the obtained results by ANFIS is 98% that is not negligible. The seismic moment is one of the important seismic indicators that use for evaluating an earthquake. Prediction of this indicator can help to predictors to find other qualities of the next earthquake such as magnitude or released energy.

Earthquake Prediction based on Spatio-Temporal Data Mining: An LSTM Network Approach	*Transactions on Emerging Topics in Computing (IEEE)*	Long Short Term Memory Neural Network Units	The accuracy on the testing data is **74.81%, with the true positive accuracy of 68.56% and true negative accuracy of 81.31%.** The accuracy was improved by using more hidden layers. However a generic overhead was observed compared to other algorithms in obtaining and processing the input data.	The algorithm can be used to make predictions and analyse earthquake and other natural disturbances occurring. The algorithm can be used to make accurate predictions with different temporal and spatial prediction granularities

Literature Survey Report:

[1] and [5] analysed by Om Ashish Mishra 16BCE0789

[2],[3],[4] analysed by Shivam Bhagwani 16BCE0717

[6], [7], [8] analysed by Faraz Ahmad 16BCE0920

II. Conclusion

The main motive of doing this project is to experience the excellence of Hadoop which defiantly is a part of the parallel and distributing computing. The aim to obtain better models of prediction than the existing system can be achieved so that the whole idea can be used by the community.

From the data obtained and visualisations done till now it was seen that most of the earthquakes occur in the pacific ring also known as the pacific ring of fire. The different metrics have been analysed and a thorough study conducted also show signs of nuclear testing causing disturbances in the tectonics plates which can have severe repercussions. Thus, proper methods are to be researched and algorithms are to be implemented for performing predictions and obtaining a model that is efficient across all platforms and metrics. The **Pig Hive** optimization technique is used. Pig is used here rather than other Hadoop tools like Hive or Zookeeper as it is fast and it processes the data faster. Just like the literal meaning of the pig, an animal who eats garbage, similarly here also the pig automatically processes the data(arranges) irrespective of the arranged or not. The KNN algorithm is inefficient for large scale processing thus the proposed work aims at studying other areas such as evolutionary computing, high performance computing, artificial intelligence, swarm particle optimization etc to find a highly efficient algorithm that can be used.

References

[1] Wyss, Max. "Why is earthquake prediction research not progressing faster?." Tectonophysics 338.3-4(2001):217-223.

[2] Zhang, Tian, Raghu Ramakrishnan, and Miron Livny. "BIRCH: an efficient data clustering method for very large databases." ACM Sigmod Record. Vol. 25. No. 2. ACM, 1996.

[3] Dittrich, Jens, and Jorge-Arnulfo Quiané-Ruiz. "Efficient big data processing in Hadoop MapReduce." Proceedings of the VLDB Endowment 5.12 (2012): 2014-2015.

[4] Shi, Xuanhua, et al. "Mammoth: Gearing hadoop towards memory-intensive mapreduce applications." IEEE Transactions on Parallel and Distributed Systems 26.8 (2015): 2300-2315.

[5] Moustra, Maria, Marios Avraamides, and Chris Christodoulou. "Artificial neural networks for earthquake prediction using time series magnitude data or Seismic Electric Signals." Expert systems with applications 38.12 (2011): 15032-15039.

[6] Asencio–Cortés, G., et al. "Earthquake prediction in California using regression algorithms and cloud-based big data infrastructure." Computers & Geosciences 115 (2018): 198-210.

[7] Mirrashid, M., et al. "Performance investigation of neuro-fuzzy system for earthquake prediction." (2016): 213-223.

[8] Wang, Qianlong, et al. "Earthquake prediction based on spatio-temporal data mining: an LSTM network approach." IEEE Transactions on Emerging Topics in Computing (2017).