# Earthquake Prediction Based on Spatio-Temporal Data Mining: An LSTM Network Approach

**Sooryaraj. S. T[1], Harini. B[2], Mrs.V. Stella Mary[3]**
[1, 2]Dept of Information and Technology
[3]AP, Dept of IT
[1, 2, 3] Sri Muthukumaran Institute of Technology

***Abstract-*** *This project investigates the importance of seismological research around the globe is very clear. Therefore, new tools and algorithms are needed in order to predict magnitude, time, and geographic location, as well as to discover relationships that allow us to better understand this phenomenon and thus be able to save countless human lives. However, given the highly random nature of earthquakes and the complexity in obtaining an efficient mathematical model, efforts until now have been insufficient and new methods that can contribute to solving this challenge are needed. In this work, a novel earthquake magnitude prediction method is proposed, which is based on the composition of a known system whose behavior is governed according to the measurements of more than two decades of seismic events and is modeled as a time series using machine learning, specifically a network architecture based on LSTM (long short-term memory)cells.*

***Keywords***- Earthqauake, LSTM, Disaster, Prediction.

## I. INTRODUCTION

Some works employ mathematical or statistical tools to make earthquake prediction. Predicts earthquake epicenters according to spatial connections theory, i.e., earthquakes occurring within a fault zone are related to one another. Particularly, predictions are made by taking advantages of Poisson range identifier function (PRI), Poisson distribution, etc. improve the Fibonacci, Dual and Lucas (FDL) method and propose an scheme to predict earthquakes by using a trigger planetary aspect date prior to a strong earthquake as a seed for the unfolding of FDL time spiral. However, these works are only tested with very limited amount of data and do not provide good results (the success rate is low).

## II. OBJECTIVE

In this work, we divided the whole observation period in non-overlapping windows of 1 h duration and considered only the event with the largest magnitude which occurred in each hourly window. As example, the first 100 h of our magnitude time series. Some windows have magnitude 0, which means that no earthquakes occurred. Actually, these values are not part of the behavior of the system and they make the training of the network very difficult.

## III. LITERATURE SURVEY

The existing literature related to the analysis of social media data for crisis response and disaster management is rapidly growing. This paper, however, focuses more narrowly on the state-of-the-art machine learning techniques that have been employed to detect and map disaster events and their geographical information. Hence, this section discusses related work in the extant literature focusing on this aspect. The role of social media data for information propagation in disasters has been recognized in devastating natural disasters such as the 2010 Haiti Earthquake and the 2012 Hurricane Sandy. The majority of studies focus on processing disaster-related social media data to gain situational awareness. One branch of existing studies focus on developing topic modeling and event detection approaches to extract frequently mentioned keywords corresponding to disaster events. For example, Mishler et al. modeled the temporal changes in topics in the Ukrainian Crisis using Structural Topic Modeling. Kireyev et al. conducted topic modeling on tweets during two earthquakes in American Samoa and Sumatra, emphasizing dynamic corpus refinement. In addition, researchers also designed a number of platforms to perform automatic classification of crisis-related microblog communications. AIDR (Artificial Intelligence for Disaster Response) is a platform that aims to classify social media posts into humanitarian categories in real-time. The majority of the existing machine learning pipelines focus on grabbing and classifying real-time posts to improve the efficiency of situational awareness in disasters. However, these approaches have an important limitation related to providing detailed information about specific disaster events (such as which road is closed, when it will open, which building is damaged, and how long service was interrupted). In addition, existing classifiers and event detection pipelines

## IV. METHODOLOGY

**LSTM (Long Short Term Memory)**

Automatic text classification or document classification can be done in many different ways in machine learning as we have seen before.This article aims to provide an example of how a Recurrent Neural Network (RNN) using the Long Short Term Memory (LSTM) architecture can be implemented using Keras. We will use the same data source as we did Multi- Class Text Classification with Scikit-Lean, the Consumer Complaints data set that originated from data.gov.We will use a smaller data set, you can also find the data on Kaggle. In the task, given a consumer complaint narrative, the model attempts to predict which product the complaint is about. This is a multi-class text classification problem. Let's roll!Reducing capital expenditure and leveraging existing assets

**Label Consolidation**

After first glance of the labels, we realized that there are things we can do to make our lives easier.

Consolidate "Credit reporting" into "Credit reporting, credit repair services, or other personal consumer reports".

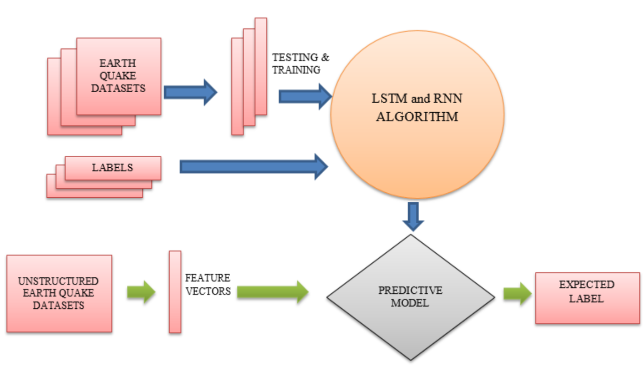
Consolidate "Credit card" into "Credit card or prepaid card".

Consolidate "Payday loan" into "Payday loan, title loan, or personal loan".

Consolidate "Virtual currency" into "Money transfer, virtual currency, or money service".

"Other financial service" has very few number of complaints and it does not mean anything, so, I decide to remove it.

**BLOCK DIAGRAM**




## V. MODULES

Data collection.:Data collection is defined as the procedure of collecting, measuring and analyzing accurate insights for research using standard validated techniques. A researcher can evaluate their hypothesis on the basis of collected data. In most cases, data collection is the primary and most important step for research, irrespective of the field of research. The approach of data collection is different for different fields of study, depending on the required information. The most critical objective of data collection is ensuring that information-rich and reliable data is collected for statistical analysis so that data-driven decisions can be made for research. Earthquakes have over the course of time, evolved into an effective alternative to expensive mail or telephone surveys. There are a few conditions which need to be met to online surveys however that you must be aware of. If you are trying to survey a sample which represents the target population, please keep in mind that not everyone is online.

**Data cleaning**: Data cleaning is the process of preparing data for analysis by removing or modifying data that is incorrect, incomplete, irrelevant, duplicated, or improperly formatted. This data is usually not necessary or helpful when it comes to analyzing data because it may hinder the process or provide inaccurate results. There are several methods for cleaning data depending on how it is stored along with the answers being sought. Data cleaning is not simply about erasing information to make space for new data, but rather finding a way to maximize a data set's accuracy without necessarily deleting information. Most importantly, the goal of data cleaning is to create data sets that are standardized and uniform to allow business intelligence and data analytics tools to easily access and find the right data for each query.

**Exploratory data analysis (EDA):** Finally, I am ready to start building my bridge between my experience with seismic earthquakes data set and my new skills in EDA. For making it I will use some clues as the building blocks. The first block is related to what the seismic earthquakes data set tell us about the subsurface. When a seismic earthquakes is observed it brings to the observer a perception of the subsurface as a collage of diverse visual patterns, called seismic earthquakes reflections in the Population, arranged in such way that after an appropriate interpretation of it a meaningful geological model appears as formed by layers defining structures. How do we get geological sense from seismic earthquakes data set. Because the seismic earthquakes data is usually recorded in time, what we are seen on a seismic earthquake is just a record of the arriving times to the ground of all the reflected seismic earthquakes waves travelling across the subsurface. As explained previously, each seismic reflected amplitude is a response to a particular elastic contrast in depth. Then, each recorded arriving time is only a transformed representation of the seismic response in depth. For this reason, the coherent distribution of all those reflected amplitudes, arranged as a function of their associated travel times, is what give us the geological perception of it. But this fact also introduces, as we will see later, a bias in the DL training stage which must be removed or mitigated before EDA.

**IMPLEMENTATION OF LSTM**: The idea of this post is to teach you how to build your first Recurrent Neural Network (RNN) for series prediction. In particular, we are going to use the Long Short Term Memory (LSTM) RNN, which has gained a lot of attention in the last years. LSTM solve the problem of vanishing / exploding gradient in typical RNN. Basically, LSTM have an internal state which is able to remember data over big periods of time, allowing it to outperform typical RNN. There is a lot of material on the web regarding the theory of this network, but there are a lot of misleading articles regarding how to apply this algorithm. In this article we will get straight to the point, building an LSTM network, training it and showing how it is able to make predictions based on the historic data it has seen.

Time-series modeling: Time series data often arise when monitoring industrial processes or tracking corporate business metrics. The essential difference between modeling data via time series methods or using the process monitoring methods discussed earlier in this chapter is the following:

Time series analysis accounts for the fact that data points taken over time may have an internal structure (such as autocorrelation, trend or seasonal variation) that should be accounted for.

This section will give a brief overview of some of the more widely used techniques in the rich and rapidly growing field of time series modeling and analysis.

## VI. CONCLUSION

We proposed a new earthquake prediction system from the Spatio-temporal perspective which overcomes the limitations in the existing system such as: The problem of challenge that large earthquakes cannot be reliably predicted for specific regions over time scales that span less than decades have been resolved. Still successfully applied to many problems but This system can capture long term dependencies. It can explode gradients problem in which gradients are either squashed to zero or increase without bound during back propagation through a large number of time steps Primarily to overcome the problem of vanishing gradients. Also we have following merits in our system which enhances the efficiency and accuracy of prediction. Prediction of earthquakes well in advance can alert the public in saving lives as well as resources at the right time will be our merit to address the people nearby. In this research, we studied the impact of earthquakes which have happened in the past through the historical data therefore we used all of the data to the right resource and predicted accurately. The traditional processes like preprocessing of data, training the neural network, weight adjustments, testing on test data are done and result was approximately to the exact prediction by other systems.

Specifically, we have designed 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.

## REFERENCES


[1] Y. Kryvasheyeu et al., “Rapid assessment of disaster damage using social media activity,” Sci. Adv., vol. 2, no. 3, 2016.

[2] H. Seppänen and K. Virrantaus, “Shared situational awareness and information quality in disaster management,” Saf. Sci., vol. 77, pp. 112–122, 2015.

[3] L. Palen and K. M. Anderson, “Crisis informatics-new data for extraordinary times: Focus on behaviors, not on fetishizing social media tools,” Science, vol. 353, no. 6296. American Association for the Advancement of Science, pp. 224–225, 15- Jul-2016.

[4] S. A. Shah, D. Z. Seker, M. M. Rathore, S. Hameed, S. Ben Yahia, and D. Draheim, “Towards Disaster Resilient Smart Cities: Can Internet of Things and Big Data Analytics Be the Game Changers?,” IEEE Access, vol. 7, pp. 91885–91903, 2019.

[5] C. Fan, A. Mostafavi, A. Gupta, and C. Zhang, “A System Analytics Framework for Detecting Infrastructure-Related Topics in Disasters Using Social Sensing,” in Advanced Computing Strategies for Engineering, 2018, pp. 74–91.