www.ijiarec.com

ISSN:2348-2079

Volume-6 Issue-2

International Journal of Intellectual Advancements and Research in Engineering Computations

Earthquake Prediction System by LSTM

[1]PARAMESWARAN.S, [2]NANDHA GOPAL.A, [3]REVATHI.D, [4]SENTHAMARAI.M

[1-3] UG Scholars, Department of computer science and engineering, Nandha Engineering college ,Erode

[4]Assistant Professor, Department of computer science and engineering, Nandha Engineering college ,Erode

***Abstract*—Earthquake prediction is a branch of the science of seismology concerned with the specification of the time, location, and magnitude of future earthquakes within stated limits and particularly the determination of parameters for the next strong earthquake to occur in a region.Prediction can be further distinguished from earthquake warning systems, which upon detection of an earthquake, provide a real-time warning of seconds to neighboring regions that might be affected.Early Response Warning System is currently responsible for alerting airports,trains, fire stations, etc. coal mine workers are exposed to alife-threatening danger in a form of seismic events. The existing earthquake prediction algorithm based on mathematical analysis, machine learningalgorithms like decision trees and support vector machines, and precursors signal. The existing system do not have very accuracy due to the seemingly dynamic and unpredictable nature of earthquakes. In this proposed systemdeep learning technique called long short-term memory with gradient descent optimization algorithms(LSTM-AdaGrad) networks predicting future earthquakes using data of past earthquakes using Long Short-Term Memory neural networks algorithm. The experimental result shows the better performance compare with existing algorithm like spatial temporal analysis.**

***Index Terms*— Earthquake prediction, long short-term memory, Recurrent Neural Network**

I. INTRODUCTION

Predictions are deemed significant if they can be shown to be successful beyond random chance. Therefore, methods of statistical hypothesis testing are used to determine

happen anyway (the null hypothesis). The predictions are then evaluated by testing whether they correlate with actual earthquakes better than the null hypothesis.Generally, there are two different aspects of earthquake prediction:long-term forecasting and short-term forecasting. Whereas short-term

forecasting is supposed to predict the exact time, location, and magnitude of an earthquake event, we focushere on long-term forecasting, which aims at predicting a large earthquake a year or even severalyears in advance. The most existing methods of short-term earthquake predictionare looking for recurrence patterns in the sequence of tectonic events occurring in thesame region. However, large earthquakes often fail to occur around their expected recurrencetimes.

Earthquake - A natural disaster is the effect of a natural hazard (e.g., Flood, tornado earthquake, heat wave, orlandslide). Earthquakes, landslides, tsunamis and volcanoes are complex physical phenomenon that leads to,environmental or human losses. Prediction of such geological disasters is the need of the day. Also, prediction of thesedisasters is a very complex process that depends on many physical and environmental parameters. Many approachesexit based on analysis for analyzing earthquake data. Data mining techniques can also be used for prediction of thesenatural hazards. The processing of post-earthquake raw data is generally done using Excel sheet which is very much timeconsuming. This project overcomes this drawback; in which data will be uploaded just by clicking on a button. Itinvolves application of data mining technique- visualization, which may serve as time effective solution for visualizingearthquake data. There are two waves that cause an earthquake that are P-wave and S-wave the P-waves (Primary orPressure wave) is a pulse of energy that travels through the earth and through the liquid. It forces the ground to moveback and forth as it is compressed and expanded. The S-wave (Secondary or Sheer wave) follows more slowly withswaying, rolling motion that causes the sudden moment on the ground back and forth perpendicular to the direction ofthe wave.

Due to the uncertainty of the occurrence of theearthquake, finding meaningful relations between theevents is important to save the human life. Nowadays,using of the computer computation and intelligent systemmethods has already been increased. Using these methodsto extract the relationships between different earthquakesat different times and locations can help us to know thisphenomenon more to make a prediction of earthquake'shappening. The progresses in seismology have provided uswithin valuable knowledge of earthquakes. The modern equipment and the improvements inthe earthquake engineering have helped us to gain thevaluable information. On the other hand, the artificialintelligence systems require large suitable amount of datafor designing such systems. Pattern recognitionmethods that are used to analyze the earthquake data areconsisting of algorithms and tools based on statisticalinformation or a pre-knowledge-based data.

The task is an instance of a classification problem withunbalanced data provided in a form of multivariate, nonstationarytime series. We present a solution based on RecurrentNeural Network with Long Short-Term Memory cells.

II. Related Work

G. Asencio-Cortes et.,al.,methodology to systematically identify those values for certain parameters, somehow hidden in a set of seismicity indicators, that generate better results in terms of average accuracy when predicting earthquakes. In this sense, a set of parameters that may deeply influence the accuracy for predictions has been first identified. Later, a sensitivity analysis over such parameters has been conducted in order to determine how a wrong setup may lead to the occurrence of a major loss of accuracy in predictions.

F. Mart´ıneza et., al., In this work, the use of principal component analysis to reduce data dimensionality and generate new datasets is proposed. In particular, this step is inserted in a successfully already used methodology to predict earthquakes.The cities mostly threatened by large earthquakes occurrence in Chile, are studied. Several well-known classifiers combined with principal component analysis have been used. Noticeable improvement in the results is reported.

A. Boucouvalas et., al., present an improvement modification to the FDL method, the MFDLmethod, which performs better than the FDL. We use the FDL numbers to develop possibleearthquakes dates but with the important difference that the starting seed date is a triggerplanetary aspect prior to the earthquake. Typical planetary aspects are Moon conjunct Sun, Moonopposite Sun, Moon conjunct or opposite North or South Modes. In order to test improvement ofthe method we used all +8R earthquakes recorded since 1900, (86 earthquakes from USGS data).

J. Fan et., al., In order to solve the problem, a new predicting earthquakes method based on extract the texture and emergence frequency of the earthquake cloud is proposed in this paper. First, strengthen the infrared cloud images. Second, extract the texture feature vector of each pixel. Then, classified those pixels and converted to several small suspected areas. Finally, tracking the suspected area and estimate the possible location. The inversion experiment of earthquake show that this approach can forecast the seismic center feasible and accurately.

E. Florido et., al., he prediction of earthquakes is a task of utmost difficulty that has been widely addressed by using many different strategies, with no particular good results thus far. Seismic time series of the four most active Chilean zones, the country with largest seismic activity, are analyzed in this study in order to discover precursory patterns for large earthquakes. First, raw data are transformed by removing aftershocks and foreshocks, since the goal is to only predict main shocks. New attributes, based on the well-known b-value, are also generated. Later, these data are labeled, and consequently discretized, by the application of a clustering algorithm, following the suggestions found in recent literature. Earthquakes with magnitude larger than 4.4 are identified in the time series. Finally, the sequence of labels acting as precursory patterns for such earthquakes are searched for within the datasets.

M. HayakawaThis paper deals with the review on this short-term EQ prediction, including the impossibility myth of EQs prediction by seismometers, the reason why we are interested in electromagnetics, the history of seismo-electromagnetics, the ionospheric perturbation as the most promising candidate of EQ prediction, then the future of EQ predictology.

M. Hayakawa et., al., so we have tried to find whether there existed any precursors to this EQ, especially abnormal animal behavior (milk yield of cows), observed at Kagawa, Shikoku, near the EQ epicenter. The milk yield of cows has been continuously monitored at Kagawa, and it is found that the milk yield exhibited an abnormal depletion about 10 days before the EQ. This behavior has been extensively compared with the former electromagnetic precursors (ULF radiation, ionospheric perturbation). This leads to the discussion on the sensory mechanism of unusual behavior of mild yield of cows, and it may be suggested that ULF radiation among different electromagnetic precursors is a mostly likely driver, at least, for this EQ.

M. Jiang Low power consumption long time offset magnetic field detector (earthquake prediction). The design ofthe hardware circuit of the magnetic field detector seismic geomagnetic acquisition and preprocessing module mainlyincludes. Electronic compass, compass. monitoring device while the magnetic azimuth for monitoring andanalyzing the object, GSM, but it can also be applied to other seismic precursor information analysis, such asearthquake precursory infrasound abnormality, only need infrasound abnormality intelligent sensor replacegeomagnetic anomaly intelligent sensor and modify the re levant parameters can be.

S. KannanAn Innovative Mathematical Model analysis was carried out based on twenty years of earthquake data from California, Central USA, Northeast USA, Hawaii, Turkey, and Japan fault zones using Latitude, Longitude and Magnitude as variables. Using Poisson's distribution and spatial connection model, an identifiable pattern was found within the random occurrences of the earthquakes around each fault zone. This research provides an effective contribution to seismology by improving probability of successful prediction.

M. LastThis paper explores several data mining and time series analysis methods for predicting the magnitude of the largest seismic event in the next year based on the previously recorded seismic events in the same region. The methods are evaluated on a catalog of 9,042 earthquake events, which took place between 01/01/1983 and 31/12/2010 in the area of Israel and its neighboring countries. The data was obtained from the Geophysical Institute of Israel. Each earthquake record in the catalog is associated with one of 33 seismic regions. The data was cleaned by removing foreshocks and aftershocks. In our study, we have focused on ten most active regions, which account for more than 80% of the total number of earthquakes in the area. The goal is to predict whether the maximum earthquake magnitude in the following year will exceed the median of maximum yearly magnitudes in the same region. Since the analyzed catalog includes only 28 years of complete data, the last five annual records of each region (referring to the years 2006–2010) are kept for testing while using the previous annual records for training. The predictive features are based on the Gutenberg-Richter Ratio as well as on some new seismic indicators based on the moving averages of the number of earthquakes in each area. The new predictive features prove to be much more useful than the indicators traditionally used in the earthquake prediction literature. The most accurate result (AUC = 0.698) is reached by the Multi-Objective Info-Fuzzy Network (M-IFN) algorithm, which takes into account the association between two target variables: the number of earthquakes and the maximum earthquake magnitude during the same year.

III. PROPOSED APPROACH

The propose earthquake prediction scheme by adjusting a long short-termmemory (LSTM) network, which is an advanced RNN(gradient descent optimization algorithms) and hasstrong nonlinear learning capability even on the data containinglong-term interval correlations. The consider as a whole the earthquakes in an area of interestto be an input element to the LSTM-RNN network,which is different from common deep learning approaches thatonly consider the data in one particular location as an input.LSTM-RNN network with two-dimensional input that canlearn the correlations among earthquakes in different locations andat different time and exploit it to make predictions.

A. Recurrent Neural Networks

Recurrent Neural Network (RNN) is a type of artificialneural network in which dependencies between nodes forma directed cycle. This allows the network to preserve a statebetween subsequent time steps. We focus on a simple RNNwith a single, self-connected hidden layer.

RNNs process all elements from a sequence one-by-one,and the output at every time step depends on all previousinputs. This is a fundamental difference from feedforward networks,where the network’s output depends only on the currentelement. It has an important theoretical implication: RNNsare capable of approximating arbitrary well any measurablesequence-to-sequence mapping.Since RNNs contain loops, the standard backpropagationalgorithm does not work. Instead, a backpropagation throughtime algorithm is used. The idea behind this method is tounroll the network over N time steps and copy the parametersN times. The RNN parameters are shared across all time steps,which makes them trainable and allows generalization.Since the number of unrolled steps can be arbitrary, RNNsare particularly suited for modeling sequential data, where thelength of the input is not fixed or can be very long. Recurrentnets have shown impressive results in many NLP tasks.

B. Long Short-Term Memory

One important problem with training RNNs is the vanishinggradient, which can occur when values smaller than 1.0are multiplied at each time step during the backpropagationthrough time. For some activation functions, the maximalvalue of the derivative is small. For example, the derivative ofcommonly used sigmoid function is never bigger than 0.25. Asa result, after N time steps the gradient is multiplied by a valueless than or equal to 0.25N, which quickly becomes very smallas N increases. While using some activation functions can reduce the likelihood of vanishing gradients,there is a special architecture designed to address this problem:Long Short-Term Memory (LSTM).The LSTM is better at storing and accessing informationthan standard RNN. The LSTM block consists of a self-connectedmemory cell and 3 gates named: input, output andforget. The gates control the access to the cell and can beinterpreted as "read", "write" and "reset" operations in thestandard computer’s memory. The network learns to controlthe gates and decides to update and/or use the value at anygiven time step. Since all the components are built from differentiable functions, the gradients can be computed forthe whole system and it is possible to train it end-to-endusing backpropagation. There are several variants of LSTMthat slightly differ in connectivity structure and activationfunctions. Below we describe the definitions of the input,output and forget gates that we used.

IV. EARTHQUAKE PREDICTION BASED ON LSTM

The task is an instance of a classification problem withunbalanced data provided in a form of multivariate, nonstationarytime series. We present a solution based on Recurrent Neural Network with Long Short-Term Memory cells.1) earth is connected, and hence the seismicactivities in one location will naturally lead to seismic activities inother locations, and 2) the seismic activities tend to have certainpatterns in the time domain.

A. Preprocessing

This is a standard Machine Learningprocedure, and as such it should be applicable to almostany problem. The normalization makes easier bothoptimization of the loss function and the regularization, because all feature values are at the same scale.

The Min Max Scaler is the probably the most famous scaling algorithm, and follows the following formula for each feature:

$$\frac{x_i - \min(x)}{\max(x) - \min(x)} \quad (1)$$

This estimator scales and translates each feature individually such that it is in the given range on the training set, i.e. between zero and one.

It essentially shrinks the range such that the range is now between 0 and 1 (or -1 to 1 if there are negative values). The normalizer scales each value by dividing each value by its magnitude in n-dimensional space for n number of features. Rescale each feature individually to a common range [min, max] linearly using column summary statistics, which is also known as min-max normalization or Rescaling. The rescaled value for feature E is calculated as,

B. Recurrent Neural Networks (AdaGrad)

Recurrent neural networks are particularly suited to model time-series data such as temporal magnitude recordings because they incorporate a time delay in their operations

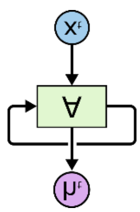

Fig. 1. Recurrent Neural Networks have loops

through a feedback connection between the output layer and the hidden layer(s). In a recurrent neural network, during every iteration the network output is passed through a recurrent layer and the output of the recurrent layer is added to the output of the hidden layer and the sum is used as the argument of the transfer function to obtain the network output in the succeeding iteration. Network output is obtained as

In the above diagram, a chunk of neural network, A, looks at some input xt and outputs a value ht A loop allows

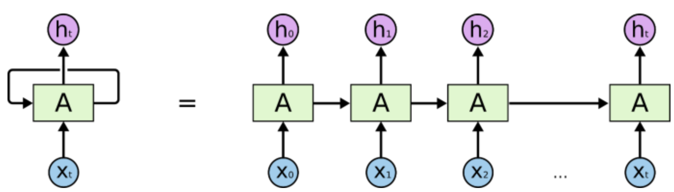

Fig. 2. An unrolled recurrent neural network

information to be passed from one step of the network to the next. These loops make recurrent neural networks seem kind of mysterious. However, if you think a bit more, it turns out that they aren't all that different than a normal neural network. A recurrent neural network can be thought of as multiple copies of the same network, each passing a message to a successor. Consider what happens if we unroll the loop.

This chain-like nature reveals that recurrent neural networks are intimately related to sequences and lists. They're the natural architecture of neural network to use for such data.The recurrent neural network model for predicting the occurrence of an earthquake of threshold magnitude or greater during the following time period consists of an input layer with eight nodes representing the eight aforementioned seismicity indicators. The number of hidden layers and the number of nodes in the hidden and recurrent layers are determined by numerical. The neural network output is either 1 or 0 indicating the occurrence or nonoccurrence of an earthquake of a certain threshold magnitude (different from the threshold magnitude used to define significant seismic events) or greater, respectively. The magnitude of the largest earthquake in the following time-period is determined by gradually increasing the threshold magnitude in increments of 0.5 Richter until the network output changes from 0 to 1 (In this article, an earthquake of a given magnitude m means all seismic events in the magnitude range $m \leq M < (m + 0.5)$).

Training the recurrent neural network is performed by comparing the network output to the actual occurrence or nonoccurrence of an earthquake of threshold magnitude or greater during the following time-period.

C. LSTM

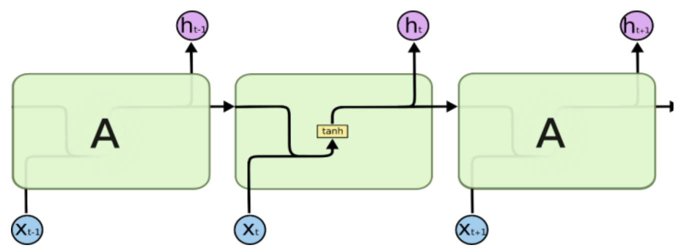

Fig. 3. The repeating module in a standard RNN contains a

LSTMs are explicitly designed to avoid the long-term dependency problem. Remembering information for long periods of time is practically their default behavior, not something they struggle to learn. All recurrent neural networks have the form of a chain of repeating modules of neural

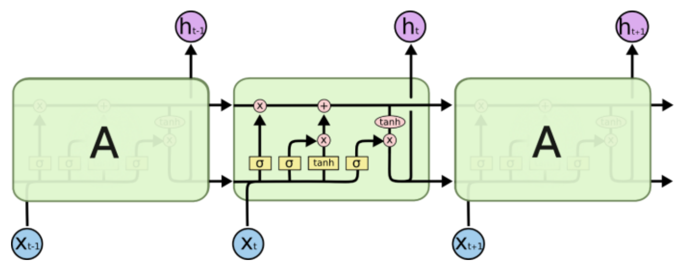

Fig. 4. The repeating module in an LSTM contains four interacting layers.

network. In standard RNNs, this repeating module will have a very simple structure, such as a single tanh layer.

LSTMs also have this chain like structure, but the repeating module has a different structure. Instead of having a single neural network layer, there are four, interacting in a very special way.

- Loss Function

To obtain the final output of the system, we choose soft max as the activation function and apply it to the output of the dense network. Particularly, the activation function maps the output vector into a vector of elements between 0 and 1, each of which represents earthquake probability in a sub-region and the sum of which equals to 1. The softmax function can be calculated as:

$$y_t^m = \frac{e^{z^m}}{\sum_{i=1}^{M} e^{z^i}} form = 1,,,M \quad (2)$$

Here, we use z to represent the output of the dense network h_t^D for simplicity.z^m and y_t^m represents the y_t^melement in vector z, and that in the output y_t, respectively. Note that the result is a vector of probabilities between 0 and 1 but not binary results that we need for crime prediction yet. To map the probabilities into 0s or 1s, we obtain an optimal probability threshold in the training process that minimizes the sum of the absolute value of the differences between the predicted label values and the real label values, which are either 0s or 1s.

D. Adaptive Gradient Algorithm

The consider the convex optimization setting. The algorithm iteratively makes a prediction earthquake xt∈ X, where X ⊆ Rd is a closed convex set, and then receives a convex loss function ft. Define the regret with respect to the (optimal) predictor x∗∈ X as

$$R(T) = \sum_{t=1}^{T} [f_t(x_t) - f_t(x^*)] \quad (3)$$

To achieve low regret, standard sub gradient algorithms move the predictor xt in the opposite direction of the sub gradient $g_t \in \partial f_t(x_t)$of the loss via the projected gradient update

$$\mathrm{x}_{\mathrm{t+1}} = \prod\nolimits_{\mathrm{x}} (\mathrm{x_t} - \eta \mathrm{g_t}) \quad (4)$$

Makes a second-order correction to the predictor using the previous loss functions.

Denote the projection of a point y onto X by $\prod_x^A(y) = argmin_{x \in X} \|x - y\|$(where $\|x\|_A = \sqrt{\langle x, Ax \rangle}$). In this notation, our adaptation of gradient descent employs the update

$$x_{t+1} = \prod_{x}^{G_t^{1/2}} \left(x_t - \eta G_t^{-1/2} g_t\right) \quad (5)$$

where the matrix $G_t = \sum_{\tau=1}^{t} g_\tau g_\tau^T$is the outer product of all previous sub gradients. The above algorithm may be computationally impractical in high dimensions since it requires computation of the matrix square root of Gt, the outer product matrix. We therefore also analyze a version in which we use diag(G_t), the diagonal of the outer product matrix, instead of G_t:

$$\boldsymbol{x_{t+1}} = \prod_{x}^{\boldsymbol{diag(G_t)^{1/2}}} \left(\boldsymbol{x_t} - \eta diag(G_t)^{-\frac{1}{2}} gt\right) \quad (6)$$

This latter update rule can be computed in linear time. Moreover, as we discuss later, when the vectors $g_t$are sparse the update can often be performed in time proportional to the support of the gradient.

Let us compare the regret bounds attained by both variants of gradient descent. Let the diameter of X be bounded, so $sup_{x,y \in X} \|x - y\|_2 \le D_2$. Then analysis of online gradient descent with the optimal choice in hindsight for the step size η achieves regret.

$$\boldsymbol{R(T) \le \sqrt{2D_2} \sqrt{\sum_{t=1}^{T} \|g_t\|_2^2}} \quad \mathbf{(7)}$$

The important parts of the bound are the infimum under the root, which allows us to perform better than using the identity matrix, and the fact that the step size is easy to set a priori.

V. Experimental Results

The data that we use is gathered from the USGS (US Geological Survey) website. In particular, we use Conterminous U.S earthquake data from 2006 to 2016 with magnitudes greater than 2.5 in our simulations. We set one-time slot to one month. In each time slot, the input is the number of earthquakes that happened in this time slot in a certain sub-region. We have 120 data items when one-time slot is one month. As usual, we divide the data into two parts: training data and testing data. Particularly, the first 2/3 of data would be used for training and the rest would be used for testing.

TABLE I
Accuracy comparison LSTM and LSTM-AdaGrad

Techniques	Performance		
	Accuracy	**True Positive**	**True Negative**
LSTM	85.12	77.07	93.49
LSTM-AdaGrad	91.34	80.43	95.74

The accuracy of a test is its ability to differentiate the patient and healthy cases correctly. To estimate the accuracy of a test, we should calculate the proportion of true positive and true negative in all evaluated cases. Mathematically, this can be stated as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

VI. Conclusion

In this proposed system deep learning technique called long short-term memory with gradient descent optimization algorithms (LSTM- AdaGrad) networks predicting future earthquakes using data of past earthquakes using Long Short-Term Memory neural networks algorithm. The experimental result shows the better performance compare with existing algorithm like spatial temporal analysis.

REFERENCES

[1] G. Asencio-Corte´s, F. Mart´ınez-A´ lvarez, A. Morales-Esteban, and J. Reyes. A sensitivity study of seismicity indicators in supervised learning to improve earthquake prediction. Knowledge-Based Systems, 101:15–30, 2016.
[2] G. Asencio-Corte´s, F. Mart´ınez-A´ lvarez, A. Morales-Esteban, J. Reyes, and A. Troncoso. Improving earthquake prediction with principal component analysis: application to chile. In International Conference on Hybrid Artificial Intelligence Systems, pages 393–404. Springer, 2015.
[3] A. Boucouvalas, M. Gkasios, N. Tselikas, and G. Drakatos. Modifiedfibonacci- dual-lucas method for earthquake prediction. In Third International Conference on Remote Sensing and Geoinformation of the Environment, pages 95351A–95351A. International Society for Optics and Photonics, 2015.
[4] J. Fan, Z. Chen, L. Yan, J. Gong, and D. Wang. Research on earthquake prediction from infrared cloud images. In Ninth International Symposium on Multispectral Image Processing and Pattern Recognition (MIPPR2015), pages 98150E–98150E. International Society for Optics and Photonics, 2015.
[5] E. Florido, F. Mart´ınez-A´ lvarez, A. Morales-Esteban, J. Reyes, and J. Aznarte-Mellado. Detecting precursory patterns to enhance earthquake prediction in chile. Computers & Geosciences, 76:112–120, 2015.
[6] M. Hayakawa. Earthquake prediction with electromagnetic phenomena. In THE IRAGO CONFERENCE 2015: 360 Degree Outlook on Critical Scientific and Technological Challenges for a Sustainable Society, volume 1709, page 020002. AIP Publishing, 2016.
[7] M. Hayakawa, H. Yamauchi, N. Ohtani, M. Ohta, S. Tosa, T. Asano, A. Schekotov, J. Izutsu, S. M. Potirakis, and K. Eftaxias. On the precursory abnormal animal behavior and electromagnetic effects for the kobe earthquake (m˜ 6) on april 12, 2013. Open Journal of Earthquake Research, 5(03):165, 2016.
[8] M. Jiang. Easily magnetic anomalies earthquake prediction. In MATEC Web of Conferences, volume 63, page 01020. EDP Sciences, 2016.
[9] S. Kannan. Innovative mathematical model for earthquake prediction. Engineering Failure Analysis, 41:89–95, 2014.
[10] M. Last, N. Rabinowitz, and G. Leonard. Predicting the maximum earthquake magnitude from seismic data in israel and its neighboring countries. PloS one, 11(1):0146101, 2016.