Available online www.jsaer.com

***Journal of Scientific and Engineering Research*, 2025, 12(1):1-12**

Research Article

ISSN: 2394-2630
CODEN(USA): JSERBR

Constructing an Earthquake Prediction Model by using the Self-Organizing Map Neural Network and Support Vector Regression

Chih-Ming Hsu

Department of Business Administration, Minghsin University of Science and Technology, Hsinchu, Taiwan

Abstract Since ancient times, humans have continued to face various types of natural disasters. Among these disasters, the earthquakes are arguably the most terrifying since humans cannot fully grasp the timing and scale of their occurrence. Furthermore, the huge losses in life, property, safety, economy, society, culture and politics etc. are usually suffered around the earthquake-affected areas when a strong earthquake occurs. Especially, there are quite frequent earthquakes occurred in Taiwan locating in the Circum-Pacific seismic belt that is the largest and most active seismic zone around the world, thus leading the intersection and collision of the Eurasian and the Philippine Sea Plates, as well as the subduction of the plates. Therefore, Taiwan's government has invested a lot of resources for predicting the occurrence of earthquakes because the effects of preventive works beforehand can far outweigh those of remedial measures afterwards. However, it is difficult and impossible to accurately predict the scale, depth, location and timing of earthquakes in terms of current technologies available to mankind. A two-stage procedure is proposed to predict the scales of earthquakes by using the Self-Organizing Map (SOM) and Support Vector Regression (SVR) based on the historical earthquakes in this study. The effectiveness and efficiency of the proposed procedure are illustrated by a case study based on the earthquakes occurred from 1995/1/1/ to 2024/12/14 in Taiwan. The experimental results show that the proposed procedure of this study can yield adequate prediction performance through evaluating by the MAPE (Mean Absolute Percentage Error), MSE (Mean Squared Error) and R^2.

Keywords: Earthquakes, Prediction, Self-organizing map (SOM), Support vector regression (SVR)

1. Introduction

Earthquake is a natural disaster that humans cannot avoid since ancient times. Whenever a strong earthquake occurs, countries around the earthquake-affected areas often suffer huge losses in life, property, safety, economy, society, culture and politics. For example, the 2004 Sumatra-Andaman earthquake caused the South Asian tsunami. The earthquake triggered a huge tsunami with a height of 15 meters to about 30 meters. Coupled with the Christmas holiday, at least 300,000 people were killed and missing. The Miyagi earthquake, also named Great East Japan Earthquake, in March 2011, including the accompanying huge tsunami and large-scale disasters caused by aftershocks, flooded areas several kilometers away from the coast. Many coastal cities and facilities were destroyed, and the economic losses are even more incalculable. The huge tsunami caused by the earthquake eventually led to the accident at the Fukushima Daiichi Nuclear Power Plant. In addition, many countries have imposed "import restrictions" on Japan's Fukushima foods for several years, causing Japan's export trade setbacks and losses that cannot be ignored. Taiwan locates in the Circum-Pacific seismic belt, the largest and most active seismic zone in the world. Due to the intersection and collision of the Eurasian and the Philippine Sea Plates, as well as the subduction of the plates, Taiwan's strata are subjected to the crustal stress, causing the strata to deform and fracture, thus leading to quite frequent earthquakes. Therefore, Taiwan's government has been allocating a large amount of funds for earthquake-related researches and earthquake

recovery works. At the same time, the government also clearly understands that the preventive works beforehand will far outweigh the effects of remedial measures afterwards. However, the mechanisms that trigger earthquakes and their influencing factors are extremely complex, making it impossible for humans to accurately predict the scale, depth, location and timing of earthquakes. Therefore, the prediction of earthquake-related information has always been a research topic of great interest and enthusiasm around the world. For example, [1] propose an approach for efficiently predicting the number of the wounded in a very short time, i.e. an "S-shape" curve for the numbers of the sick and wounded by utilizing a continuous interval grey discrete Verhulst model that is based on the kernels and measures (CGDVM-KM). The authors apply an interval whitening method to convert the continuous interval sequence into the kernel, as well as to measure the sequences with equal information quantity. In other words, the continuous interval sequence is first converted into the kernel and measure sequences with equal information quantity through the interval whitening method. The kernel and measure sequences are then combined with the classical grey discrete Verhulst model to present the grey discrete Verhulst models for the kernel and measure sequences, respectively. Based on these procedures, they develop the CGDVM-KM model to overcome the systematic errors caused by the discrete form equation for parameter estimation, as well as continuous form equation for simulation and prediction in classical grey Verhulst model. Therefore, the prediction accuracy thus can be improved. In addition, several examples are used to verify the rationality and validity of their model. According to the comparison with other forecasting models indicates that their proposed model can provide higher prediction accuracy, yield the better simulation effect while forecasting the wounded in massive earthquake disasters. [2] integrate an alarm-based model and binary classification to explore the precursory behavior of geoelectric signals before large earthquakes. They remove a time parameter used for coarse-graining of earthquake occurrences, as well as extend the single-station method into a joint-stations approach to improve the original methodology. The optimal frequency bands of earthquake-related geoelectric signals featuring the highest signal-to-noise ratio can be determined by analyzing the filtered geoelectric data with different frequency bands. Furthermore, the relationship between geoelectric signals and seismicity is proven according to the significance tests. Therefore, the machine learning methods are suggested to extract the underlying relationship, that could be used as a tool for quantifying probabilistic forecasts of impending earthquakes and getting closer prediction of the operational earthquakes. [3] create a hybrid decision-making framework for deciding amongst potential loss mitigation actions (or taking no action) by combining a multiple criteria decision-making approach and the cost-benefit analysis. Three hypothetical case studies by using Patras (Greece), that is a high seismicity location, are taken as an example to demonstrate their proposed framework. According to the simulation results, their proposed approach is flexible enough for dealing with new problems, end-users and stakeholders. Furthermore, it also shows that reasonable mitigation actions are viable and financially beneficial during periods of increased seismic hazard for reducing the potential consequences of earthquakes. The framework is very sensitive to the inputs that involving end users to help constrain these inputs when making calculations is a crucial work. [4] predict the labquake by deep learning (DL) methods, forecast fault zone shear stress through introducing an autoregressive (AR) forecasting DL model, as well as expand the range of lab fault zones for the purpose of laboratory earthquake prediction and fault zone stress forecasting. In the AR methods, previous measurements are utilized to forecast stress in the future by predicting iteratively. In addition, their DL methods can perform better than the existing ML approaches, and also can predict by using limited training. Besides, the forecasts beyond a single seismic cycle for aperiodic failure are also investigated in their study. The experimental results show that there are significant improvements for predicting the labquake, and demonstration can be made for the followings: (1) DL models based on Long-Short Term Memory and Convolution Neural Networks can predict labquakes in these conditions including pre-seismic creep, aperiodic events and alternating slow/fast events, as well as (2) the acoustic energy is a fingerprint of fault zone stress since the fidelity can predict the fault zone stress well. Both the time to start of failure (TTsF) and time to the end of Failure (TTeF) for labquakes can be predicted. Notably, the TTeF can be successfully predicted in all seismic cycles, but the TTsF prediction can vary along with the amount of preseismic fault creep. They apply a forecasting modelling framework with three sequences including the LSTM, Temporal Convolution Network and Transformer Network. The AR can predict only a target variable at a specific time, that is distinct from the existing prediction models. Furthermore, the forecasts beyond a single seismic cycle are limited, but are encouraging. Their proposed ML/DL models can outperform the state-of-the-art and

autoregressive model and improve the performance of forecasting earthquakes by using the current methods. [5] replace the instantaneous triggering by a mean time-to-failure depending on the absolute stress value to develop a modified Coulomb Failure (CF) model to forecast the time-dependent earthquakes. Notably, absolute Coulomb stress is required and instantaneous triggering if stress exceeds a threshold is assumed in the traditional CF model. Therefore, the CF model is suited to predict background earthquake rates and time-dependent stress effects. They show that their modified model can lead to identical results as the rate-and-state (RS) model, as well as can reproduce the Omori-Utsu relation for aftershock decays and stress-shadowing effects by specifically choosing an exponential dependence on stress and a stationary initial seismicity rate. [6] forecast the probability of a strong aftershock of Greek seismicity from 1995 to 2022 by using the NESTORE machine learning approach. The NESTORE classifies aftershocks into two types, named Type A and Type B, based on their magnitude difference between the mainshock and the strongest aftershock. There is smaller difference between the mainshock and strongest aftershock in Type A clusters. In their algorithm, region-dependent training is required as input as well as the forecasting performance is evaluated on an independent test set. According to the best test result obtained by 6 hours after the mainshock, the 92% of clusters corresponding to 100% of Type A clusters and more than 90% of Type B clusters can be correctly forecasted. The experimental results show that the algorithm can be applied in this Greek area, and their approach is particularly attractive for seismic risk mitigation since the required forecasting time is short. [7] forecast the frequency and magnitude for earthquakes above Mw 4.0 in Northeastern Algeria through parametric and non-parametric time series forecasting approaches. In their study, the Autoregressive Integrated Moving Average (ARIMA) model is used in the parametric approach, as well as the Singular Spectrum Analysis (SSA) approach utilized in the non-parametric method. The annual number of earthquakes and maximum magnitude events occurred in Northeastern Algeria from 1910 to 2019 that including 287 main events larger than Mw 4.0 are used to train and test the ARIMA and SSA models. Notably, the SSA method takes as the function of forecasting algorithm, and the obtained results are also compared to those acquired by the ARIMA model. The implementation results indicate that the SSA forecasting model can perform better than the ARIMA model based on the evaluation criterion through root mean square error (RMSE). A statistical N-test in terms of the total number of events is also applied to analyze the consistency between the observation and forecast. According to their findings, the annual maximum magnitude in Northeastern Algeria between 2020 and 2030 will range from Mw 4.8 to Mw 5.1, and there are four and six occurred events that have a magnitude of at least Mw 4.0 annually. [8] utilize a deep-learning model based on neural networks to develop the Recurrent Earthquake foreCAST (RECAST) system. The greater volume and diversity of earthquake observations access can be made to overcome the theoretical and computational limitations in the traditional approaches. Their proposed approach is benchmarked with a temporal Epidemic Type Aftershock Sequence model. The RECAST can accurately model earthquake-like point processes directly from cataloged data according to the tests on synthetic data. Next, the fit and forecast accuracy compared to the benchmark can be improved while testing on earthquake catalogs in Southern California if the training set is sufficiently long (>10(4) events). Furthermore, the forecasting performance need not be sacrificed while adding flexibility and scalability for earthquake forecasting by the basic components in RECAST. [9] design an earthquake forecasting framework, tested it in seismogenic regions in southwestern China, to forecast earthquakes in real time. The features recorded using the electromagnetic (EM) and geo-acoustic (GA) sensors of the multi-component seismic monitoring system acoustic electromagnetic to AI (AETA) in each station take as the input data. By giving the data of the current week, the forecasting framework intends to forecast the location and magnitude of the earthquake that might occur during the next week. In addition, a principal component analysis is utilized to reduce the dimension from massive EM and GA data, and the random-forest-based classification is then applied. The authors gather the available data from 2016 to 2020 to train the proposed algorithm, that is evaluated later based on the real-time data in 2021. According to the experimental results, the testing accuracy, precision, recall and F1-score can reach 70%, 63.63%, 93.33% and 75.66%, respectively. Besides, the mean absolute error regarding the predicted distance is 381 km, as well as the mean absolute error of magnitude is 0.49. [10] propose a hybrid SARIMA-XGBoost methodology to predict earthquake magnitude for addressing the challenges given by earthquakes probabilistic nature. Their proposed approach consists of two steps. The first step is an exploration procedure that uses exploratory data analysis, that includes descriptive statistics and data visualization. Then, a method

focusing on forecasting future earthquakes forms the second step. They use an earthquake dataset during 1965 to 2023 to acquire insights and lessons for more effective earthquake prediction methods. In addition, the implementation results of SARIMA-XGBoost model are compared to those obtained by the traditional methods including ARIMA and SARIMA models. Based on evaluating by the mean absolute error (MAE), mean squared error (MSE) and root mean squared error (RMSE), their developed hybrid SARIMA-XGBoost model can provide superior prediction performance. Therefore, the prediction accuracy can be significantly improved through yielding the notably low values of MAE, MSE and RMSE. Furthermore, the hybrid model can reduce the forecasting errors more effectively by integrating SARIMA's time series (TS) analysis with XGBoost's machine learning (ML). [11] apply the fully convolutional networks (FCN) to forecast future earthquakes by using the spatial map of the logarithm of past estimated released earthquake energies as neural networks' inputs. The earthquakes data of California is implemented to their developed model, that is also compared with the elaborate version of the epidemic type aftershock sequence (ETAS) model. Based on the long-term earthquake forecast simulation, whose performance is evaluated by the Molchan diagram, the FCN model is close to the ETAS model while forecasting earthquakes with a magnitude that is greater than or equal to 3.0, equal to 4.0, and equal to 5.0. Furthermore, it is 2000-4000 times faster than calibrating the ETAS model and generating its probabilistic forecasts through training and implementing the FCN model. Extensive knowledge of statistical seismology or the analysis of earthquake catalogue completeness is not required. The forecasting model's performance in some time-magnitude forecasting windows can be enhanced by serving the earthquake catalogue with a magnitude greater than or equal to 0 as FCN input. At the same time, the neural network structure and feature engineering make the FCN model in their study straightforward. [12] proposes a model for forecasting the potential seismogenic earthquake or earthquake fault zones by using the complete earthquake catalog data and spatio-temporal analysis. The author utilizes the complete shallow earthquake catalog between 1963-1999 and 1963-2006 to constructs and compares three types' models based on evaluating their reliability via the delta AIC. The experimental results show that the Model-3, that is constructed based on the product of the normalized model of the combined smooth seismicity model of a relatively small to moderate complete earthquake catalog data with a relatively uniform background model and weighted by the normalized seismic moment rate derived from the surface strain rate, yields the highest reliability. Therefore, a more extended observation period and using a complete, albeit relatively small-to-moderate, earthquake catalog is suggested for constructing a more reliable and accurate model. In addition, it is considered successful by implementing the Probabilistic Seismic Hazard Function (PSHF) window using the b-value of a 5-year window length with a 1-year sliding window prior to a significant seismic event. The importance of the temporal "b-value" in conjunction with the reliable seismicity rate and spatial probabilistic earthquake forecasting models in earthquake forecasting is also proven. The large changes in the PSHF prior to giant and large earthquakes and the finding of a correlation between decreased b-value time window length and earthquake magnitude are also shown in his study.

Based on the above literature review, there are numerical traditional methods or models for dealing with the problems of forecasting earthquakes, e.g. the Verhulst model, multiple criteria decision-making, autoregressive (AR) forecasting, Coulomb Failure (CF) model, and Autoregressive Integrated Moving Average (ARIMA) model etc. With the rapid development of AI in recent years, the artificial neural networks and bionic algorithms have been extensively used in the field of earthquake prediction and provide adequate results. Therefore, this study develops a two-stage procedure to predict the scales of earthquakes in the future based on the historical earthquake data by using the self-organizing map (SOM) and support vector regression (SVR) algorithm. The earthquake data collected from 1995/1/1/ to 2024/12/14 in Taiwan are used to demonstrate the effectiveness and efficiency of the proposed procedure. The rest of this paper is organized as follows. Section 2 briefly introduces the methodologies, including the SOM and SVR algorithms, that are used to develop the prediction procedure in this study. The proposed two-stage earthquake prediction procedure is then presented in Section 3. Our proposed procedure is validated by a case study on predicting the scales of earthquakes occurred in Taiwan between 1995/1/1 and 2024/12/14 in Section 4 where the prediction performance is evaluated through MAPE (Mean Absolute Percentage Error), MSE (Mean Squared Error) and R2. Finally, Section 5 concludes this study and gives a direction for the future research.

2. Methodologies

Self-Organizing Map

A neural network (NN) inspired by the structure and function of biological neural networks in animal brains is a model in the domain of machine learning. There are several connected units or nodes, called artificial neurons, to model the neurons of the brain in a neural network. In addition, these neurons are connected by edges that simulate the synapses in the brain. The neurons are arranged into several layers depending on their connecting relationship. The typical neural network consists of three layers, including the input, hidden and output layers. For each neuron, the signals coming from all connected neurons in the previous layer are summed through weighting the signal by the weight corresponding to the connected edge and fed into this neuron. Each neuron then sends a signal to all connected neurons in the next layer after processing the received signal of this neuron. Notably, the signal is a real number, as well as the output of each neuron is calculated by a non-linear function, called the activation function, of the sum of a neuron's inputs. Furthermore, the weights that can adjust during the learning process in the connected edges determine the strength of the signal at each connection. The neural networks can be broadly divided into two categories: supervised learning and unsupervised learning. In the supervised learning, a paradigm where the input objects and their corresponding desired output values are used to train a model. The training data are applied to build a function that can map new input data to an expected output value. The optimal neural model is determined by allowing the algorithm of neural network to correctly determine the output values for these unseen cases. Therefore, the learning algorithm is required to be able for generalizing from the training data to the unseen situations through a reasonable method. The quality of an algorithm is measured by the statistical measure, called generalization error. On the other hand, the unsupervised neural networks train on unlabeled data without explicit input-output pairs. The unsupervised learning is not under the guidance of features. Instead, the unsupervised neural network is provided with unlabeled data sets containing only the input data and left to discover the patterns existed in the data to build a new model from these data. The self-organizing map (SOM) developed by [13] is a well-known unsupervised machine learning technique that is used to produce a low-dimensional, typically two-dimensional, representation for a higher-dimensional data set while preserving the topological structure existed within the original data as shown in Figure 1. The learning goal of the self-organizing map is to create different parts of the neural network that can respond similarly to certain input patterns. The motivation of SOM is partly coming from how visual, auditory or other sensory information is handled in separate parts of the cerebral cortex in the human brain. In general, the weights of the neurons are initialized randomly or based on sampling from the subspace spanned by the two largest principal component eigenvectors. The competitive learning is utilized while training the SOM. The Euclidean distance to all weight vectors is computed when a training example is fed to the network, and the neuron whose weight vector is most similar to the training example is considered as a winner, i.e. the best matching unit (BMU). The weights of the BMU and neurons close to the BMU in the SOM grid are then adjusted towards the input vector of the training example. In addition, the magnitude of the change will decrease along with learning time, as well as with the grid-distance from the BMU. The learning algorithm of SOM is as follow:

Step 1: Initialize the node weight vectors in a map randomly.

Step 2: For $t = 0,1,2,m$

Step 2-1: Randomly pick an input vector $V_i(t)$.

Step 2-2: Find the node that most closes to the input vector in the map as the best matching unit (BMU), denoted by s.

Step 2-3: For each node j, update its vector by pulling closer to the input vector:

$$V_j(t+1) = V_j(t) + NF(s,j,t) \cdot LRS(t) \cdot (V_i(t) - V_j(t)) \quad (1)$$

Here,

t is the current iteration;

m is the iteration limit;

$V_i(t)$ is the vector of the input data;

s is the index of the best matching unit (BMU) in the map;

$V_j(t)$ is the current weight vector of node j;

$NF(s,j,t)$ is the neighborhood function;

$LRS(t)$ is the learning rate schedule.

The SOM has been broadly applied in various fields, e.g. [14-16].

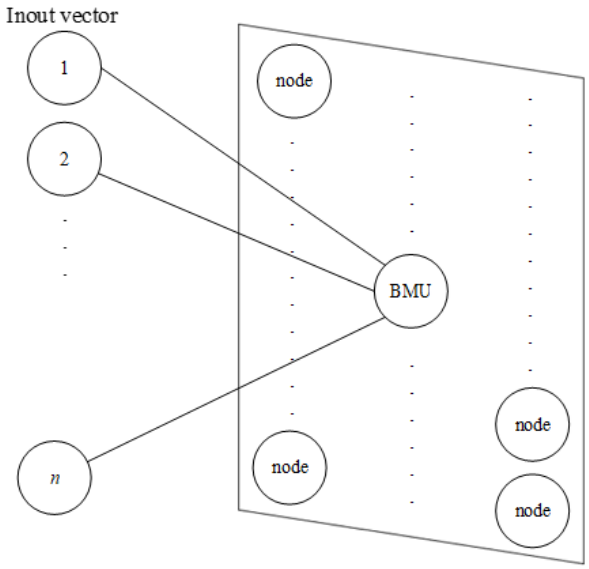

Figure 1: Self-organizing map

Support Vector Regression

The support vector machine (SVM) invented by [17] is a tool widely used to resolve classification problems in the field of machine learning. Based on the same principles as SVM, [18] proposed the support vector regression (SVR) to focus on predicting continuous outputs rather than classifying data points. Unlike the typical regression models, the SVR transforms the input features into spaces in a higher dimension for locating the ideal hyperplane to accurately represent the data. Through this method, the SVR can effectively manage both linear and non-linear relationships that renders it a powerful tool in various fields. Furthermore, the SVR can attain high accuracy and robustness since even though when dealing with complex datasets since the distinctive features of support vector machine are utilized.

Suppose there are *N* pairs of inputs $X_k = (x_{k1}, x_{k2}, \ldots, x_{kn}) \in \mathbf{R}^n$ and output $y_k \in \mathbf{R}$. A regression model is built to represent the functional relationship between X_k and y_k. For separating more easily, the original *n*-dimensional inputs $X_k = (x_{k1}, x_{k2}, \ldots, x_{kn})$ are first transformed into $\varphi_i(X_k)$ in a higher-order dimensional space. The linear

regression model thus can be described as:

$$y'_k = f(X_k, W) = \sum_i^m w_i \varphi_i(X_k) + w_0 = W^T \Phi(X_k) + w_0, k = 1,2, \ldots N \quad (2)$$

where y'_k, W, $\Phi(X_k)$ and w_0 are the predicted output, weight vector, feature vector and bias, respectively.

[19] introduced an ε-insensitive loss function to evaluate the predicted error as:

$$L_\varepsilon(y_k, y'_k) = \begin{cases} 0 & if\ |y_k - y'_k| \le \varepsilon \\ |y_k - y'_k| - \varepsilon & otherwise \end{cases} \quad (3)$$

to illustrate that there is no loss arisen when the predicted output locates within an acceptable distance from the actual (target) output. This can be re-stated as follows:

$$y_k - W^T \Phi(X_k) - w_0 - \varepsilon \le \xi_k, k=1,2,\ldots,N \quad (4)$$

$$W^T \Phi(X_k) + w_0 - y_k - \varepsilon \le \xi'_k, k=1,2,\ldots,N \quad (5)$$

$$\xi_k \ge 0, k = 1,2, \ldots, N \quad (6)$$

$$\xi'_k \ge 0, k = 1,2, \ldots, N \quad (7)$$

In other words, ξ_k denotes the error term if the actual output y_k is larger than the predicted value y'_k and the error is represented by ξ'_k when the actual output y_k is less than the predicted value y'_k. Therefore, a problem by minimizing the empirical risk can be formulated as follows:

Minimize

$$\frac{1}{2}||W||^2 + C\left(\sum_{k=1}^{Q}\xi_k + \sum_{k=1}^{Q}\xi'_k\right) \quad (8)$$

subject to the constraints shown in Equations (4)-(7). In Equation (8), the parameter C used to balance the complexity and loss must be specified by users in advance. The Lagrangian by using primal variables thus can be formulated as follows:

$$L_P(W, w_0, \Xi, \Xi', \Lambda, \Lambda', \Gamma, \Gamma')$$
$$= \frac{1}{2}W^T W + C\left(\sum_{k=1}^{Q}\xi_k + \sum_{k=1}^{Q}\xi'_k\right) - \sum_{k=1}^{Q}\lambda_k\left(W^T\Phi(X_k) + w_0 - y_k + \varepsilon + \xi_k\right) -$$
$$\sum_{k=1}^{Q}\lambda'_k\left(y_k - W^T\Phi(X_k) - w_0 + \varepsilon + \xi'_k\right) - \sum_{k=1}^{Q}(\gamma_k\xi_k + \gamma'_k\xi'_k). \quad (9)$$

Here, $\Xi = (\xi_1, \ldots, \xi_Q)^T$ and $\Xi' = (\xi'_1, \ldots, \xi'_Q)^T$ are vectors consisting of slack variables. In addition, $\Lambda = (\lambda_1, \ldots, \lambda_Q)^T$、 $\Lambda' = (\lambda'_1, \ldots, \lambda'_Q)^T$、 $\Gamma = (\gamma_1, \ldots, \gamma_Q)^T$ and $\Gamma' = (\gamma'_1, \ldots, \gamma'_Q)^T$ are the Lagrangian multiplier vectors of Equations (4)-(7). The optimality thus can be found by taking the partial derivative of LP with respect to the primal variables to its saddle point, as follows:

$$\frac{\partial L_P(W, w_0, \Xi, \Xi', \Lambda, \Lambda', \Gamma, \Gamma')}{\partial W} = 0 \Rightarrow W = \sum_{k=1}^{Q}(\lambda_k - \lambda'_k)\,\Phi(X_k) \quad (10)$$

$$\frac{\partial L_P(W, w_0, \Xi, \Xi', \Lambda, \Lambda', \Gamma, \Gamma')}{\partial w_0} = 0 \Rightarrow \sum_{k=1}^{Q}(\lambda_k - \lambda'_k) = 0 \quad (11)$$

$$\frac{\partial L_P(W, w_0, \Xi, \Xi', \Lambda, \Lambda', \Gamma, \Gamma')}{\partial \xi_k} = 0 \Rightarrow \gamma_k = C - \lambda_k \quad (12)$$

$$\frac{\partial L_P(W, w_0, \Xi, \Xi', \Lambda, \Lambda', \Gamma, \Gamma')}{\partial \xi'_k} = 0 \Rightarrow \gamma'_k = C - \lambda'_k \quad (13)$$

Through substituting Equation (10), (11), (12), and (13) into Equation (9), the simplified dual form LD then can be obtained as follows:

Maximize

$$L_D(\Lambda, \Lambda') = \sum_{k=1}^{N} d_k\,(\lambda_k - \lambda'_k) - \varepsilon\sum_{k=1}^{N}(\lambda_k + \lambda'_k) - \frac{1}{2}\sum_{k=1}^{N}\sum_{l=1}^{N}(\lambda_k - \lambda'_k)(\lambda_l - \lambda'_l)\Phi(X_k)\Phi(X_l). \quad (14)$$

Subject to

$$\sum_{k=1}^{N}(\lambda_k - \lambda'_k) = 0 \quad (15)$$
$$0 \le \lambda_k \le C, k = 1,2,\ldots,N \quad (16)$$
$$0 \le \lambda'_k \le C, k = 1,2,\ldots,N \quad (17)$$

Here, the inner product $\Phi(X_k)\Phi(X_l)$ can also be expressed in terms of $K(X_k, X_l)$, called the kernel function, according to Mercer's Theorem. The commonly used kernel functions include linear, polynomial (homogeneous), polynomial (inhomogeneous), radial basis function, and hyperbolic tangent, etc. The $\widehat{W}$, i.e. the optimal approximation for the weight vector W is therefore acquired by optimizing the Lagrangian as follows:

$$\widehat{W} = \sum_{k=1}^{n_s}(\hat{\lambda}_k - \hat{\lambda}'_k)\,\Phi(X_k) \quad (18)$$

Here, the n_s is the total number of support vectors and the index k only runs over the support vectors 1 to n_s. The support vectors are X_ks that lie nearest to the max-margin hyperplane in SVR.

SVR had been extensively applied in various fields, e.g. [20-22].

Silhouette Coefficient

Silhouette proposed by [23] is a method for interpretating and validating the consistency within clusters of data. How similar an object is to its own cluster compared to other clusters can be evaluated by the silhouette value. The silhouette ranges between -1 and +1, and a high silhouette indicates that the object is well matched to its own cluster and poorly matched to the neighboring clusters. In other words, the clustering configuration is appropriate when most objects can have high silhouettes. On the other hand, there are too many or too few in the clustering configuration if many points have a low or negative silhouette value. The clustering can be considered as "strong", "reasonable", and "weak" when an average silhouette is over 0.7, over 0.5, and over 0.25, respectively. Notably, it becomes difficult to achieve high silhouette values with increasing dimensionality of data since the curse of dimensionality, as the distances become more similar. The silhouette can be calculated based on any distance metric, such as the Euclidean distance or the Manhattan distance.

Assume the data have been clustered by using any technique into k clusters.

For data point i in the cluster C_I, i.e. $i \in C_I$, let

$$a(i) = \frac{1}{|C_I|-1}\sum_{j \in C_I, i \neq j} d(i,j) \quad (19)$$

be the mean distance between i and all other data points in the same cluster.

Where,

$|C_I|$ is the number of points belonging to cluster C;

$d(i,j)$ is the distance between data points i and j in the cluster C.

$a(i)$ is a measure of how well i is assigned to its cluster. The smaller the $a(i)$ value implies the better assignment.

The mean dissimilarity of point i to some cluster C_J is defined as the mean of the distance from i to all points in C_J that differs from C_I, i.e. $C_J \neq C_I$.

For each data point $i \in C_I$, define

$$b(i) = \min_{J \neq I} \frac{1}{|C_J|}\sum_{j \in C_J} d(i,j) \quad (20)$$

Therefore, a silhouette (value) of data point i can be defined as

$$s(i) = \frac{b(i)-a(i)}{max\{a(i),b(i)\}}, if\ |C_I| > 1 \quad (21)$$

and

$$s(i) = 0, if\ |C_I| = 1 \quad (22)$$

The $s(i)$ ranges from -1 to +1, and is set as 0 when the cluster's size equals 1. The cluster that has the smallest mean dissimilarity is called as the "neighboring cluster" of i since because the cluster is the next best fit cluster for point i. Later, [24] introduced the silhouette coefficient for the maximum value of the mean $s(i)$ over all data of the entire dataset as

$$SC = \max_k \tilde{s}(k) \quad (23)$$

where $\tilde{s}(k)$ is the mean $s(i)$ over all data of the entire dataset for a specific number of clusters k. A larger silhouette coefficient indicates the better clustering results.

Proposed Procedure

In this study, the self-organizing map (SOM) and support vector regression (SVR) algorithm are used to propose a systematic approach for predicting the scales of earthquakes in the future as briefly depicted in Figure 2 and illustrated as follows:

Step 1: Data Collection

Step 1-1: Collect the scale and depth of each earthquake occurred during the research period.

Step 1-2: Normalize the scales and depths into [-1, 1] based on their corresponding maximum and minimum values.

Step 2: Data Clustering

Step 2-1: Cluster the normalized earthquake data obtained in Step 1-2 into clusters of different numbers by using SOM with various structures.

Step 2-2: Determine the best number of clusters based on the clustering performance evaluated by the silhouette coefficient.

Step 3: Prediction Data Preparation

Step 3-1: Set the scale of each earthquake in each cluster as the dependent variable

Step 3-2: Set the scales of earthquakes of a certain number, e.g. 10, occurred previously as the corresponding independent variables for the dependent variable set in Step 3-1.

Step 3-3: Merge each dependent variable's value (set in Step 3-1) along with its corresponding independent variables' values (set in Step 3-2) to form the prediction data.

Step 4: Data Segmentation

Divide the prediction data prepared in Step 3-3 into training and test data according to a pre-determined ratio, e.g. 3:1, randomly or based on a certain principle.

Step 5: SVR Model Construction

For each cluster, construct a prediction model by using the SVR technique to the training and test data determined in Step 4.

Step 6: Prediction Performance Evaluation

Evaluate the prediction performance of SVR models for clusters of various numbers when implementing these SVR models on the test data that are never seen before.

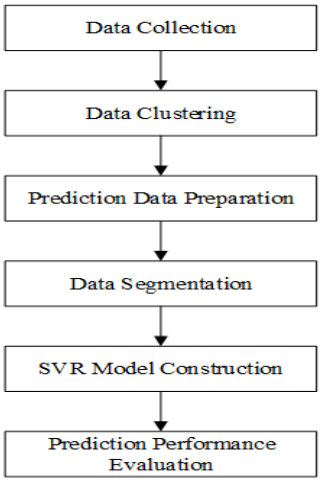

Figure 2: Proposed procedure

3. Case Study

Data Collection

Earthquakes of more than a hundred occurs per year in Taiwan. There are even 466 earthquakes in 2023. However, most of these earthquakes are noninductive. Furthermore, the detailed observation data of earthquakes can only be available on the web since 1995. Therefore, the scales and depths of 4292 inductive earthquakes occurred from 1 January 1995 to 14 October 2024 are collected as the raw data in this case study. The scales and depths are then normalized into [-1, 1] based on (7.3, 2.1) and (268.6, 0.1) that are the maximum and minimum values of scales and depths, respectively.

Data Clustering

The SOM neural network is applied to cluster the normalized earthquake data. Here, the SOM function is implemented by utilizing Weka 3.9.5 [25] data mining software. The topology of SOM is set as a rectangle, consisting of the scale and depth, with a height of 1 and a various width from 1 to 10 whose parameters' setting is depicted in Figure 3. The rest parameters are fixed based on their default settings. The clustering performance is evaluated by the silhouette coefficient. The implementation results are summarized in Table 1 where an asterisk denotes the best clustering. Hence, the original data are divided into 5 groups with 360, 1232, 242, 1772 and 687 items, respectively.

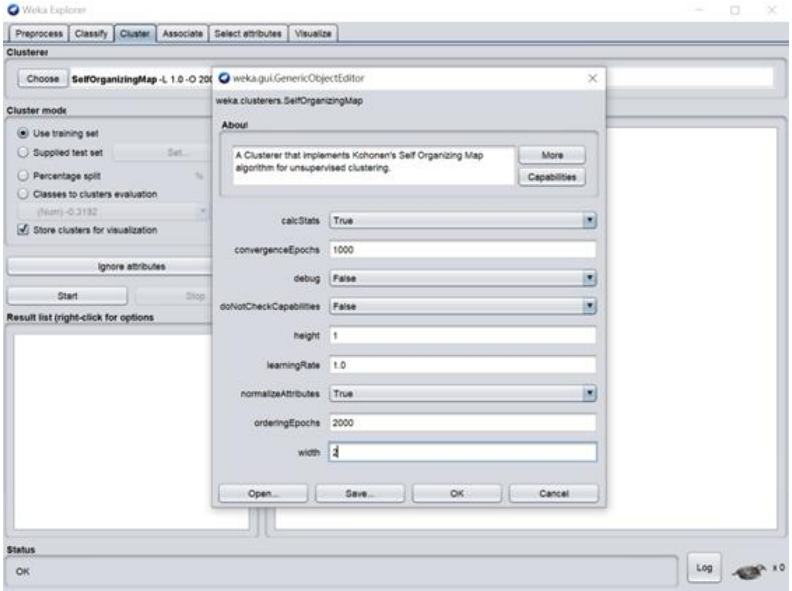

Figure 3: Parameters' setting in Weka.

Table 1: Clustering results of SOM

Number of clusters	2	3	4	5	6	7	8	9	10
Silhouette coefficient	0.5147	0.4534	0.4534	0.5317*	0.4843	0.4841	0.4667	0.4631	0.4785

Prediction Data Preparation

For each cluster, the scale of each earthquake clustered in this group is set as the dependent variable and the scales of 10 earthquakes occurred previously are set as the corresponding independent variables. Therefore, there are 350, 1222, 232, 1762 and 677 items in clusters 1 to 5, respectively. These data form the prediction data for SVR.

Data Segmentation

Randomly divide the prediction data into training and test data based on a 3:1 ratio. For example, the training and test data include 263 and 87 items, respectively, for the first cluster.

SVR Model Construction

The SVR technique that is implemented by the LIBSVM software [26] for the training and test data in each cluster to construct a prediction model. A grid search method is applied to determine the optimal parameters of SVR including C, Gamma and epsilon. In addition, an epsilon-SVR is utilized in this study thus the parameter s of LIBSVM is set as shown in Figure 4 that illustrates the best parameters' setting obtained by the grid search for the first cluster. Therefore, the best SVR prediction model then can be acquired according to the found parameters' setting.

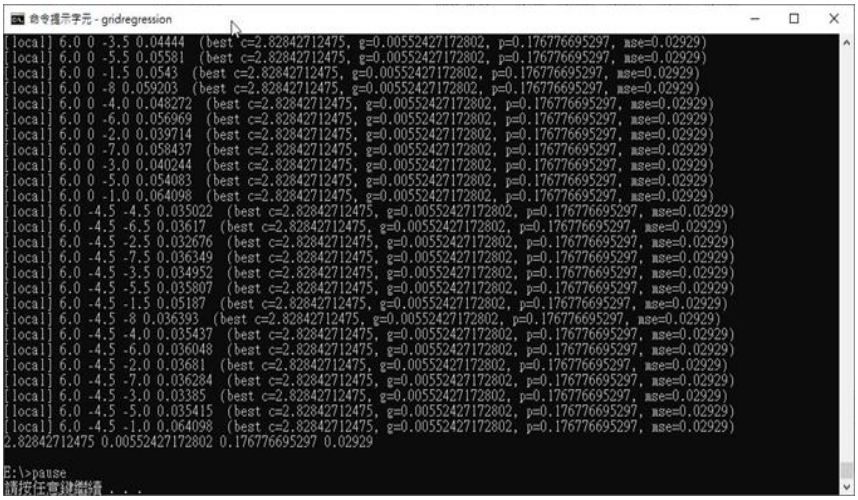

Figure 4: Best parameters' setting of SVR for the first cluster

Prediction Performance Evaluation

The test data that are never met before are fed into the obtained SVR prediction models to realize the generalization capability of these SVR models. The prediction performance is evaluated by the MAPE (Mean Absolute Percentage Error), MSE (Mean Squared Error) and R2. Table 2 shows the implementation results. According to Table 2, the maximum value of R2 for the training and test data can only attain 0.08227 and 0.02573, respectively.

The minimum MAPEs are 3.76% and 3.88% regarding the training and test data, respectively. The SVR model bring the minimum MSEs as 0.00654 and 0.00570 when implementing on the training and test data, receptively. Based on these results, the prediction performance for predicting the scales of earthquakes can be considered adequate since it is essentially a very hard problem even though the performance indices do not look well.

Table 2: Results of implementing SVR models on the test data

Cluster 1				Cluster 4			
Index	MAPE	MSE	R^2	Index	MAPE	MSE	R^2
Training data	0.06172	0.02852	0.08227	Training data	0.04316	0.00654	0.04226
Test data	0.05470	0.02061	0.00009	Test data	0.03968	0.00570	0.01785

Cluster 2

Index	MAPE	MSE	R^2
Training data	0.03757	0.00683	0.01855
Test data	0.03877	0.00742	0.00124

Cluster 5

Index	MAPE	MSE	R^2
Training data	0.06278	0.01055	0.06407
Test data	0.07005	0.01301	0.02573

Cluster 3

Index	MAPE	MSE	R^2
Training data	0.09841	0.05628	0.03518
Test data	0.08640	0.04471	0.00142

4. Conclusions

Since ancient times, earthquakes always have been a kind of natural disaster that humans scare very much due to their uncertainties about occurring timing, scales, depths and locations. Furthermore, the countries around the earthquake-affected areas will face large amount of losses in life, property, safety, economy, society, culture and politics when a strong earthquake occurs. However, human had not fully realized the factors that affect the occurrence of earthquakes and the mechanisms trigging earthquakes. Although it is a difficult problem for accurately predicting the information of earthquakes, e.g. the scale, depth, location and timing, in advance, earthquake prediction problems have always attracted the interests of researchers and been an enthusiasm scientific issue around the world. Especially, Taiwan locates in the Circum-Pacific seismic belt that is the largest and most active seismic zone of the earth. Taiwan must face earthquakes quite frequently since the intersection, collision and subduction of the Eurasian and Philippine Sea Plates that trigger the strata to deform and fracture. The earthquake prediction problem therefore particularly becomes an important issue for the people and government of Taiwan. In this study, the self-organizing map (SOM) neural network and support vector regression (SVR) are utilized to develop a systematic procedure for predicting the scales of earthquakes. A case study on predicting the scales of earthquakes during 1995/1/1/ and 2024/12/14 in Taiwan is used to demonstrate the effectiveness and efficiency of the proposed procedure. According to the experimental results, the procedure proposed in this study can provide adequate prediction performance by evaluating the prediction performance based on the MAPE (Mean Absolute Percentage Error), MSE (Mean Squared Error) and R2. Prediction of the depths, timing and locations of future earthquakes more precisely might be a direction in the future research.

Acknowledgment

The author would like to thank the partial support of Minghsin University of Science and Technology, Taiwan, R.O.C. under Contract No. MUST-114BA-03.

References

[1]. Zhang, J., Wang, T. Y., Chang, J. P., & Gou, Y. (2021). Forecasting the number of the wounded after an earthquake disaster based on the continuous interval grey discrete Verhulst model. Discrete Dynamics in Nature and Society, 2021, Document No. 6654288.

[2]. Chen, H. J., Chen, C. C., Ouillon, G, & Sornette, D. (2021). A paradigm for developing earthquake probability forecasts based on geoelectric data. European Physical Journal-Special Topics, 230(1), 381-407.

[3]. Azarbakht, A., Rudman, A., & Douglas, J. (2021). A decision-making approach for operational earthquake forecasting. International Journal of Disaster Risk Reduction, 66, DOI: 10.1016/j.ijdrr.2021.102591.

[4]. Laurenti, L., Tinti, E., Galasso, F., Franco, L., & Marone, C. (2022). Deep learning for laboratory earthquake prediction and autoregressive forecasting of fault zone stress. Earth and Planetary Science Letters, 598, DOI: 10.1016/j.epsl.2022.117825.

[5]. Dahm, T., & Hainzl, S. (2022). A Coulomb stress response model for time-dependent earthquake forecasts. Journal of Geophysical Research-Solid Earth, 127(9), DOI: 10.1029/2022JB024443.

[6]. Anyfadi, E. A., Gentili, S., Brondi, P., & Vallianatos, F. (2023). Forecasting strong subsequent earthquakes in Greece with the machine learning algorithm NESTORE. Entropy, 25(5), Document No. 797.
[7]. Merdasse, M., Hamdache, M., Peláez, J. A., Henares, J., & Medkour, T. (2023). Earthquake magnitude and frequency forecasting in northeastern Algeria using time series analysis. Applied Science-BASEL, 13(3), Document No. 1566.
[8]. Dascher-Cousineau, K., Shchur, O., Brodsky, E. E., & Günnemann, S. (2023). Using deep learning for flexible and scalable earthquake forecasting. Geophysical Research Letters, 50(17), Document No. e2023GL103909.
[9]. Saad, O. M., Chen, Y. F., Savvaidis, A., Fomel, S., Jiang, X. X., Huang, D. N., Yong, S. S., Wang, X'., Zhang, X., & Chen, Y. K. (2023). Earthquake Forecasting Using Big Data and Artificial Intelligence: A 30-Week Real-Time Case Study in China. Bulletin of The Seismological Society of America, 113(6), 2461-2478.
[10]. Kaushal, A., Gupta, A. K., & Sehgal, V. K. (2024). Exploiting the synergy of SARIMA and XGBoost for spatiotemporal earthquake time series forecasting. Earth Surface Processes and Landforms, 49(14), DOI: 10.1002/esp.5992.
[11]. Zhang, Y., Zhan, C. X., Huang, Q. H., & Sornette, D. (2024). Forecasting future earthquakes with deep neural networks: application to California. Geophysical Journal International, 240(1), 81-95.
[12]. Triyoso, W. (2024). Applying the Akaike Information Criterion (AIC) in earthquake spatial forecasting: a case study on probabilistic seismic hazard function (PSHF) estimation in the Sumatra subduction zone. Geomatics Neural Hazards & Risk, 15(1), Document No. 2319187.
[13]. Kohonen, T. (1990). The self-organizing map. Proceedings of the IEEE, 78(9), 1464-1480.
[14]. Xu, N., Zhu, W., Wang, R., Li, Q., Wang, Z. W., Finkelman, R. B. (2023). Application of self-organizing maps to coal elemental data. International Journal of Coal Geology, 277, Document No. 104358.
[15]. Song, Y. C., & Hirose, A. (2022). Phasor-quaternion self-organizing-map-based ground penetrating radar systems. IEEE Transaction on Geoscience and Remote Sensing, 60, Document No. 4501611.
[16]. Macshane, J., & Ahmadinia, A. (2024). Construction of trail networks based on growing self-organizing maps and public GPS data. International Journal of Knowledge-Based and Intelligent Engineering Systems, 28(2), 247-262.
[17]. Cortes, C., & Vapnik, V. N. (1995). Support-vector networks. Machine Learning, 20(3), 273-297.
[18]. Drucker, H., Burges, C. C., Kaufman, L., Smola, A. J., & Vapnik, V. N. (1997). Support Vector Regression Machines. In M. C. Mozer, Jordan, M. I., & Petsche, T. (Eds.), Advances in Neural Information Processing Systems, 9, 155-161. Cambridge, MA: MIT Press.
[19]. Vapnik, V. N. (1998). Statistical Learning Theory. New York: Wiley.
[20]. Nie, R. J., Che, J. X., Yuan, F., & Zhao, W. H. (2024). Forecasting peak electric load: robust support vector regression with smooth nonconvex ϵ-insensitive loss. Journal of Forecasting, 43(6), 1902-1917.
[21]. Lin, W. M. (2023). Generator fault diagnosis with bit-coding support vector regression algorithm. Energies, 16(8), Document No. 3582.
[22]. Huang, H. J., Wei, X. X., & Zhou, Y. Q. (2022). A binary PSO-based model selection for novel smooth twin support vector regression. International Journal of Swarm Intelligence Research, 13(1), DOI:10.4018.
[23]. Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Computational and Applied Mathematics, 20, 53-65.
[24]. Kaufman, L., & Rousseeuw, P. J. (1990). Finding Groups in Data: An Introduction to Cluster Analysis. Hoboken, NJ: Wiley-Interscience.
[25]. Weka (2024). https://ml.cms.waikato.ac.nz/index.html.
[26]. Chang, C. C., & C.-J., Lin (2011). LIBSVM : a library for support vector machines. ACM Transactions on Intelligent Systems and Technology, 2(3), 27:1-27:27.