

Artificial Intelligence for Project Management

# Using Machine Learning Technique to Predict Earthquake and PGA

王琪驛  
R06521607

## TABLE OF CONTENT

COVER.....	I
TABLE OF CONTENT.....	II
TABLE OF FIGURE.....	IV
LIST OF TABLES .....	V
<b>1 INTRODUCTION .....</b>	<b>1</b>
1.1 BACKGROUND .....	1
1.2 RESEARCH QUESTION .....	2
1.3 RESEARCH OBJECTIVE .....	2
1.4 LIMITATION .....	2
<b>2 LITERATURE REVIEW .....</b>	<b>3</b>
2.1 WHAT IS AN EARTHQUAKE EARLY WARNING SYSTEM?.....	3
2.2 EARLY WARNING SYSTEM (EEW) IN TAIWAN.....	4
2.3 MACHINE LEARNING .....	6
2.3.1 <i>Decision Tree</i> .....	6
2.3.2 <i>Support Vector Machine (SVM)</i> .....	6
2.3.3 <i>Artificial Neural Network (ANN)</i> .....	7
<b>3 METHODOLOGY.....</b>	<b>8</b>
3.1 OVERVIEW.....	8
3.2 RESEARCH METHODOLOGY.....	8
3.3 DATA RETREVING .....	9
3.4 BUILDING MODEL.....	9
3.4.1 <i>Earthquake classifier</i> .....	11
3.4.2 <i>PGA predictor</i> .....	12
3.5 MODEL VALIDATION .....	13
<b>4 RESULT .....</b>	<b>14</b>
4.1 DATA PRESENTING.....	14
4.2 EARTHQUAKE CLASSFIER .....	15
4.3 PGA PREDICTOR .....	15
4.3.1 <i>LR, SVR, and CART model</i> .....	15
4.3.2 <i>ANN model</i> .....	17
4.4 SUMMARIZE THE PERFORMANCE OF MODELS .....	18

4.5	MODEL PRSENTATION .....	19
<b>5</b>	<b>DISCUSSION .....</b>	<b>21</b>
5.1	OVERVIEW THE RESEARCH GOAL.....	21
5.2	MODEL PERFORMANCE .....	21
<b>6</b>	<b>CONCLUSION .....</b>	<b>23</b>
<b>7</b>	<b>REFERENCES.....</b>	<b>24</b>

## TABLE OF FIGURE

FIGURE 2.1 EARTHQUAKE EARLY WARNING BASICS SOURCE: CURRIE & WISE, 2016.....	3
FIGURE 2.2 DISTRIBUTION OF FREE-FIELD STRONG-MOTION STATIONS AND REAL-TIME STATINS .....	4
FIGURE 2.3 THE DISTRIBUTION OF MEASURED PGA AND PREDICTED PGA FOR THE 16 NEEWS STATIONS .....	5
FIGURE 3.1 FRAMEWORK OF RESEARCH METHODOLOGY .....	8
FIGURE 3.2 PROCESS OF TRAINING MODEL .....	9
FIGURE 4.1 SIX ROWS AT FIRST OF PREDICTIONS OF EARTHQUAKE PARAMETERS.....	14
FIGURE 4.2 3D SCATTER PLOT OF DATA NORMALIZATION.....	14
FIGURE 4.3 THE PARAMETER SETTING IN SVC MODEL.....	15
FIGURE 4.4 TEST SCORE IN SVC MODEL.....	15
FIGURE 4.5 REAL-PREDICTED LOGLOG SCATTER PLOT-CART .....	16
FIGURE 4.6 REAL-PREDICTED LOGLOG SCATTER PLOT-SVR .....	16
FIGURE 4.7 REAL-PREDICTED LOGLOG SCATTER PLOT-LR.....	16
FIGURE 4.8 10 TIMES OF ANN MODEL (10+15) REAL-PREDICTED LOG SCATTER PLOT .....	18
FIGURE 4.9 USER INTERFACE FOR PREDICTING EARTHQUAKE .....	20
FIGURE 4.10 USER INTERFACE FOR PREDICTING PGA .....	20
FIGURE 5.1 ERROR TOLERANCE RANGE .....	22

## LIST OF TABLES

TABLE 3.1 GRIDSEARCH IN SVC MODEL PROCESS.....	10
TABLE 3.2 ADJUSTABLE PARAMETERS IN DIFFERENT MODELS.....	11
TABLE 4.1 MODEL EVALUATION RESULTS .....	16
TABLE 4.2 ANN MODELS COMBINATIONS RESULTS.....	17
TABLE 4.3 INTEGRATION RESULTS OF FOUR MODELS.....	19
TABLE 5.1 INTEGRATION RESULTS OF FOUR MODELS.....	22

# 1 INTRODUCTION

## 1.1 Background

The earthquake disasters in Taiwan are mostly come from active faults. In the last century, there are nearly a dozen large magnitude earthquakes have occurred in Taiwan, including the Meishan earthquake in 1906 (ML = 7.1, 1,258 death), the Hisnchu-Taichung earthquake in 1935 (ML = 7.1, 3,276 death), and the Chi-Chi earthquake in 1999 (ML = 7.3, 2,455 death). In addition to causing casualties, the loss of economy and property cannot be underestimated. Since Taiwan is located on the western circum Pacific seismic belt which is one of the most active seismic regions in the world (Wen, Shin, Wu, Hsiao, & Wu, 2009), it needs an earthquake early warning systems.

Earthquake early warning systems can detect the location of an earthquake in a few seconds and issue a warning to the target area before the damaging waves arrive (Allen & Kanamori, 2003). These systems can reduce the fatalities, injuries, and damages caused by earthquake by alerting people so that they can stop trains, run to a safer place, take cover and many other applications (Allen, 2011). Nowadays, smartphones can be found anywhere, thus can be a great tool if can be used as an early warning systems.

In this research, we will use smartphones as a tool to make earthquake early warning system. By using the sensors in smartphones, we only need to pass software from the network operatos to users, which is relatively cheaper than building stations to detect earthquake. We will propose an algorithm to determine if an earthquake has occurred by using smartphone's accelerometers. The proposed algotithm will consist of a trigger using Short-Term Average / Long-Term Average (STA/LTA), a classifier using Artificial Neural Network (ANN), and a predictor using Support Vector Machine (SVM).

Our source of information is provided by the Taiwan Strong-Motion Instrument Program (TSMIP), as the subject of the research on the seismic warning system. We use the analytical method to perform the machine learning by using known and unprocessed parameters to find the best mode. We hope that the results of prediction can be accurate.

## **1.2 Research Question**

1. How to determine if a waveform is an earthquake?
2. How to predict the PGA of the earthquake?
3. What is the most effective model for this particular research?

## **1.3 Research Objective**

1. Finds out the way to classify waveforms.
2. Finds out the way to predict earthquake's PGA.
3. Finds out the best model to be waveforms classifier and earthquake's PGA predictor in this particular research.

## **1.4 Limitation**

1. STA/LTA algorithm as the trigger.
2. Artificial neural network.
3. Support Vector Machine.
4. Coded in Python.
5. Predicting PGA.
6. Classifying waveforms.

## 2 LITERATURE REVIEW

### 2.1 What is an earthquake early warning system?

The idea of an earthquake early warning system began in the late nineteenth century. Published in the San Francisco Bulletin in November 1868, just after the San Francisco earthquake on the Hayward fault, Dr. J. D. Cooper developed the idea of having an earthquake bell in the center of San Francisco that would ring after telegraph cables sent signals after detecting ground shaking (Allen, Gasparini, Kamigaichi, & Böse, 2009). Then in 1950, Japan started installing accelerometers to ensure that its train function properly, this led to the development of the early warning system. This made Japan to have the first operational earthquake early warning system in the world (Wenzel, Baur, Fiedrich, Ionescu, & Oncescu, 2001). Figure 2.1 shows the basics of an earthquake early warning system. Mainly, a system contains four principal components (Asgary, Levy, & Mehregan, 2007):

1. A monitoring system composed of sensors
2. A real-time communication link that transmits data from the sensors to a computer
3. A processing facility that converts data into information
4. A system that issues and communicates the warning

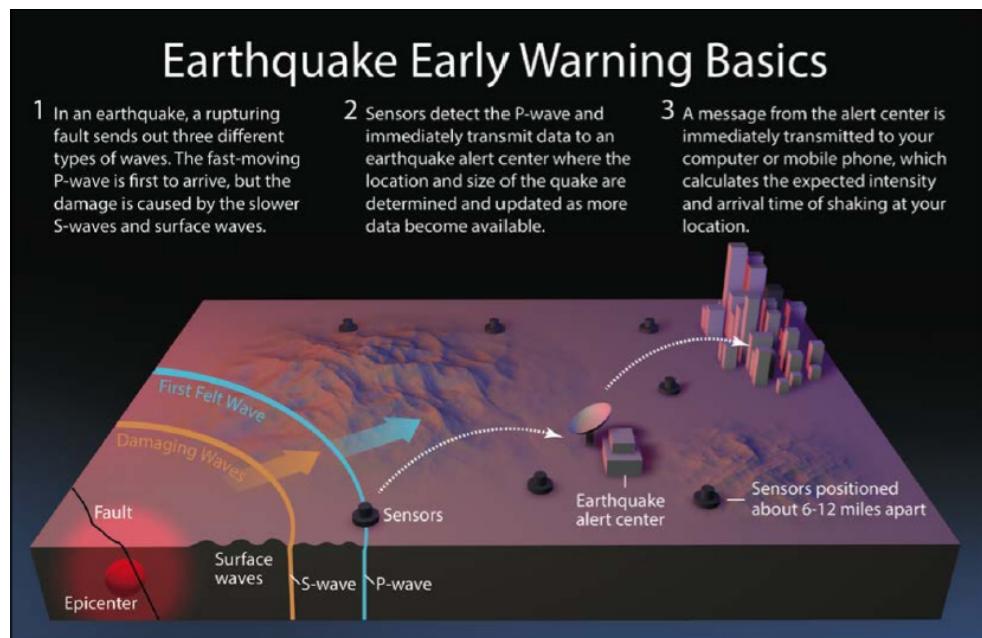


Figure 2.1 Earthquake Early Warning Basics

Source: Currie & Wise, 2016

## 2.2 Early warning system (EEW) in Taiwan

In Taiwan there are two types of EEW being used, they are regional warning and onsite warning. Regional warning provide more reliable EEW messages but requires longer processing time while onsite warning used to speed up EEW processing (Wu, Hsiao, Lee, Teng, & Shin, 2007).

There are 688 free-field strong-motion stations constructed in the TSMIP, currently 109 are telemetered for real-time monitoring as shown in Figure 2.2 (Wen, Shin, Wu, Hsiao, & Wu, 2009). The real-time stations are connected to Taipei Headquarter via dedicated telephone lines, which become the base for the EEW system. The EEW system will provide warnings within a few seconds to several dozen seconds of lead-time before destructive S waves arrive.

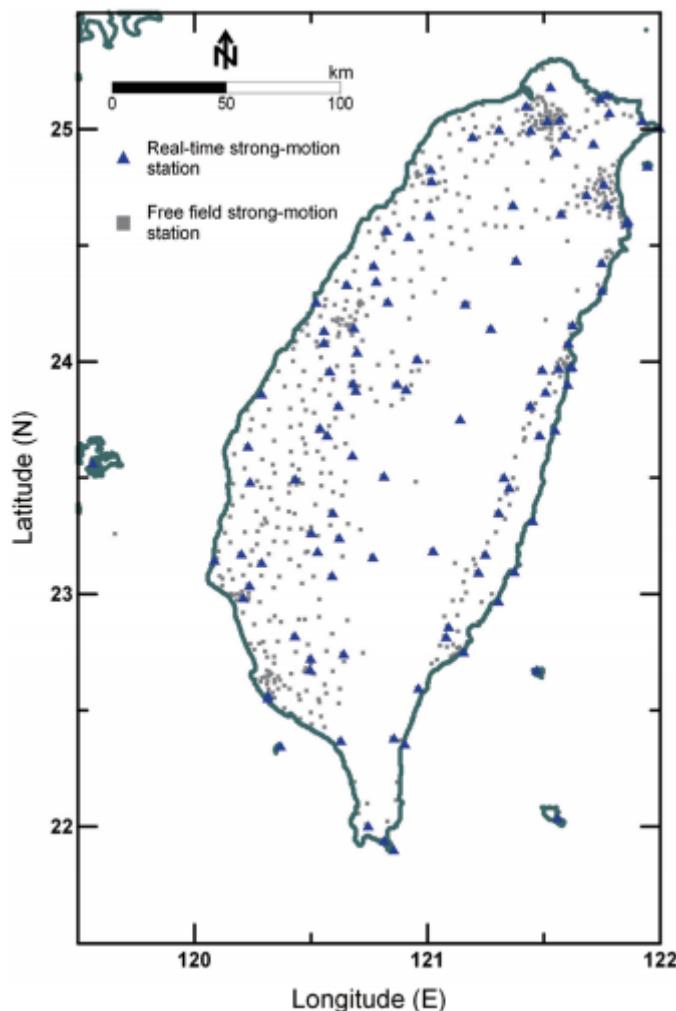


Figure 2.2 Distribution of free-field strong-motion stations and real-time statins

source: Wen, Shin, Wu, Hsiao, & Wu, 2009

The research as reported in (Hsu, et al., 2016) indicates that the accuracy of the predicted PGA during the Meinong earthquake is considered. The comparison of the predicted PGA and the measured PGA of all the 16 stations is plotted in Figure 2.3. It can be observed that the predicted PGA corresponds to the measured PGA very well in a logarithm scale. The accuracy of predicted intensity was 93.75% and as for accuracy of issued alarms was 100%.

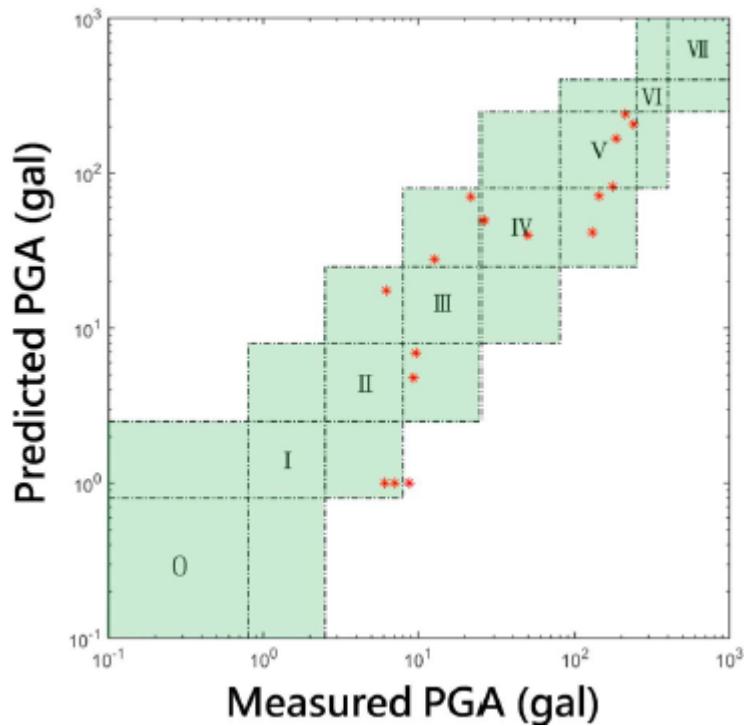


Figure 2.3 The distribution of measured PGA and predicted PGA for the 16 NEEWS stations

Source: Hsu, et al., 2016

## **2.3 Machine learning**

Machine learning is a branch of artificial intelligence that allows computers to learn naturally like humans. Analyze data from detectors or databases and extract useful rules and features to improve the algorithm. Common machine learning algorithms include support vector machines, neural networks, decision trees, random forests, etc. This study will apply these methods to explore the model, the following introduction

### **2.3.1 Decision Tree**

Decision tree is a predictive model; each decision tree is a tree structure. Classification of target data by attribute by its branch. Each decision tree can be tested for data in terms of segmentation of the data set and can recursively trim the tree. When it is no longer possible to split or a single class can be applied to a branch, the recursion process is complete. The advantage of decision trees is that they are easy to understand and implement. After explaining, people have the ability to understand the meaning expressed by the decision tree. The shortcoming is that the single decision tree often has an over-fitting phenomenon.

### **2.3.2 Support Vector Machine (SVM)**

Vladimir N. Vapnik and Alexey Y. Chervonenkis invented the original SVM algorithm in 1963. In 1992, Bernhard E. Boser et al. proposed an improved method that can be expanded into a nonlinear classifier. SVM is often used for data classification. It is assumed that the data points in the data set belong to one of the two classes, and the goal is to find out which class the new data points will be in. The SVM will look for a hyperplane that can separate the data points into two categories. Because SVM is very good at generalization, SVM can get much better results than other classifiers when the number of samples is small. On the contrary, when the number of samples is large, it will consume a lot of resources and time, and it is not suitable for solving multi-class problem.

### **2.3.3 Artificial Neural Network (ANN)**

Artificial Neural Network is a mathematical model or computational model that mimics the structure and function of a biological neural network. The neural network routes multiple neurons. After the feature vectors X1, X2, and X3 are input into the model, they are given corresponding weights, respectively W1, W2, and W3. If the inner product result of the feature vector and the weight is greater than a certain threshold the output result is 1 otherwise if it is less than a certain threshold the output will be 0. Like a real neural network, it has the ability to learn from mistakes. First let the data signal pass through the network, output the result, calculate the error with the real situation, then pass the error signal back, and adjust the weight to each neuron in the correct direction. Therefore, after many iterations, the accuracy of the model can be greatly improved. Although it is a powerful predictive model, the disadvantage is that it cannot explain its own reasoning base and process and it requires a lot of parameters to tune.

### 3 METHODOLOGY

#### 3.1 Overview

This project used STA/LTA algorithm as a trigger and used a classifier model to classify if a waveform is an earthquake or not as well as predicting the PGA of an earthquake. This two-step approach was implemented so that the power requirements were not increased, since the STA/LTA method is a simple and low cost computation method.

#### 3.2 Research methodology

First, the data was collected. Since machine-learning technique was still related to data mining, the availability of needed data was important. The recorded earthquake data was collected from Taiwan Strong Motion Instrumentation Program (TSMIP). After data collection, the raw collected data was processed in pre-processing stage. The pre-processed data was the complete and final dataset that was used in this project.

The next step was building the machine-learning algorithm. SVM Classifier was used in this project to determine if a motion was an earthquake or not and SVR, LR, CART, and ANN was used for predicting PGA. The dataset in the previous step was used for training, validating, and testing. The whole project framework is summarized in Figure 3.1.

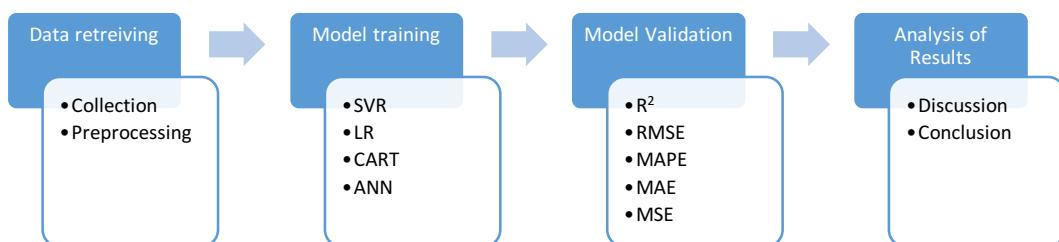


Figure 3.1 Framework of Research Methodology

### 3.3 Data retrieving

This project used three types of data for training, validating, and testing purposes. They were normal human activity data, smartphone recorded earthquake data, and TSMIP's earthquake data.

The data consisted of recorded earthquake from 1992-2006 that were collected from TSMIP. The data contained wave arrival time, sample rate, acceleration as the input parameters. The smartphone recorded earthquake data will be obtained by placing smartphones on shaking table that was simulating past earthquakes.

The earthquake data from TSMIP needed to be modified to replicate waveforms recorded from a smartphone. To do this, the sampling rate was converted to 50Hz, and then the smartphone noises were added.

### 3.4 Building model

The algorithm of the proposed methods was developed in Python programming language. The algorithm was split into two parts, trigger part and Earthquake classifier & PGA predictor. The training follows Figure 3.2:



Figure 3.2 Process of training model

First, the required kits were installed and loaded and then the data was loaded. The dataset was split multiple times with the proportion of 70% of the data for training and 30% for testing using cross-validation tests. At this step, the training of the model was started.

In this study, we used gridsearch to optimize the parameters of SVM, LR, and CART. First, the searching range was set & changed until the best parameter was found. The important point to note was that if the found parameter was in the searching range, then it could be considered as the best parameter. However, if the best parameter appeared at the boundary of the searching range, then searching range needs to be changed to do another GridSearch. Table 3.1 demonstrates the

SVC algorithm GridSearch process, three parameter ranges adjusted until the found parameters were within the searching ranges.

Table 3.1 GridSearch in SVC Model Process.

Times	GridSearch result
1	<pre>from sklearn.model_selection import GridSearchCV param_grid = {'C': [0.09,0.1,0.2,0.3,0.4],               'gamma': [0.09,0.01,0.02,0.3,0.4]} grid = GridSearchCV(model,param_grid)  grid.fit(Xtrain, ytrain) print(grid.best_params_)  {'C': 0.4, 'gamma': 0.02}</pre>
2	<pre>from sklearn.model_selection import GridSearchCV param_grid = {'C': [0.4,0.5,0.6],               'gamma': [0.01,0.02,0.3]} grid = GridSearchCV(model,param_grid)  grid.fit(Xtrain, ytrain) print(grid.best_params_)  {'C': 0.6, 'gamma': 0.02}  !true</pre>
3	<pre>param_grid = {'C': [0.7,0.8,0.9],               'gamma': [0.01,0.02,0.3]} grid = GridSearchCV(model,param_grid)  grid.fit(Xtrain, ytrain) print(grid.best_params_)  {'C': 0.9, 'gamma': 0.02}  !True</pre>
4	<pre>from sklearn.model_selection import GridSearchCV param_grid = {'C': [0.9,1,1.1],               'gamma': [0.01,0.02,0.3]} grid = GridSearchCV(model,param_grid)  grid.fit(Xtrain, ytrain) print(grid.best_params_)  {'C': 1, 'gamma': 0.02}  !True</pre>

Different model has different parameters. Table 3.2 summarizes the parameters of the three models:

Table 3.2 Adjustable Parameters in Different Models

MODEL	SVM	LR	CART
PARAMETER	<ul style="list-style-type: none"> <li>• C</li> <li>• Gamma</li> </ul>	<ul style="list-style-type: none"> <li>• fit_intercept</li> <li>• normalize</li> </ul>	<ul style="list-style-type: none"> <li>• criterion</li> <li>• splitter</li> <li>• min_samples_split</li> <li>• min_samples_leaf</li> </ul>

The ANN model was a more complex compared to the other models, in which the adjustment that needed to be included are the combination of the neurons and the number neron layers. Since this is a hyperparameter, the most suitable units and parameters combinations cannot be found using gridSearch but using literature and experiment instead. First, a single layer of 5 nodes, a single layer of 10 nodes, a single layer of 15 nodes, a single layer of 20 nodes, and two layers of 5 nodes, two layers of 5 nodes plus 10 nodes, two layers with 10 nodes each, two layers of 10 nodes plus 15 nodes model were used for training. Moreover, the data was normalized first before inputed to the ANN model.

### 3.4.1 Earthquake classifier

Simple short-term average/long-term average (STA/LTA) method was used for the trigger part. The trigger would be activated if the STA/LTA exceed a threshold. After the STA/LTA triggered, 2-sec data windows with a 1-sec step was used to calculate the three key features up to 10 sec after the STA/LTA trigger.

The three features used are Interquartile range (IQR) between the 25<sup>th</sup> and 75<sup>th</sup> percentile of the acceleration vector sum, the zero crossing rate from the component with the highest value (ZC), and the cumulative absolute velocity (CAV) of acceleration vector sum. IQR is an amplitude parameter that shows the middle 50% range of amplitude of the movement. ZC is a simple frequency measure. These three features will be first scaled to a range of 0 to 1. CAV is a cumulative measure of amplitude on the three components in the time window and is determined as follows:

$$CAV = \int_0^2 |a(t)| dt \quad (1)$$

where  $a(t)$  is the vector sum of the 3 components acceleration.

### 3.4.2 PGA predictor

This study intends to find the best model to predict PGA. According to a paper published by (Zollo, et al., 2011) a paper published by (許丁友, 2017). Six kinds of initial wave characteristics were used as input data for the support vector regression model: acceleration extremum (Pa) of vertical seismic history over  $t_p$  seconds after P wave arrival, velocity extreme value (Pv), displacement extreme value Pd, main period c, cumulative absolute velocity (CAV) and velocity square integral (IV2). The decision of the arrival time of the arrival wave is determined by the short-term average and long-term average difference method (short-term average/long-term average, STA/LTA). The mathematical formulas for each of the characteristics of the first wave summarized as follows:

$$\tau_c = 2\pi/\sqrt{r}, \quad \text{where } r = \int_0^{t_p} \dot{u}^2(t) dt / \int_0^{t_p} u^2(t) dt \quad (2)$$

$$CAV = \int_0^{t_p} |\ddot{u}(t)| dt \quad (3)$$

$$IV2 = \int_0^{t_p} \dot{u}^2(t) dt \quad (4)$$

In the formula,  $u(t)$  is the surface displacement duration. Due to the acceleration of the original seismic record system, the above-mentioned characteristics of the first wave was calculated. It is necessary to integrate it into the velocity and displacement. In addition, the high-pass filter of Butterworth is used to remove the low-frequency offset after integration, and the number of stages and the angle of the filter. The frequencies are 2 and 0.075 Hz, respectively.

In this study, the support vector regression model is constructed by using “representing seismic data”. The input data of the support vector regression model are the characteristics of the above six earthquakes, and were normalized to values between -1 and 1, and the output data is the maximum value of the absolute value of the surface acceleration of the seismic wave in three directions.

In order to validate and train the support vector regression model, representative seismic data was divided into 10 sub-samples in a cross-validation manner, which is a single sub-sample retained as the data of the verification model, and the other 9 samples were used to training. This was repeated 10 times, each subsample was

used for verification, and the average error of 10 results was finally taken as the final results.

The undetermined parameters of the support vector regression model include  $\sigma$ ,  $\varepsilon$ , and  $C$ . Since  $\varepsilon$  directly affecting the number of support vectors, the smaller  $\varepsilon$  is selected in this study, and the number of support vectors is larger, so that the model could cover earthquakes regardless of size. The situation for the remaining parameters  $\sigma$  and  $C$ , the search method was based on the grid point. The search range was first considered to be larger but the grid spacing is wider. Then, the range and grid spacing are gradually narrowed down, and the grid with the smallest cross-validation mean square error is found to be the best parameters.

### 3.5 Model validation

Certain validation technique must be used to determine whether the model can produce a reliable and acceptable results. Because of that reason, this project will be validated using the human activity data and earthquake data from shake table tests. The classification model mainly uses accuracy as a criterion for training the strengths and weaknesses of the model. The Regression model used RMSE (Root mean square error) and MAPE (Mean absolute percentage error).

## 4 RESULT

### 4.1 Data presenting

Using the parameters extracted from Raw data to identify the earthquake, the data was first labeled. The TSMIP data had an earthquake label of "1" and anusus data of '-1'. Train the model using the machine learning algorithm and find the model with the highest accuracy to further adjust the parameters.

The data that had been preprocessed as shown below. The parameters for predicting whether the earthquake was used are limited to the range of [0, 1] by min-max normalization. The normalized data is shown in the Figure 4.1 and the parameters are scattered as shown in the Figure 4.2.

head(train_data)				
	ZC <dbl>	IQR <dbl>	CAV <dbl>	labels <fctr>
1	0.6282051	0.004771575	0.005029633	1
2	0.4487179	0.006134882	0.005325221	1
3	0.5897436	0.002044961	0.004023799	1
4	0.7051282	0.010906457	0.012975304	1
5	0.7051282	0.017722992	0.033041946	1
6	0.6923077	0.010224803	0.010160266	1

Figure 4.1 Six Rows at First of Predictions of Earthquake Parameters

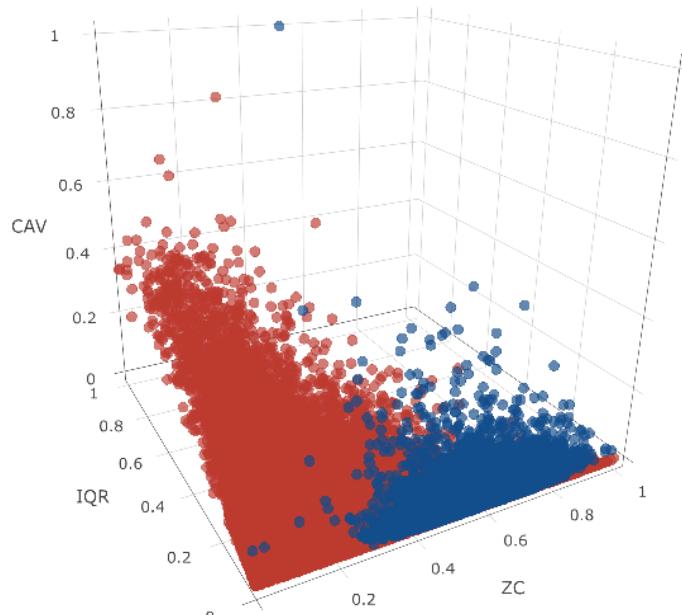


Figure 4.2 3D Scatter Plot of Data Normalization

## 4.2 Earthquake classifier

From the three-dimensional distribution of the data shown, it could be judged that the data could be segmented by an irregular hyperplane. Based on such data characteristics, first the Support Vector Classifier was used for training. After using Gridsearch, the parameters in the model was set as shown in Figure 4.3. The final testing accuracy was 99%, which should be a very accurate model. The result is shown in Figure 4.4.

```
model = SVC(kernel='rbf', class_weight='balanced', C=1, gamma=0.02)
```

Figure 4.3 The parameter setting in SVC model

	precision	recall	f1-score	support
-1	0.99	1.00	0.99	27795
1	1.00	0.99	0.99	27324
micro avg	0.99	0.99	0.99	55119
macro avg	0.99	0.99	0.99	55119
weighted avg	0.99	0.99	0.99	55119

Figure 4.4 Test score in SVC model

## 4.3 PGA predictor

PGA predictions was more difficult than earthquake predictions. First, LR, SVR, and CART are used to do GridSearch to find the best combination of parameters. In addition, the ANN model was also used. Eight combinations of hyperparameters were constructed for training and the results were checked.

### 4.3.1 LR, SVR, and CART model

First, the prediction results using LR, SVR, and CART are shown in Table 4.1 from the residual results of the predicted values, it could be seen that the SVR is better. However, if the log distribution from the predicted-actual value is observed as shown in Figure 4.5, Figure 4.6, and Figure 4.7. It can be observed that the SVR prediction results underestimate the original PGA, which is less acceptable in earthquake prediction. Therefore, we have adopted a more complex ANN model.

Table 4.1 Model Evaluation Results

	<i>R sqaure</i>	MAPE	MAE	MSE	RMSE
<i>LR</i>	-65.005	73.8	28.335	93537.4	305.839
<i>SVR</i>	0.259	62.683	7.694	793.227	28.164
<i>CART</i>	0.172	98.777	10.355	886.403	29.773

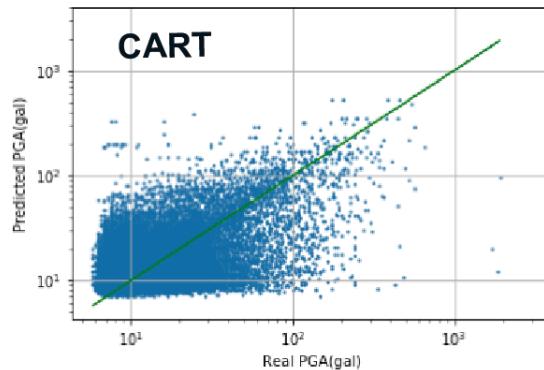


Figure 4.5 Real-Predicted loglog scatter plot-CART

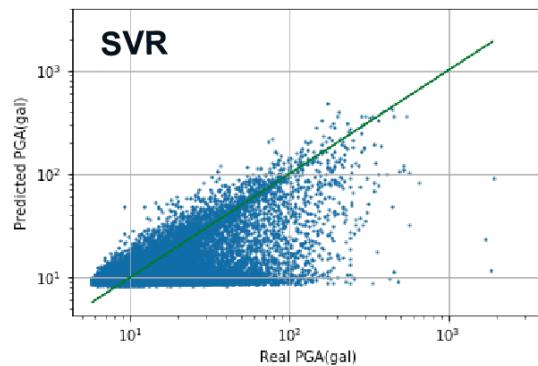


Figure 4.6 Real-Predicted loglog scatter plot-SVR

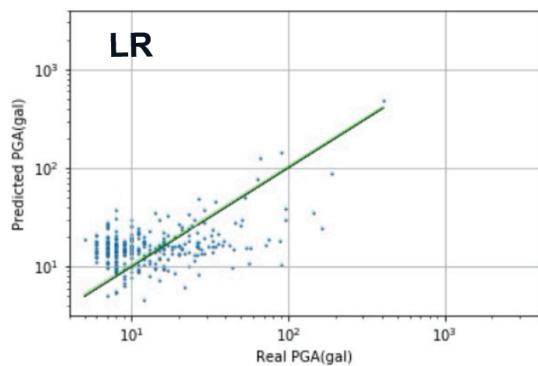


Figure 4.7 Real-Predicted loglog scatter plot-LR

### 4.3.2 ANN model

The training results are presented in 8 models we decided to be as shown in Table 4.2. We choose a model with two layers of 10+15 nodes with the smallest RMSE. After changing the training data to random, we reseparated the 70% of the train data and the 30% of the test data, and trained for ten times.

Table 4.2 ANN Models Combinations Results

Layer	Units	RMSE
1	5	30.59
	10	30.298
	15	30.229
	20	30.217
2	5+5	30.76
	5+10	30.371
	10+10	30.221
	10+15	<b>30.129</b>

Finally, ten models were trained using 10+15 and the results of each prediction are presented in Figure 4.8. The horizontal axis in all figures is the real PGA and the vertical axis is the predicted PGA. The lowest model of RMSE is 25.91, but most of the data in the distribution of predicted results are under the slash, which means that most of the prediction results are lower than the actual value. If converted into earthquake, it had the tendency to be slightly lower than the actual earthquake, instead of going higher than the actual earthquake. Comparing the model with RMSE of 25.91 with the model with RMSE of 30.0412 was less distributed under the oblique line, so it is chosen as the most appropriate model.

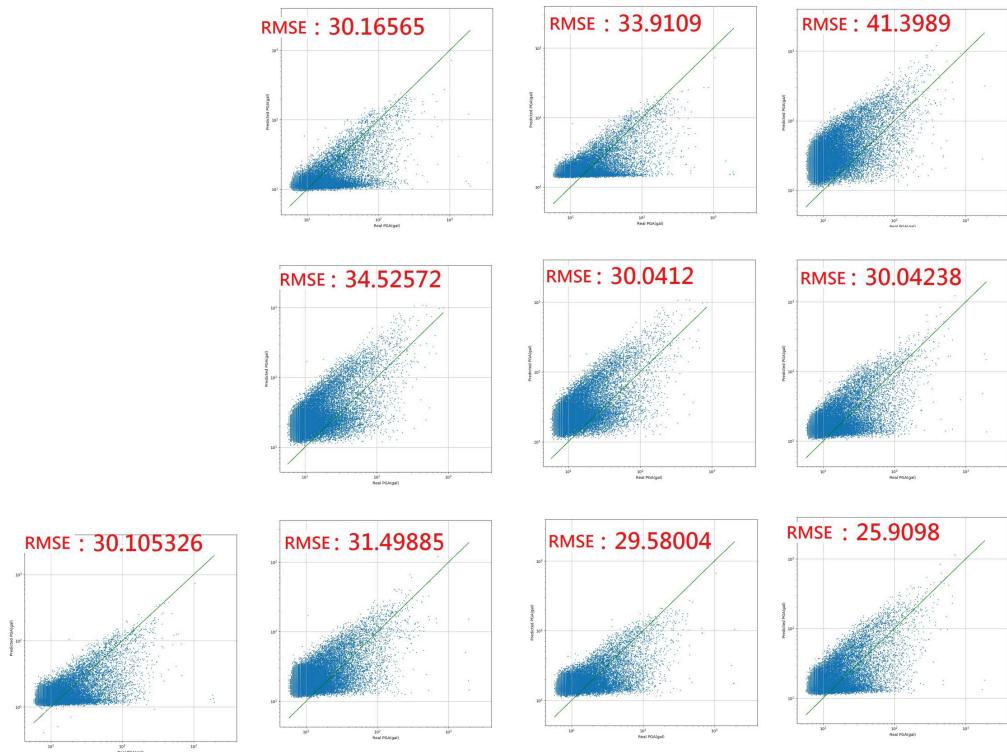
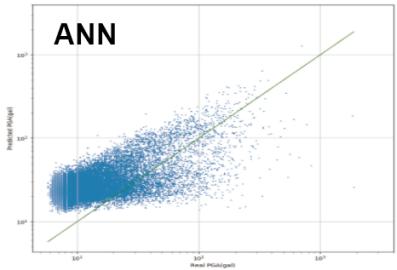
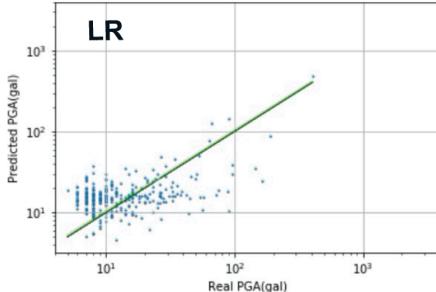
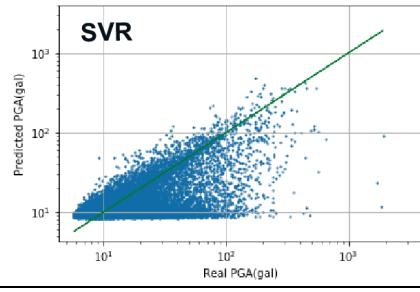
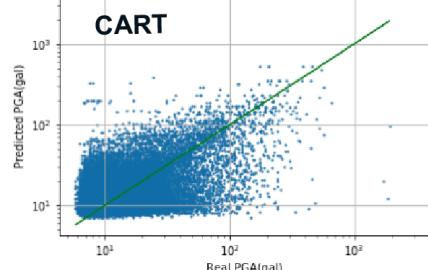


Figure 4.8 10 Times of ANN Model (10+15) Real-Predicted Log Scatter Plot

#### 4.4 Summarize the performance of models

From summarizing all the trained models, the accuracy of classifying earthquake is up to 99% with the best model. In the PGA prediction model, the results predicted by the four algorithms are presented in Table 4.3. Although the RMSE of the ANN was not lower than the other three models, the MAPE was significantly less than the other three models. It can also be seen from the Table 4.3 that the data distribution of ANN is relatively average, located near the oblique line, and less data is below the oblique line. In summary, the ANN model is the most appropriate one for predicting PGA.

Table 4.3 Integration Results of Four Models.

model	Result	Real-Predict comparison
ANN	<b>R sqaure:</b> 0.260951697284 <b>MAPE:</b> 48.9160105293 <b>MAE:</b> 8.98543173097 <b>MSE:</b> 906.330676624 <b>RMSE:</b> 30.105326383	 <p><b>ANN</b></p>
LR	<b>R sqaure:</b> -65.005721879 <b>MAPE:</b> 73.7998494223 <b>MAE:</b> 28.3345861527 <b>MSE:</b> 93537.3949343 <b>RMSE:</b> 305.838838172	 <p><b>LR</b></p>
SVR	<b>R sqaure:</b> 0.258624715042 <b>MAPE:</b> 62.6834597667 <b>MAE:</b> 7.69434428767 <b>MSE:</b> 793.227359357 <b>RMSE:</b> 28.1642922751	 <p><b>SVR</b></p>
CART	<b>R sqaure:</b> 0.171540010021 <b>MAPE:</b> 98.7773002125 <b>MAE:</b> 10.3554109709 <b>MSE:</b> 886.402802355 <b>RMSE:</b> 29.7725175683	 <p><b>CART</b></p>

#### 4.5 Model prsentation

In order to allow users to better use these predictive models, we developed two user interfaces using tkinter. Just enter the predicted parameters to know the magnitude of the earthquake and whether it is an earthquake, as shown in Figures 4.9 and 4.10.



Figure 4.9 User interface for predicting earthquake



Figure 4.10 User interface for predicting PGA

## 5 DISCUSSION

This section revealed the answers three of the research questions: (i) Q1: How to determine if a waveform is an earthquake? (ii) Q2: How to predict the PGA of the earthquake? (iii) Q3: What is the most effective model for this particular research? It also describes the future direction of the study, including how to better train the direction of the model and the practical application of the model.

### 5.1 Overview the research goal

#### **Q1: How to determine if a waveform is an earthquake?**

Simple short-term average/long-term average (STA/LTA) method is used for the trigger part. The trigger is activated if the STA/LTA exceed a threshold. After the trigger is activated, an SVC model is used to further classify if it is a true earthquake.

#### **Q2: How to predict the PGA of the earthquake?**

Some model (ANN, LR, SVR, and CART) were made in this research and it was found that the PGA can be predicted using regression model.

#### **Q3: What is the most effective model for this particular research?**

The earthquake classifier uses SVC, and the test accuracy reaches 99%; the most suitable model for predicting PGA is ANN, and RMSE is 30.0412.

### 5.2 Model performance

Although the residual of the last selected model is still high, according to the grading index of the research (Hsu, et al., 2016), after the predicted result is converted into the level, the difference between the predicted value and the actual value is +/- 1 level is an acceptable accuracy range, and the predicted result conversion is shown in Table 5.1.

Table 5.1 Integration Results of Four Models.

realPGA	Predicted PGA	realPGA Level	Predicted PGA Level	Is there more than the range?
40.219	100.5989	4	5	0
8.1396	21.98742	3	3	0
9.0972	19.18618	3	3	0
9.0972	22.66018	3	3	0
:	:	:	:	:

According to Table 5.1, the comparison results are shown in Figure 5.1. From this figure, it can be seen that there are few points beyond the red frame range, and the accuracy rate is 91.55%.

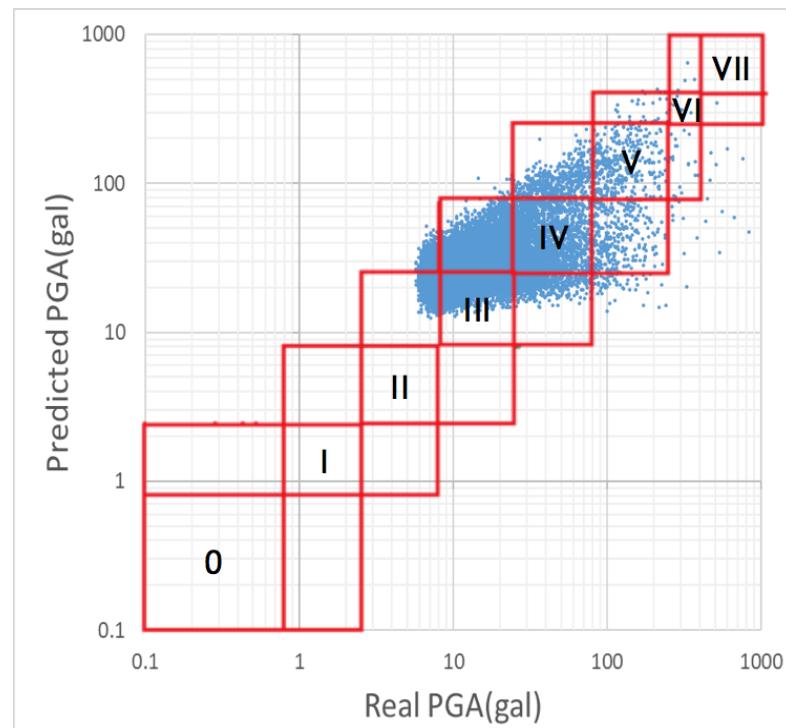


Figure 5.1 Error tolerance range

## 6 CONCLUSION

In the future, the proposed model can be used as EEW systems; this can be useful in earthquake active regions that do not have any EEW systems such as Indonesia or as an on-site warning system in place where EEW system has existed before like Taiwan. Because the proposed model has been designed and tested using smartphones (which is available in most places) it gathers data without the need of installing new traditional EEW systems, which can reduce the budget & time needed to make EEW systems.

In this research, the model is divided into two parts earthquake classifier and PGA predictor. The earthquake classifier consists of STA/LTA & SVC classifier. The STA/LTA is used in order to keep smartphone's power consumption low. After the algorithm is triggered, SVC is used to further classify if a wave is an earthquake and the result's accuracy can get as high as 99%. Whereas the best model to predict PGA is using ANN with RMSE of 30.0412, converted to a seismic indicator, accuracy has been to 91.55%.

In the training results, the PGA prediction part can be found that the model is not very good when predicting earthquakes with large magnitude. The reason may be that the earthquakes with large magnitude are less likely to occur, so there are less earthquake with large magnitude data. Although this is a regression task, the PGA itself has a seismic grading, so it can be called a data imbalance problem. In the future, in order to get better prediction results, the training should be based on the PGA grading indicators. Data from major earthquakes can be amplified by Data augmentation. Then, it should be possible to have better results.

The model trained in this study is a model for predicting earthquakes. The researchers in related fields can optimize and import the earthquake warning system, which has certain auxiliary force for Taiwan earthquake prevention system.

## 7 REFERENCES

- Allen, R. M. (2011). Seconds Before the Big One: Progress in Earthquake Alarms. *Scientific American*, 74-79.
- Allen, R. M., & Kanamori, H. (2003). The Potential for Earthquake Early Warning in Southern California. *Science* 300, 786-789.
- Allen, R. M., Gasparini, P., Kamigaichi, O., & Böse, M. (2009). The Status of Earthquake Early Warning Around the World: An Introductory Overview. *Seismological Research Letters*, 682-693.
- Asgary, A., Levy, J., & Mehregan, N. (2007). Estimating Willingness to Pay for a Hypothetical Earthquake Early Warning Systems. *Evinronmental Hazards*, 312-320.
- Currie, C. P., & Wise, D. J. (2016). *Additional Actions Needed to Identify and Mitigate Risks to Federal Buildings and Implement an Early Warning System*. Washington, D.C.: United States Government Accountability Office.
- Hsu, T.-Y., Wang, H.-H., Lin, P.-Y., Lin, C.-M., Kuo, C.-H., & Wen, K.-L. (2016). Performance of the NCREE's on-site warning system during the 5 February 2016 Mw 6.53 Meinong earthquake. *Geophysical Research Letters*, 8954-8959.
- Kong, Q., Allen, R., Schreier, L., & Kwon, Y.-W. (2016). MyShake: A smartphone seismic network for earthquake early warning. *Science Advances*, 70-79.
- Wen, K.-L., Shin, T.-C., Wu, Y.-M., Hsiao, N.-C., & Wu, B.-R. (2009). Earthquake Early Warning Technology Progress in Taiwan. *Journal of Disaster Research*, 202-210.
- Wenzel, F., Baur, M., Fiedrich, F., Ionescu, C., & Onicescu, M. C. (2001). Potential of Earthquake Early Warning Systems. *Natural Hazards*, 407-416.
- Wu, Y., Hsiao, N., Lee, W., Teng, T., & Shin, T. (2007). State of the art and progress in the earthquake early warning system in Taiwan. *Earthquake Early Warning Systems*, 283-291.

- Zollo, A., Iannaccone, G., Convertito, V., Elia, L., Iervolino, I., Lancieri, M., . . .  
Gasparini, P. (2011). Earthquake Early Warning System in Southern Italy.  
*Geophysical Research Letters*, 2395-2420.
- 許丁友. (2017). 現地型強震預警技術與應用. 土木水利, 68-75.