Soil Dynamics and Earthquake Engineering 49 (2013) 210–217

Contents lists available at SciVerse ScienceDirect

Soil Dynamics and Earthquake Engineering

journal homepage: www.elsevier.com/locate/soildyn

Rapid on-site peak ground acceleration estimation based on support vector regression and P-wave features in Taiwan

Ting-Yu Hsu [a,*], Shieh-Kung Huang [a], Yu-Weng Chang [a], Chun-Hsiang Kuo [a], Che-Min Lin [a], Tao-Ming Chang [a], Kuo-Liang Wen [a,b], Chin-Hsiung Loh [a,c]

[a] National Center for Research on Earthquake Engineering, Taipei 106, Taiwan, ROC
[b] Department of Earth Sciences and Institute of Geophysics, National Central University, Jhongli 32054, Taiwan, ROC
[c] Department of Civil Engineering, National Taiwan University, Taipei 106, Taiwan, ROC

ARTICLE INFO

Article history:
Received 13 September 2012
Received in revised form
21 January 2013
Accepted 3 March 2013
Available online 27 March 2013

ABSTRACT

This study extracted some P-wave features from the first few seconds of vertical ground acceleration of a single station. These features include the predominant period, peak acceleration amplitude, peak velocity amplitude, peak displacement amplitude, cumulative absolute velocity and integral of the squared velocity. The support vector regression was employed to establish a regression model which can predict the peak ground acceleration according to these features. Some representative earthquake records of the Taiwan Strong Motion Instrumentation Program from 1992 to 2006 were used to train and validate the support vector regression model. Then the constructed model was tested using the whole earthquake records of the same period as well as the 2010 Kaohsiung earthquake with 6.4 M_L. The effects on the performance of the regression models using different P-wave features and different length of time window to extract these features are studied. The results illustrated that, if the first 3 s of the vertical ground acceleration was used, the standard deviation of the predicted peak ground acceleration error of the whole tested 15-years earthquake records is 20.89 gal.The length of time window could be shortened, e.g. 1 s, and the prediction error is slightly sacrificed, in order to prolong the lead-time before destructive S-waves reaches.

1. Introduction

In the last two decades, earthquake early warning (EEW) techniques have been emerged crediting to the advances in digital seismology, communications, automatic processing and the algorithms for rapid estimation of earthquake parameters [13]. Based on the requirement of information for the algorithms to estimate earthquake parameters, earthquake early warning techniques can be divided into two groups: regional warning and on-site warning. Generally, since the regional warning leverages the information of several stations next to the epicenter, the accuracy on earthquake parameters estimation of regional warning is usually higher than the on-site warning. However, for the region close to the epicenter, where seismic intensity is usually much higher than the region outside, the lead-time before destructive wave arrives provided by the regional warning can be null. On the other hand, on-site warning can provide more lead-time at the region close to an epicenter since only the seismic information on the target site is required. Therefore, raise of the accuracy and lead-time of on-site warning is one of the key points to improve the effect of earthquake early warning techniques.

On-site warning issues an alarm in a few seconds after triggering based on the initial P-wave motion at a single station. Nakamura [9] developed an earthquake early warning system named UrEDAS which predicted potential damage according to the estimation of magnitude and location based on the calculated predominant frequency, back azimuth, vertical-to-horizontal ratio and amplitude level. Nakamura [10] further developed Compact UrEDAS which simply estimated potential damage based on the calculation of destructive intensity defined as the logarithm of absolute value of the inner product of acceleration and velocity. Odaka et al. [12] tried another approach to estimate magnitude and epicentral distance based on the P-wave amplitude and the fitting parameter of the waveform envelope. Kanamori [6] proposed another predominant frequency of P-wave which is similar to the one developed by Nakamura [9] to estimate the magnitude. The combination of both the predominant frequency and peak ground displacement of P-wave measured on the vertical direction is also proposed to recognize damaging earthquakes. Böse et al. [2] estimated the magnitude, epicenter distance and peak ground velocity (PGV) based on the three-component waveforms of acceleration, velocity and

* Corresponding author. Tel.: +886 2 6630 0863; fax: +886 2 6630 0858.
E-mail address: dysheu@ncree.narl.org.tw (T.-Y. Hsu).

0267-7261/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.soildyn.2013.03.001

displacement. In summary, in the above-mentioned literature, the on-site earthquake early warning is issued according to the estimated magnitude, location, PGV and/or vulnerability of the earthquake.

However, the peak ground acceleration (PGA) is also an representative parameter for earthquake early warning. After an earthquake, the seismic intensity scale measured at each station can be calculated according to the measured PGA. For example, in Taiwan, the Central Weather Bureau calculate the seismic intensity scale based on the equation with the form $I = a\log(PGA) + b$. Similarly, the Japan Meteorological Agency also reports the earthquake intensity level based on the ground acceleration. The United States Geological Survey (USGS) can also report the Modified Mercalli Intensity (MMI) scale at each station for smaller events based on the measured PGA using the relationship developed by Worden et al. [16]. Besides, for practical concern, PGA threshold is one of the main criteria for earthquake emergency reaction in many applications. For instance, the high-speed train system uses PGA as the main criterion for emergency stop. The nuclear power plant also uses PGA as the criterion for emergency shutdown. Therefore, in this study, PGA is chosen as the target to be estimated. The P-wave features including predominant frequency, peak ground acceleration, peak ground velocity, peak ground displacement, cumulative absolute velocity and integration of the squared velocity will be extracted from the initial P-wave motion of the vertical component at a single station. The support vector regression (SVR) is employed to establish a regression model to predict the PGA according to these P-wave features. In order to gain more lead-time of the on-site warning, the effect of the length of the initial P-wave motion on the performance of the SVR model is also studied.

In addition, in the literature, the database used to establish and to validate an empirical regression model between the initial P-wave motion and the final peak ground motion is usually limited to some selected records of some representative events. Sometimes the epicenter distance is also restricted. As a result, the applicability of the established empirical regression model could be restricted to only the selected events. Furthermore, since the established empirical regression model is never validated through the earthquake events of an entire period, e.g. 10 years, the estimated error in practice is never revealed. Therefore, in this study, the earthquake records between 1992 and 2006 of the Taiwan Strong Motion Instrumentation Program (TSMIP, [7]) will be used to test the proposed approach. The database includes the catastrophic Chi-Chi earthquake and several damaging earthquake events as well as more than ten thousands of non-damaging earthquake records.

2. Methodology

2.1. P-wave features

The on-site earthquake early warning (EEW) technique takes advantage of the different velocity of propagation of P- and S-waves. In other words, the expected ground shaking dominated by S-wave could be estimated based on the recorded early informative P-wave of a single station. This is usually accomplished through the empirical regressions between the P-wave features extracted from the measurements of the first few seconds and the final earthquake intensity at the same site. Satriano et al. [13] reviewed the concepts, methods and physical backgrounds of EEW. The P-wave features used to estimate the final earthquake size were also summarized in the same paper. These P-wave features includes the peak measurement, predominant period and integral quantities. The P-wave features used to estimate final earthquake intensity in this study are briefly introduced in the following paragraphs.

Firstly, the peak measurement of acceleration, velocity and displacement of the vertical direction after t_p seconds of P-wave arrival are considered. These features are denoted as P_a, P_v and P_d respectively and can be calculated straightforwardly. These parameters were found correlated to the PGA, PGV and PGD of the entire measured time history of ground motion at the same station [17].

Next, the effective predominant period proposed by Kanamori [6] is employed. This parameter is correlated to the earthquake magnitude and can be calculated as

$$T_c = 2\pi/\sqrt{r}, \text{ where } r = \int_0^{t_p} \dot{u}^2(t)dt \Big/ \int_0^{t_p} u^2(t)dt \tag{1}$$

where $u(t)$ and $\dot{u}(t)$ are the vertical component of displacement and velocity time history of ground motion after P-wave arrival, respectively.

Finally, two more integral quantities are utilized, i.e. the cumulative absolute velocity (CAV) and the integral of the squared velocity (IV2). The CAV is used as a threshold to determine whether a damaging earthquake is coming [3] and is defined as

$$\text{CAV} = \int_0^{t_p} |\ddot{u}(t)|dt \tag{2}$$

where $\ddot{u}(t)$ is the vertical component of acceleration time history of ground motion after P-wave arrival. The IV2 is correlated to earthquake magnitude and is defined as [4]

$$\text{IV2} = \int_0^{t_p} \dot{u}^2(t)dt \tag{3}$$

In this study, the above-mentioned 6 P-wave features extracted from the vertical component of ground motion after P-wave arrival, i.e. P_a, P_v, P_d, T_c, CAV and IV2, are used for rapid estimation of PGA.

2.2. Support vector regression

As presented in the papers introduced in Section 2.1, the EEW techniques usually establish empirical regression between only one or two P-wave features and the target earthquake parameter. Furthermore, a simple linear relationship between the logarithm of both the P-wave features and the target earthquake parameter is often preferred [10,12,6,17,1,4]. Sometimes, the combination of the P-wave features could lead to a better empirical regression [4]. However, although concise and sometimes physical regression models are preferred, the simplification could result in a compromised regression model.

A better regression model could be achieved by considering multiple P-wave features and a more complex regression model, which can be accomplished by artificial intelligence. Support vector regression (SVR), which is a supervised learning method based on statistical learning theory, is outstanding for solving the multivariate problem [15]. Moreover, it is highlighted due to its outstanding advantages such as no local minimum problem and reliability at underfitting, overfitting or high noise conditions. Owing to those merits, this paper employs SVR algorithm to establish a nonlinear regression model between several P-wave features and PGA.

Given a training data $(\mathbf{x},\mathbf{y}) \in \Re$, the SVR finds a regression function which can best approximate the actual output vector $\mathbf{y}$ with an error tolerance ε, and is as flat as possible. This is done by mapping the input data $\mathbf{x}$ in a higher dimensional feature space, in which the data may exhibit linearity, and then perform linear regression in this feature space. Let $\mathbf{x}$ be mapped into a feature

space by a nonlinear function $\phi(\mathbf{x})$. The regression function can be expressed as

$$f(\mathbf{w},b)=\mathbf{w}\cdot\phi(\mathbf{x})+b \tag{4}$$

where $\mathbf{w}$ is a regression coefficient vector, and b is the model offset. A Vapnik's ε -insensitive loss function is adopted in the SVR as

$$\xi_i,\xi_i^*=\begin{cases} y_i-(w\cdot\phi(x_i)+b) & \text{for } |y_i-(w\cdot\phi(x_i)+b)|>\varepsilon \\ 0 & \text{otherwise}\end{cases} \tag{5}$$

where ξ_i and ξ_i^* are training errors located at upper side and lower side of the regression function, respectively, as defined in Fig. 1. When the ε-insensitive loss function is used, absolute errors larger than ε are accepted, and smaller errors in the tube $(-\varepsilon, \varepsilon)$ are neglected.

The regression problem can be expressed as the following convex optimization problem [14].

$$\min_{\mathbf{w},b,\xi,\xi^*}\left[\frac{1}{2}||\mathbf{w}||^2+C\left(v\varepsilon+\frac{1}{m}\sum_{i=1}^{m}(\xi_i+\xi_i^*)\right)\right] \tag{6}$$

subjected to

$$\mathbf{w}\cdot\phi(\mathbf{x}_i)+b-y_i\le\varepsilon+\xi_i$$
$$y_i-\mathbf{w}\cdot\phi(\mathbf{x}_i)-b\le\varepsilon+\xi_i^*$$
$$\xi_i,\xi_i^*\ge 0\quad i=1,\cdots,m$$

where m is the number of training data. Eq. (6) is called the v-support vector regression where v is the lower bound of the fraction of the number of support vectors; and C is a positive constant that determines the degree of penalized loss when a training error occurs. Because the size of variable ε is traded off against model complexity and training errors via the constant v, it will be determined automatically after training if v is chosen. The SVR avoids underfitting and overfitting the training data by minimizing both the regularization term $\|\mathbf{w}\|^2/2$ and the training error term $C(v\varepsilon+1/m\sum_{i=1}^{m}(\xi_i+\xi_i^*))$ in Eq. (6). Minimizing the first term is equivalent to maximize the margin of the tube, and minimizing the second term corresponds to minimizing the empirical risk.

Introducing a dual set of Lagrange multipliers, α_i and β_i, enables the optimization problem to be solved more easily in the dual form, by applying the standard quadratic programming algorithm.

$$\max_{\alpha,\beta}\left[-\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}(\alpha_i-\beta_i)(\alpha_j-\beta_j)k(\mathbf{x}_i,\mathbf{x}_j)+\sum_{i=1}^{m}y_i(\alpha_i-\beta_i)\right] \tag{7}$$

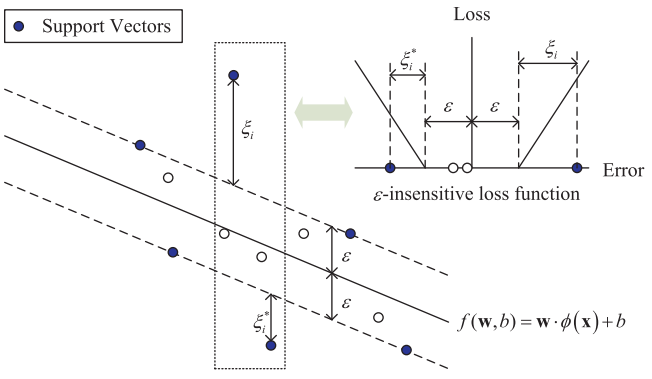

Fig. 1. Basic concepts of the SVR technique and the Vapnik's ε-insensitive loss function, where ε is the error tolerance. The absolute errors larger than ε are accepted, and smaller errors in the tube $(-\varepsilon, \varepsilon)$ are neglected. The parameters ξ_i and ξ_i^* are training errors located at upper side and lower side of the regression function, respectively.

subjected to

$$\sum_{i=1}^{m}(\alpha_i-\beta_i)=00\le\alpha_i,\quad \beta_i\le\frac{C}{m}\quad\text{for}\quad i=1,2,\cdots,m$$
$$\sum_{i=1}^{m}(\alpha_i+\beta_i)\le Cv$$

where $k(\mathbf{x}_i,\mathbf{x}_j)$ is the kernel function to yield the inner products in feature space; m is the number of training data. The radial basis function kernel, $k(\mathbf{x}_i,\mathbf{x}_j)=\exp(-||x_i-x_j||/2\sigma^2)$, is used in this paper. After the Lagrange multipliers, α_i and β_i, are determined. The parameters $\mathbf{w}$ and b can be estimated under Karush–Kuhn–Tucker complementarity conditions [5]. Therefore, the prediction function can be expressed as

$$f(\mathbf{x})=\left[\sum_{i=1}^{J}(\alpha_i-\beta_i)k(\mathbf{x},\mathbf{x}_i)\right]+b \tag{8}$$

where J is the number of nonzero $(\alpha_i-\beta_i)$, i.e. support vectors. The input data $\mathbf{x}$ and output data $\mathbf{y}$ of the SVR model are the six P-wave parameters and PGA, respectively. Because the PGA of the training data can cover a wide range, e.g. from less than 1 gal to more than 1000 gal, the training data with small PGA tends to be neglected when calculating the error. As a result, the estimation accuracy for small PGA could be quite poor. In order to estimate the PGA with acceptable accuracy for both small amplitude and large amplitude, the parameter v is chosen as 0.95 which makes most of the training data support vectors, including the one with small PGA. This also improves the estimation accuracy for data with small PGA.

There are two more parameters of the SVR model remain undetermined, i.e. parameter C and parameter σ in the kernel function. The simple grid search method with grids range from 2^{-10} to 2^{10} in intervals 2^1 for both parameters is applied to optimize these two parameters. In order to prevent overfitting, 10-fold cross validation process is employed. The parameters with the smallest cross validation root mean squared error are selected to construct the SVR model. After the best parameters C and σ are chosen, the SVR model can be constructed.

3. Earthquake data and preprocessing

The TSMIP network has been carried out by the Central Weather Bureau (CWB) to collect high-quality instrumental recordings of strong ground motions caused by earthquakes around Taiwan. About 700 free-field stations have been installed and are presently operating throughout Taiwan. The TSMIP station signals are digitized at 200 or 250 samples per second with 16 bit resolution or higher. Most accelerographs have a dynamic range of $\pm 2g$. Approximately 15-years of TSMIP data between 29 July 1992 to 31 December 2006 are employed in this paper. The original number of these TSMIP data are 105,360 three-components records. However, some of the data are obviously distorted and are therefore excluded from the data set. Moreover, some of the data lengths are shorter than 3 s which are also excluded. The remaining avaliable 91,142 data, named as "Testing Earthquake Data", are employed in this study to estimate the performance of the proposed approach in general.

In order to reduce the time consuming for training the SVR model, 71 earthquake events with local magnitude, M_L, between 3.0 and 7.3 are selected for all focal depth (2.8–282.8 km). Although Taiwan is a seismically active zone where 54 earthquake events with $M_L>6.0$ occured within 15 years, the strong ground motion records of large seismic intensity in the "Testing Earthquake Data" are still limited. If all the destructive ground motion records are used as the base for constructing the

SVR model, the estimation of accuracy of the SVR model may not be convincing because a question may be arised to doubt if the SVR model can predict other destructive ground motion records, even though the SVR model is believed robust to overfitting. Therefore, in this study, only half of the strong ground motion records of the 71 earthquake events will be used to train the SVR model. As a result, the database for training the SVR model consists of 4166 data, which is named as "Representing Earthquake Data", whose histogram and distribution of magnitudes and hypocenter distance are shown in Fig. 2.

The original strong ground motion records are acceleration signals. The zero-mean normalization of the records was applied. The records were integrated once and twice to obtain velocity and displacement signals, respectively. The second-order 0.075 Hz high-pass Butterworth filter was applied to remove the low-frequency drift after integration. The Short-Term Average/Long-Term Average (STA/LTA) algorithm was applied to automatically determine the P-wave arrival time.

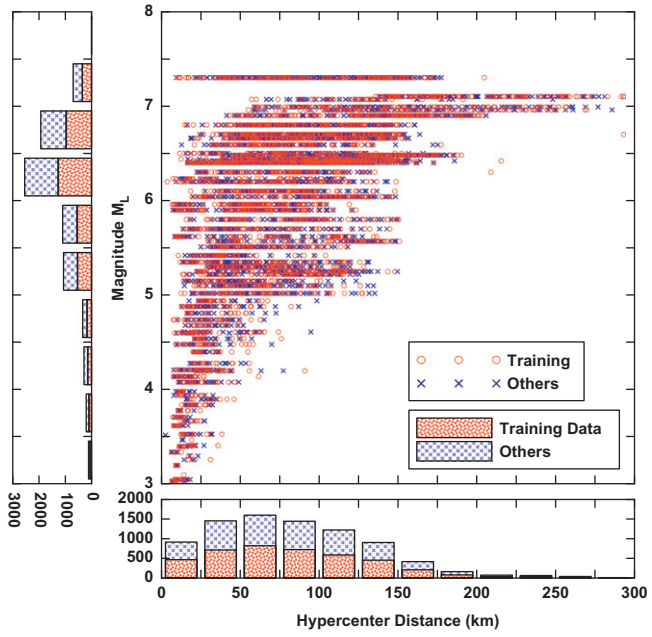

Fig. 2. Histograms and distributions of magnitudes M_L and hypocentral distances of the representing 71 earthquake events. The red circles mark the 4166 "Representing Earthquake Data" for training the SVR model, which is only half of the records of the 71 earthquake events. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4. Results and discussions

First of all, in order to find the best P-wave features to estimate PGA, the combinations of the six P-wave features used as the input data of the SVR model were studied. Totally 57 cases of all possible combinations were studied ($57=C_6^1+C_6^2+C_6^3+C_6^4+C_6^5+C_6^6$, where $C_k^n=n(n-1)(n-2)\cdots(n-k+1)/k!$). At first, a typical length of the time window t_p was choosen as 3 s, which was used most frequently in the literature. The "Representing Earthquake Data" was used to determine the parameters of the SVR models of different combinations. After the SVR model was constructed of each case, the standard deviation of the errors between real PGA and predicted PGA of the "Testing Earthquake Data" using the SVR model was calculated.

Fig. 3 depicts the standard deviations of the predicted PGA errors of the "Testing Earthquake Data" using the SVR models of

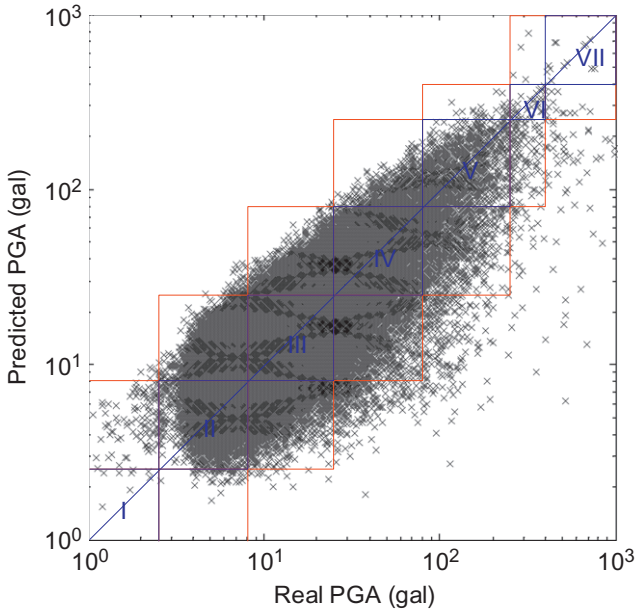

Fig. 4. Real PGA and the predicted PGA of the 91,142 "Testing Earthquake Data" using the SVR model of six P-wave features with $t_p=3$. The Roman numerals "I" to "VII" represent the seismic intensity scale in Taiwan. The regions enclosed by the blue lines represent the seismic intensity of the predicted PGA is the same as the real one, while the regions enclosed by the red lines represent the seismic intensity of the predicted PGA is within $\pm$one-scale of the real one. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

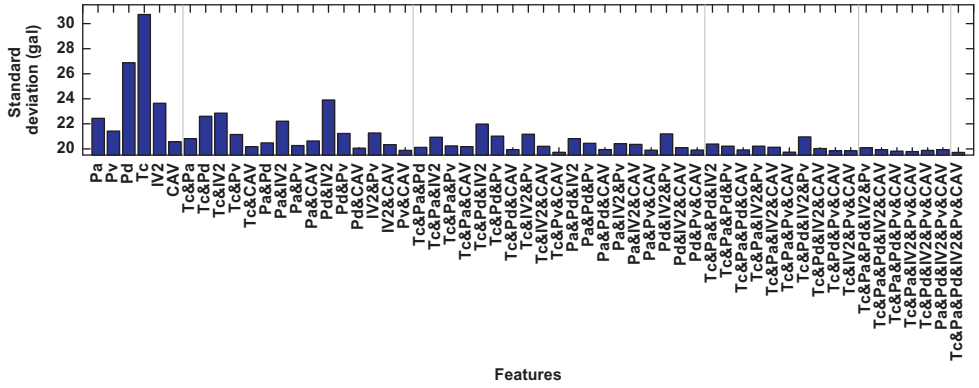

Fig. 3. The standard deviations of the predicted PGA errors using the SVR models of different combinations of P-wave features with $t_p=3$. P_a, P_v and P_d represent the peak acceleration, peak velocity and peak displacement, respectively; T_c represents the effective predominant period; CAV represents the cumulative absolute velocity and IV2 represents the integral of the squared velocity.

different combination of P-wave features as the input. In general, the predicted PGA errors decreases as more P-wave features are considered, which implies more information of the P-wave may help the SVR model predict the PGA. Among the SVR models which use only individual P-wave feature, the predicted PGA errors of the one using CAV is the smallest, followed by the one using P_v. Among the SVR models which use two P-wave features, the predicted PGA errors of the one using CAV and P_v is the smallest. If three P-wave features are used, the SVR model which uses CAV, P_v and T_c makes the smallest predicted PGA errors, which is already close to the predicted PGA errors of the one using six P-wave features. If fewer P-wave features are required to construct the SVR model, these three features could be choosen with higher priority. Among all SVR models, the one uses all six P-wave features predicts the PGA with the least error. Therefore, in this study, the SVR model using all six P-wave features is selected and employed in the following sections. The final value of the parameters C and σ were choosen as 4096 and 1.4142, respectively, which can be used to reconstruct the SVR model.

Fig. 4 compares the real PGA and the predicted PGA of the "Testing Earthquake Data" using the SVR model of six P-wave features. The regions enclosed by the blue lines and the red lines are within zero- and one-level difference of the seismic intensity scale of Taiwan for reference, respectively. It can be observed in the figure that the predicted PGA is quite approximate to the real PGA. The standard deviation of the predicted PGA errors is 20.89 gal, and the ratio of the predicted PGA located within one-level difference from the real PGA is 99.22%. Note that the predicted PGA is generally lower than the real PGA, which means the predicted PGA of the SVR model is in the non-conservative side. The predicted PGA could be scaled up to the conservative side for practical application, which is not considered in this paper.

Furthermore, in order to know the performance of the SVR model for specific earthquake events, the real PGA and the predicted PGA of the four earthquake events with largest PGA of the "Testing Earthquake Data" are compared in Fig. 5. It can be observed in the figure that the predicted PGA of the earthquake events are quite approximate to the real PGA, except the Chi-Chi earthquake on 21 September 1999, whose predicted PGA generally is almost two-levels underestimated comparing to the real PGA. Remind that EEW technique only employs the information

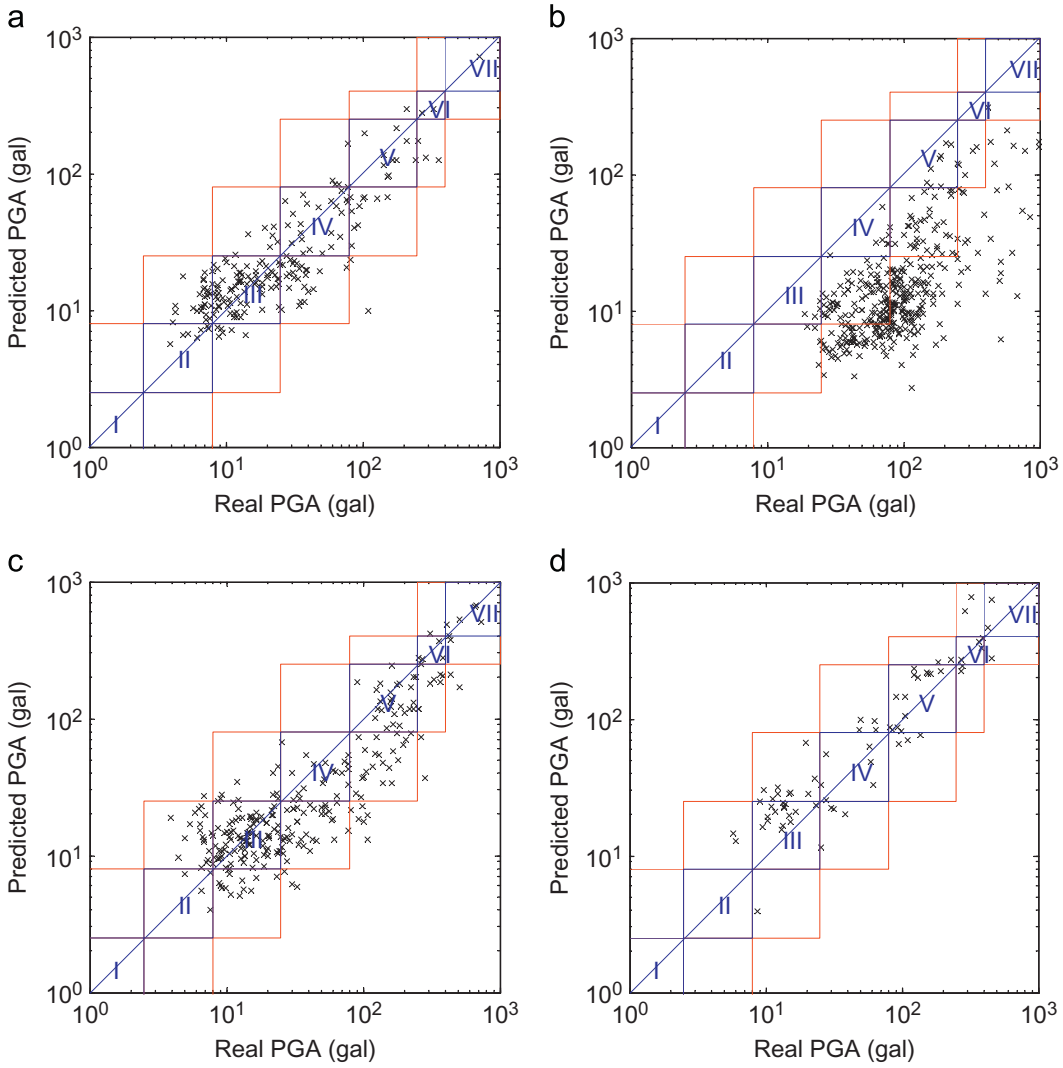

Fig. 5. Real PGA and the predicted PGA of the four earthquake events with largest PGA using the SVR model of six P-wave features with t_p=3: (a) Chai-Yi earthquake on 17 July 1998; (b) Chi-Chi earthquake on 21 September 1999; (c) Chai-Yi earthquake on 22 October 1999; (d) Tai-Tung earthquake on 1 April 2006. The Roman numerals "I" to "VII" represent the seismic intensity scale in Taiwan. The regions enclosed by the blue lines represent the seismic intensity of the predicted PGA is the same as the real one, while the regions enclosed by the red lines represent the seismic intensity of the predicted PGA is within $\pm$one-scale of the real one. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

carried by the first few seconds of P-wave, therefore the earthquake intensity with long or complex slip propagation process could be underestimated. The Chi-Chi earthquake was reported as with at least two asperities [8]. Base on the fault model proposed in the paper, one minor asperity was located in the hypocenter region with average slip of about only 3 m, while the major asperity was located about 30 km to 65 km north of the hypocenter region with average displacement about 9 m. The rupture of the north major asperity was believed to start after about 13 s of the rupture of the minor asperity at the hypocenter. As a result, the estimation of the Chi-Chi earthquake intensity, which was mainly caused by the rupture of the major asperity 13 s after the minor asperity rupture, based on the first few seconds of P-wave become not possible. Nevertheless, the estimation of PGA could be improved if longer t_p is used. An example is illustrated in Fig. 6 where the PGA is predicted by the SVR model with $t_p=10$. Note that, although the PGA is predicted much close to the real PGA in this case, the response time is sacrificed.

Next, the effect of the length of time window t_p was studied. Although longer t_p reaches higher reliability, the lead-time of early warning is sacrificed. As discussed, the lead-time is the crucial point for an effective earthquake warning system which should issue an early warning before the arrival of destructive seismic waves, instead of issuing an alarm with very

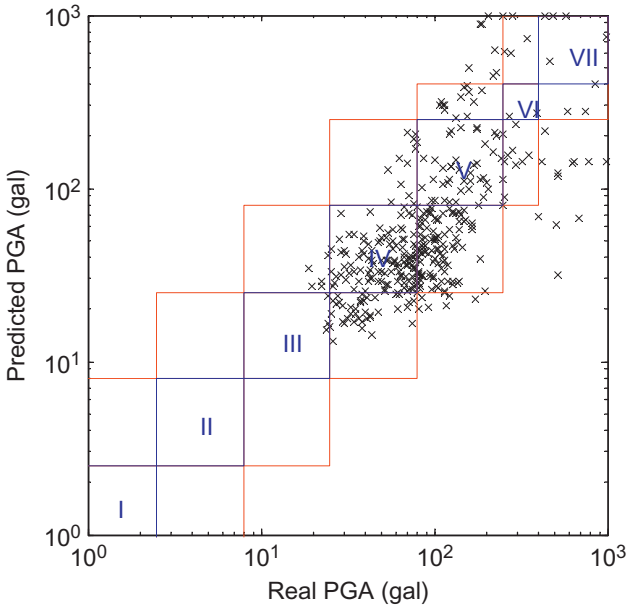

Fig. 6. Real PGA and the predicted PGA of the Chi-Chi earthquake using the SVR model of six P-wave features with $t_p=10$. The Roman numerals "I" to "VII" represent the seismic intensity scale in Taiwan. The regions enclosed by the blue lines represent the seismic intensity of the predicted PGA is the same as the real one, while the regions enclosed by the red lines represent the seismic intensity of the predicted PGA is within $\pm$ one-scale of the real one. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

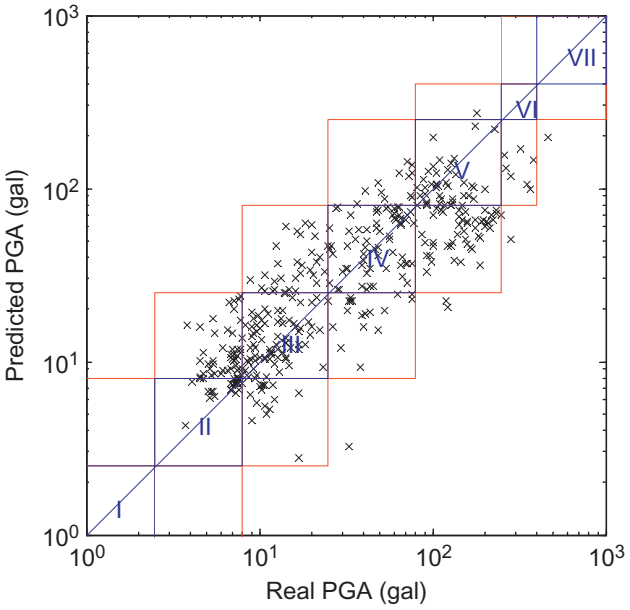

Fig. 8. Real PGA and the predicted PGA of the 2010 Kaohsiung earthquake using the SVR model with $t_p=3$. The Roman numerals "I" to "VII" represent the seismic intensity scale in Taiwan. The regions enclosed by the blue lines represent the seismic intensity of the predicted PGA is the same as the real one, while the regions enclosed by the red lines represent the seismic intensity of the predicted PGA is within $\pm$ one-scale of the real one. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

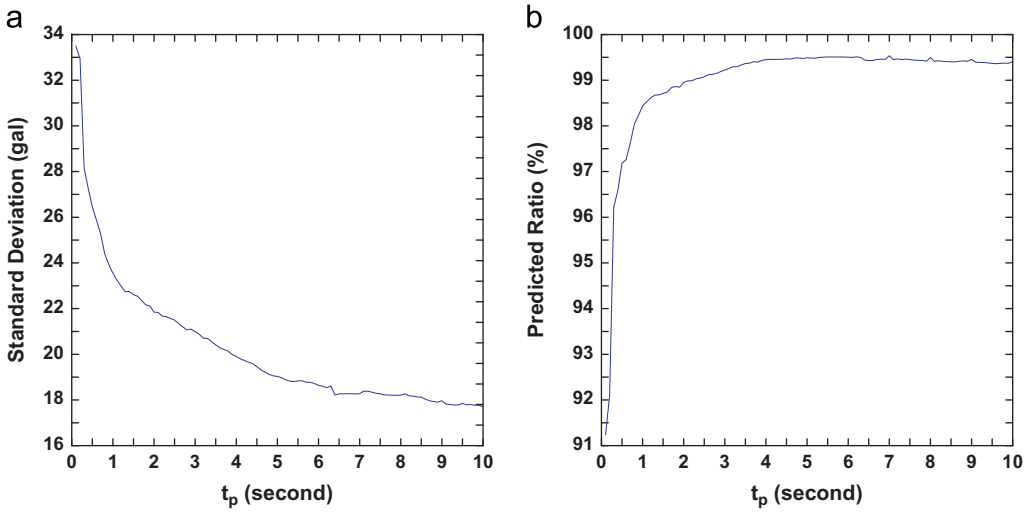

Fig. 7. The performance of the SVR model of the "Testing Earthquake Data" using different length of time window: (a) standard deviation of the predicted PGA errors; and (b) ratio of the predicted PGA located within one-level difference from the real PGA.

high reliability after or during the invading of destructuve seismic waves.

The length of the time window t_p ranges from 0.1 s to 10 s in intervals 0.1 s. The results show that the standard deviation of the predicted PGA errors of the "Testing Earthquake Data" decreases as t_p increases, as shown in Fig. 7. The standard deviation of predicted error diminishes very quickly from about 34 gal to 23 gal within the first 1 s. It continues to decrease gradually thereafter to less than 18 gal. In addition, the ratios of the predicted PGA located within one-level difference from the real PGA are also plotted in the same figure. The one-level predicted ratios increase dramatically from about 91% to 97% within the first 0.4 s and then continue to increase very fast to almost 99% till $t_p=1$, which is already quite close to the maximum value saturated after around $t_p=4$. It seems that t_p could be less than 3 s for practical application using the proposed approach, thus the extent of the region without lead-time could be reduced and valuable response time could be earned.

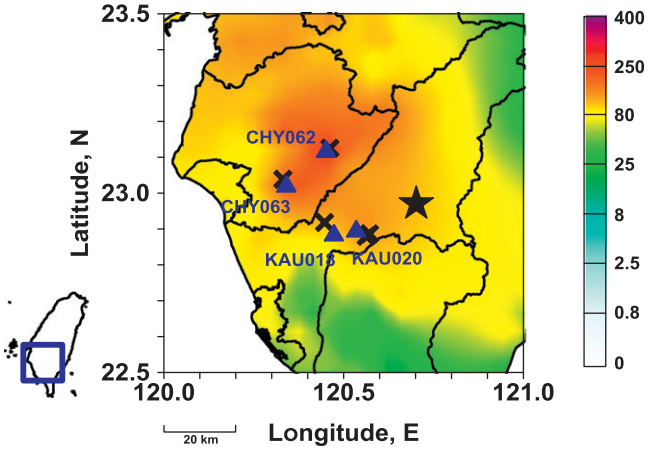

Fig. 9. The locations of the four stations, the adjacent damaged buildings and the epicenter of the 2010 Kaohsiung earthquake, plotted in the contour map of the measured PGA (gal). The black star represents the epicenter, the blue triangle represents the station and the black cross represents the adjacent damaged building. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Finally, the feasibility of the proposed SVR model is studied through application to the 2010 Kaohsiung earthquake which is absent from the "Testing Earthquake Data". The Kaohsiung earthquake with magnitude $M_L=6.4$ and focal depth 22.6 km caused 96 injured and hundreds of buildings damaged with different levels. The maximum PGA is 463.03 gal measured at CHY062 station about 30.48 km from the epicenter, while the minimum PGA is 5.16 gal measured at TAP057 station about 254.46 km from the epicenter. All the TSMIP records of the Kaohsiung earthquake are with good quality, and the real PGA and the predicted PGA of these 390 earthquake records are compared in Fig. 8 using the SVR model with $t_p=3$. Again, the overall approximate relationship between real PGA and the predicted PGA can be observed.

The performance of the SVR models at several stations close to the epicenter are also studied. According to the reconnaissance report of National Center for Research on Earthquake Engineering in Taiwan (NCREE), four stations, i.e. KAU020, KAU018, CHY062 and CHY063, with short distance from the epicenter of 18.46 km, 24.8 km, 30.5 km and 37.2 km respectively are accompanied with severe or moderate damage of nearby buildings and nonstructures

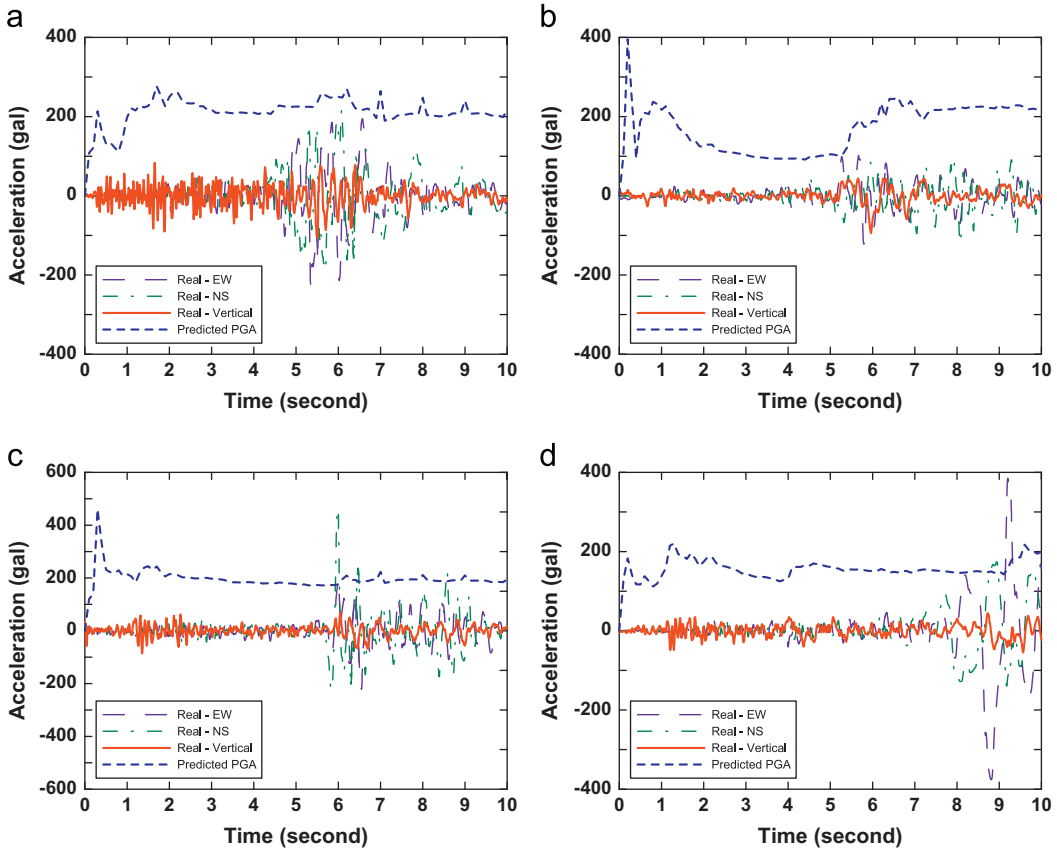

Fig. 10. The predicted PGA of the SVR models with t_p ranges between 0.1 s and 10 s in intervals 0.1 s at station: (a) KAU020; (b) KAU018; (c) CHY062; and (d) CHY063. The measured acceleration time history of three directions at each station is also plotted.

[11]. The locations of the four stations, the adjacent damaged buildings and the epicenter are marked in Fig. 9.

The predicted PGA at these four stations using the SVR models with different length of time windows, i.e. $t_p=0$ to $t_p=10$, and the measured acceleration time history of three directions are plotted together in Fig. 10. It can be observed that after the arrival of P-wave, all the predicted PGA of these four stations at different time are larger than 80 gal or 92 gal, which corresponds to intensity V of Taiwan scale and intensity VI of MMI scale [16]. In other words, if the warning is issued based on these intensities, the region close to these four stations could be alerted right after the arrival of P-wave. However, in practice, due to the higher uncertainty of the predicted PGA using less P-wave information, the warning could be postponed until the first second if higher reliability is demanded. Furthermore, if the warning is issued at the first second after the arrival of P-wave, the response time before the strike of the largest seismic wave of these four stations are 4 s to 8 s.

5. Conclusions and discussions

In this paper, a novel approach for EEW is proposed. The approach estimates PGA using the SVR models based on P-wave features extracted from the first few seconds of the vertical ground acceleration record at a single station. These features include T_c, P_a, P_v, P_d, CAV and IV2. The effects on the performance of the SVR models using different combinations of these P-wave features are studied. The results show that the SVR model which uses these six features together may have the smallest prediction error of the PGA. However, if fewer P-wave features are preferred to construct the SVR model, the features CAV, P_v and T_c could be chosen with higher priority.

The SVR models are trained using the "Representing Earthquake Data" which includes only half of the earthquake records of 71 representative earthquake events of TSMIP from 1992 to 2006 (totally 4116 records). In order to estimate the overall prediction error in practice, the SVR models are validated by the "Testing Earthquake Data" which includes the whole 91,142 earthquake records of TSMIP of the same period. Generally, the predicted PGA is quite approximate to the real PGA, but is slightly underestimated. The standard variation of the error between the predicted PGA and the real PGA for the "Testing Earthquake Data" using the SVR model with $t_p=3$ is 20.89 gal, while the ratio of the predicted PGA located within one-level difference from the real PGA is 99.22% (Taiwan seismic intensity scale).

The effects on the performance of the SVR models using $t_p=0.1$ s to $t_p=10$ s in intervals 0.1 s are also studied. As expected, higher reliability of PGA prediction could be achieved with longer t_p. However, the standard deviation of the prediction error diminishes very quickly within the first 1 s, while the one-level prediction ratio also increase dramatically and reaches a high percentage which is close the maximum value. As a result, more remaining reaction time could be earned for practical application if smaller t_p is used.

The feasibility of the proposed SVR model is also studied trough application to the damaging 2010 Kaohsiung earthquake which is excluded from the "Testing Earthquake Data". Again, the overall approximate prediction of PGA is observed. In addition, the performance of the SVR model at four stations accompanied with severe or moderate damage of adjacent buildings and nonstructures are illustrated. It seems that the warning could be issued at these four stations based on the large PGA predicted. Furthermore, even though the epicenter distance of these four stations are quite small, i.e. between 18.46 km and 37.2 km, the response time before the strike of the largest seismic wave are 4–8 s if the warning is issued at the first second after the arrival of P-wave. Note that the focal depth of the Kaohsiung earthquake is 22.6 km. The response time could be shortened if the focal depth is less.

Although the proposed SVR model seems to perform well for most of the earthquakes in the "Testing Earthquake Data", the predicted PGA of the earthquakes with long or complex slip propagation process could be underestimated if short t_p is used, due to that the information carried by the first few seconds of P-wave is limited. This phenomenon is usually observed when applying to huge subduction-zone earthquakes, such as the 2011 Tohoku earthquake in Japan and the 2010 Chile earthquake. In the application of the proposed SVR model with $t_p=3$ on the 1999 Chi-Chi earthquake in Taiwan, the predicted PGA is generally underestimated due to the long or complex two-stage slip process. The estimation of PGA could be greatly improved if longer t_p is used; however, the response time is sacrificed. Future research to overcome the challenge of EEW techniques due to long or complex slip propagation process is still required.

Acknowledgment

We thank the two reviewers for their comments and suggestions that have significantly improved the manuscript. We also thank the CWB in Taiwan for providing strong motion data.

References

[1] Böse M, Ionescu C, Wenzel F. Earthquake early warning for Bucharest, Romania: novel and revised scaling relations. Geophysical Research Letters 2007;34:L07302.
[2] Böse M, Heaton T, Hauksson E. Rapid estimation of earthquake source and ground-motion parameters for earthquake early warning using data from a single three-component broadband or strong-motion sensor. Bulletin of the Seismological Society of America 2012;102(2):738–50.
[3] Erdik M, Fahjan Y, Ozel O, Alcik H, Mert A, Gul M. Istanbul earthquake rapid response and the early warning system. Bulletin of Earthquake Engineering 2003;1(1):157–63.
[4] Festa G, Zollo A, Lancieri M. Earthquake magnitude estimation from early radiated energy. Geophysical Research Letters 2008;35(22):L22307.
[5] Fletcher R. Practical methods of optimization. 2nd ed. New York: John Wiley; 1987.
[6] Kanamori H. Real-time seismology and earthquake damage mitigation. Annual Review of Earth and Planetary Sciences 2005;33:195–214.
[7] Liu KS, Shin TC, Tsai YB. A free-filed strong motion network in Taiwan: TSMIP. Terrestrial, Atmospheric and Oceanic Sciences 1999;10(2):31–50.
[8] Ma KF, Mori J, Lee SJ, Yu SB. Spatial and temporal distribution of slip for the 1999 Chi-Chi, Taiwan, Earthquake. Bulletin of the Seismological Society of America 2001;91(5):1069–87.
[9] Nakamura Y. On the urgent earthquake detection and alarm system (UrEDAS). In: Proceedings of the 9th world conference earthquake engineering; 1988 (7):673–8.
[10] Nakamura Y. A new concept for the earthquake vulnerability estimation and its application to the early warning system. In: Proceedings of early warning conference potsdam. Germany; 1998.
[11] National Center for Research on Earthquake Engineering. Reconnaissance report on Jiaxian earthquake in Kaohsiung on March 4. NCREE-10-010. Taipei; 2010 [in Chinese].
[12] Odaka T, Ashiya K, Tsukada S, Sato S, Ohtake K, Nozaka D. A new method of quickly estimating epicentral distance and magnitude from a single seismic record. Bulletin of the Seismological Society of America 2003;93(1):526–32.
[13] Satriano C, Wu YM, Zollo A, Kanamori H. Earthquake early warning: concepts, method and physical grounds. Soil Dynamics and Earthquake Engineering 2011;31:106–18.
[14] Schölkopf B, Smola A, Williamson R, Bartlett PL. New support vector algorithms. Neural Computation 2000;12:1207–45.
[15] Vapnik VN. The nature of wtatistical learning theory. Berlin: Springer-Verlag; 1995.
[16] Worden CB, Gerstenberger MC, Rhoades DA, Wald DJ. Probabilistic relationships between ground-motion parameters and Modified Mercalli Intensity in California. Bulletin of the Seismological Society of America 2012;102(1):204–21.
[17] Wu YM, Kanamori H. Rapid assessment of damaging potential of earthquakes in Taiwan from the beginning of P waves. Bulletin of the Seismological Society of America 2005;95:1181–5.