

Research Article

Study on the Magnitude of Reservoir-Triggered Earthquake Based on Support Vector Machines

Hai Wei , Mingming Wang , Bingyue Song, Xin Wang, and Danlei Chen

Faculty of Electric Power Engineering, Kunming University of Science and Technology, Kunming, Yunnan 650500, China

Correspondence should be addressed to Mingming Wang; wang.ming.ming@163.com

Received 31 January 2018; Revised 31 May 2018; Accepted 27 June 2018; Published 24 July 2018

Academic Editor: Minvydas Ragulskis

Copyright © 2018 Hai Wei et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

An effective approach is introduced to predict the magnitude of reservoir-triggered earthquake (RTE), based on support vector machines (SVM) and fuzzy support vector machines (FSVM) methods. The main influence factors on RTE, including lithology, rock mass integrity, fault features, tectonic stress state, and seismic activity background in reservoir area, are categorized into 11 parameters and quantified by using analytical hierarchy process (AHP). Dataset on 100 reservoirs in China, including the 48 well-documented cases of RTE, are collected and used to train and validate the prediction models established with SVM and FSVM, respectively. Through numerical tests, it is found that both the SVM and FSVM models are effective in the prediction of the magnitude of RTE with high accuracy, provided that sufficient samples are collected. While the results of FSVM which is extended from SVM by introducing a fuzzy membership to reduce the influence of noises or outliers are found to be slightly less accurate than those of SVM in the current analysis of RTE cases. The reason might be attributed to the high discreteness of the sample data in the current study.

1. Introduction

Reservoir-triggered earthquake (RTE), or reservoir earthquake (RE) in brief [1–3], also called as reservoir-induced earthquake (RIE), was caused by impounding water in reservoir created by the construction of dam across rivers. The first case of RTE was pointed out by Carder, which happened at Lake Mead in the USA [4]. Over the past decades, about 130 suspected cases of RTE have been reported in the world including the USA, India, Greece, Egypt, and New Zealand [3]. Among them, about 100 cases were universally acknowledged as RTE, since they were very hard to be distinguished from natural earthquakes. China is among the countries most prone to earthquakes, where about 40 RTE cases have been reported so far. They have brought significant threats to the safety of large dams and consequently are great hazards to the life and property of local residents [5], especially in the west of Sichuan province and the northwest of Yunnan province, where large dams have been built or are under construction in high density. It is necessary for dam engineers to know a reasonable prediction for the magnitude of RTE in order to assess the dam behavior subjected to this type of

seismic load. An effective prediction on the possibility of RTE and its magnitude in case of occurrence is thus essential in the design of large dams for the safety of life and property in related areas.

Like natural earthquakes, the mechanism of RTE is extremely complicated and remains almost completely unknown to seismologists. Since the occurrence time of RTE is very difficult to be predicted, previous studies on RTE have mainly focused on the prediction of its location and magnitude. Through extensive studies by seismologists, it has been widely acknowledged that RTE was closely influenced by combined factors including lithology, mechanical parameters of deep rock mass, the stress state of rock mass, and pore water distribution [6]. However, due to the difficulty in the measurement and quantification of those factors, great challenges have been posed on the analysis of RTE [7, 8].

In spite of this, various analytical approaches have been proposed on the study of RTE in the past decades, such as mechanical model method [7], geological analogy method [7], probabilistic and statistical method [8–10], artificial neural network method (ANN) [11, 12], and artificial intelligence technique [13]. Among them, the mechanical

model method is limited by the fact that the nonlinearity of rock, parameter variation, and pore water distribution as well as boundary conditions cannot be taken into full consideration. Therefore, the geological analogy method is proposed based on analogy with RTE cases, in terms of hydrology, engineering geology, and infiltration conditions. The possibility of RTE, as well as its magnitude and location in case of occurrence, is thus able to be analyzed. However, foundational mechanical principles like the force equilibrium condition cannot be included in this method. The probabilistic and statistical method [8], proposed based on Bayes' theorem, can also be used to evaluate the occurring probability of RTE. It is based on statistical analysis of previous RTE cases, and thus the accuracy of the result is strongly dependent upon sample numbers and prior probability of RTE. The artificial neural network method has strong ability in nonlinear mapping and thus shows great advantages in dealing with complex nonlinear relations in complicated problems, whereas the uncertainty of network structures and problems related to overfitting and underfitting have limited its application in the analysis of RTE.

The support vector machines (SVM) is a powerful machine learning method firstly proposed by Vapnik in 1995 [14]. It is derived from the statistical learning theory and has strong generalization capability. The best compromise between the complexity of model and learning ability could be achieved with the employment of structure risk minimization (SRM) principle, which is developed from the traditional empirical risk minimization (ERM) method [14–16]. SVM was applied initially to pattern classification and later to data regression problems. By introducing a kernel function, the original application of SVM in the optimal classification of linear separable data is extended to nonlinear problems. They could be solved by transferring to a quadratic programming (QP) problem with constraint condition [16]. With its excellent performance in dealing with nonlinear and high-dimensional problems, SVM has been widely applied in multiple fields, such as image recognition, classification, time series prediction, inverse analysis of geotechnical parameters, and hydrological forecasting [17–21]. However, conventional SVM is found very sensitive to noises or outliers. As a result, Lin and Wang [22] proposed fuzzy support vector machines (FSVM) by applying a fuzzy membership to each input point to reduce the effect of noises or outliers. In order to reduce the dependence of membership function on the geometric shape of sample data, Tang et al. [23] and Du et al. [24] introduced a new membership function based on the distance between a sample and its class hyperplane in FSVM. To further reduce the membership of nonsupport vectors and increase that of support vectors, Ding and Gu [25] proposed a new algorithm where a dual membership based on hypersphere was employed. For multiclass classification problems, Tsang et al. [26] defined a degree of membership of a sample to different classes by introducing a fuzzy membership to each training sample, while Abe [27] classified a sample into a multilabel class whose membership function is the largest.

In this study, we attempt to extend the application of SVM and FSVM to predict the magnitude of RTE, based on

dataset collected on reservoirs and RTE cases in China. Habibagahi [12] and Samui and Kim [13] have also used artificial intelligence techniques to predict the magnitude of induced earthquakes. In their models, only limited factors on reservoir parameters such as comprehensive parameter and maximum reservoir depth are taken into account. While geological factors, which are of great significance in the triggering of earthquakes, are not considered. Therefore, an effort has been made to carry out this research covering multiple geological factors of reservoir area and reservoir parameters for providing a step forward of research on RTE. Dataset on 100 large and medium-sized reservoirs, covering the 48 RTE cases, are collected from reputed enterprises and published literatures. Computational models are established based on SVM and FSVM. The analytical hierarchy process (AHP) [28] is employed to quantify geological factors based on geological analysis, providing the input data of the computational models. Three numerical tests are carried out to test the performance of the established models. Results obtained from the SVM and FSVM models are analyzed and compared in detail. Discussions are made on the applicability of SVM and FSVM, which offers an effective approach in the field of RTE prediction.

2. Mathematical Models and Numerical Procedure

2.1. The SVM and FSVM Models. The solution of FSVM classification can also be transferred into a quadratic programming (QP) problem like conventional SVM. By solving the dual problem of QP, we can easily obtain the solutions of FSVM. The calculation algorithm was proposed by Vapnik [15] and Deng and Tian [16], which can be summarized into the following steps:

- (1) Assume a training set $T\{(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)\} \in (X \times Y)$, where $x_i \in X = R^n$, $y_i \in Y = \{-1, 1\}$, $i = 1, 2, \dots, l$. Introduce a fuzzy membership s_i ($0 \leq s_i \leq 1$, $i = 1, 2, \dots, l$) [22], which can be regarded as the attitude of corresponding training point i toward one class. If $s_i = 0$, it indicates that training point i does not absolutely belong to one class; on the contrary, $s_i = 1$ indicates that training point i is completely within one class. The method of determining the value of s_i for training point i will be described later. For this binary classification (BC) problem, it can be transformed into QP problems with constraint condition including the fuzzy membership. The corresponding dual problem can be expressed as

$$\min_{\alpha} \left(\frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l y_i y_j \alpha_i \alpha_j K(x_i, x_j) - \sum_{j=1}^l \alpha_j \right), \quad (1)$$

$$\text{s.t.} \quad \begin{cases} \sum_{i=1}^l y_i \alpha_i = 0, \\ 0 \leq \alpha_i \leq s_i C, i = 1, 2, \dots, l, \end{cases} \quad (2)$$

where C is the regularization parameter, α is the Lagrange multiplier, K is a kernel function, and s_i is the fuzzy membership. When $s_i = 1$, this dual problem is equivalent to standard SVM.

Among several kernels, polynomial kernels and radial basis function (RBF) kernels are often used for pattern classification. While in most cases RBF kernels are found to perform better [27], it can be expressed as

$$K(x, x') = \exp\left(\frac{-\|x - x'\|^2}{\sigma^2}\right), \quad (3)$$

where x is the input vector and σ^2 is the parameter of the RBF kernel function.

By solving the above QP, the optimal solution of this dual problem could be expressed as $\alpha^* = (\alpha_1^*, \alpha_2^*, \dots, \alpha_l^*)^T$, in which the component α_i corresponds to sample T_i .

- (2) The construction of optimal hyperplane based on optimal solution α^* of above QP, the optimal hyperplane can be achieved as

$$w^* \cdot x + b^* = 0, \quad (4)$$

where $w^* = \sum_{i=1}^l y_i \alpha_i^*$ is the weight of sample i , and b^* can be calculated by choosing some positive α_j^* less than C , based on the following formula

$$b^* = y_j - \sum_{i=1}^l y_i \alpha_i^* K(x_i \cdot x_j). \quad (5)$$

- (3) The determination of the decision function $f(x)$ for binary classification, the decision function $f(x)$ can be expressed as

$$f(x) = \text{sgn}\left(\sum_{i=1}^l y_i \cdot \alpha_i^* K(x_i, x) + b^*\right). \quad (6)$$

The determination of the fuzzy membership s_i is crucial to the performance of FSVM. At present, s_i is mostly determined based on the membership function, whose value is determined by the distance between the sample and its class center [22]. But this method is irrational for dataset with nonspherical-shape distribution. In the current study, we shall employ a new membership function proposed by Du et al. [24] and Ding and Gu [25], where the basic idea is to introduce the class hyperplane to substitute the cluster center. The class hyperplane is defined through the cluster center, and perpendicular to the line across the positive and negative cluster center, as shown in Figure 1. The fuzzy membership is then stated as the distance between the sample and the class hyperplane. In this way, the dependence of membership function on the geometric shape of sample data is greatly reduced, and thus the generalization of FSVM is significantly enhanced.

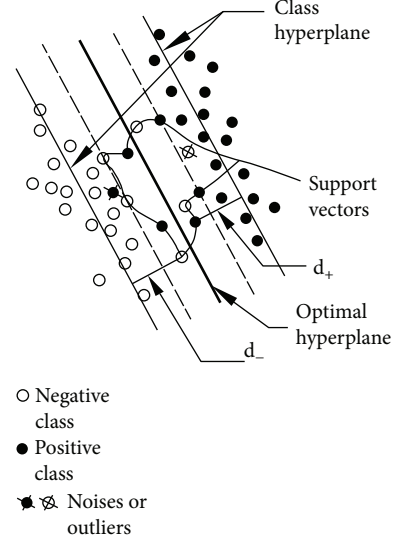


FIGURE 1: Scheme of SVM and FSVM (for 2-dimensional case).

Based on this new method, for nonlinear classification problems, the distance between the sample i and the class hyperplane, denoted by d_{i+} , d_{i-} , respectively, for samples in positive, negative class, as shown in Figure 1, can be obtained by using the kernel technique, as follows:

$$\begin{aligned} d_{i+} &= \left| \frac{1}{n_+} \sum_{j=1}^{n_+} K(x_j, x_i) - \frac{1}{n_-} \sum_{j=1}^{n_-} K(x_j, x_i) - \frac{1}{n_+^2} \sum_{i=1}^{n_+} \sum_{j=1}^{n_+} K(x_j, x_i) \right. \\ &\quad \left. + \frac{1}{n_+ n_-} \sum_{i=1}^{n_+} \sum_{j=1}^{n_-} K(x_j, x_i) \right|, \\ d_{i-} &= \left| \frac{1}{n_+} \sum_{j=1}^{n_+} K(x_j, x_i) - \frac{1}{n_-} \sum_{j=1}^{n_-} K(x_j, x_i) + \frac{1}{n_+^2} \sum_{i=1}^{n_+} \sum_{j=1}^{n_+} K(x_j, x_i) \right. \\ &\quad \left. - \frac{1}{n_+ n_-} \sum_{i=1}^{n_+} \sum_{j=1}^{n_-} K(x_j, x_i) \right|, \end{aligned} \quad (7)$$

where n_+ denotes the number of samples in positive class, n_- denotes the number of samples in negative class, and K is the kernel function.

The maximum distance between positive samples and the class hyperplane was denoted as $D_+ = \max(d_{i+})$, and that for negative samples is denoted as $D_- = \max(d_{i-})$.

Consequently, the membership degree of samples in each class can be defined as:

$$s_i = \begin{cases} 1 - \frac{d_{i+}}{D_+ + \delta} & \text{(for } x_i \text{ within positive class),} \\ 1 - \frac{d_{i-}}{D_- + \delta} & \text{(for } x_i \text{ within negative class),} \end{cases} \quad (8)$$

where δ is a very small positive number to avoid the case $s_i = 0$ and to guarantee $0 < s_i \leq 1$.

It can be seen from (8) that s_i gets smaller when sample x_i locates farther away from the class hyperplane and vice

versa. However, if a sample x_i is far away from its class hyperplane and not within its own class, it will most likely become a noise or outlier point and can severely affect the position of optimal hyperplane in standard SVM. While in FSVM, with the introduction of the fuzzy membership s_i , which becomes very small for noises or outliers, the effect of noises or outliers could then be eliminated on the position of the optimal hyperplane.

In addition, the position of the optimal hyperplane obtained with SVM is significantly influenced by the choice of the kernel function, the regularization parameter, and the relaxation coefficient, which need to be optimized first before applying [17, 18, 29].

2.2. Numerical Procedure

2.2.1. Quantitative Analysis on Parameters and Grading for the Magnitude of RTE.

The geological influence factors on RTE include the composition and structure of rocks at the base of the reservoir, the distribution, attitude, mechanical characteristics, growth level, and the present activity of the fault, as well as the hydrological conditions [30]. Nevertheless, those factors are very difficult to quantify, which brings great difficulty in the quantitative analysis of RTE. In the current study, the analytical hierarchy process (AHP) [28], which is an efficient way to deal with unquantifiable parameters, is employed to quantify these influence factors. Those geological factors are represented by natural number after employing AHP.

2.2.2. Data Normalization.

To prevent the models from domination by input variables with large value for different dimension, the original value of sample data is normalized by the following equation.

$$x_{ij}' = \frac{x_{ij} - \mu_{x_i}}{\sigma_{x_i}}, \quad (9)$$

where x_{ij} is the original value of sample, x_{ij}' is the scaled value of sample, μ_{x_i} is the expectation of sample, and σ_{x_i} is the standard deviation of sample.

2.2.3. The Establishment of the SVM and FSVM Models.

Since there are multiple classes in the prediction models of RTE, the SVM and FSVM models involve multiple classifications. For multiple classifications, SVM and FSVM models could be established by “one-against-all” and “one-against-one” ways [16].

Based on previous mathematical models and numerical procedure, the algorithm of SVM and FSVM is shown in the flow chart of Figure 2, for which computing codes can be programmed.

3. The Prediction Models of RTE Magnitude Based on SVM and FSVM

3.1. Data Construction.

Dataset on 100 large and medium-sized reservoirs in China, covering the 48 well-documented cases of RTE, are collected from reputed enterprises and published literatures for the current study. Based on them, 11

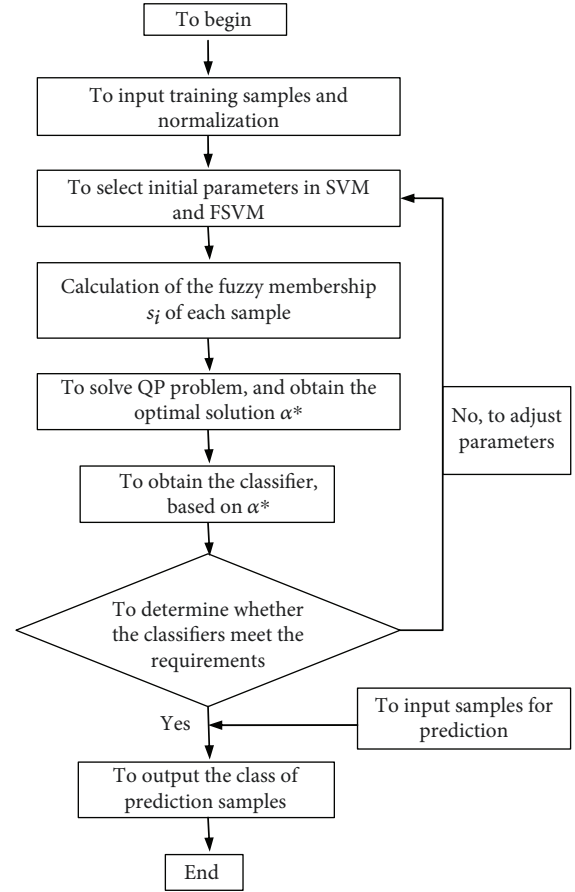


FIGURE 2: Flow chart of FSVM and SVM algorithm.

influence factors, including the geological factors and reservoir parameters, are taken into account in the analysis of RTE, as discussed below.

3.1.1. Dam Height and the Capacity of Reservoir.

Through statistical analysis of RTE cases, it has been found by Jiang et al. [30] that RTE has no obvious relation to dam height and the capacity of reservoir. This means that RTE may happen in any reservoir area regardless of its capacity. For example, the Dengjiaqiao reservoir, located in Hubei province in central China, has triggered an earthquake of Ms 2.2, while the capacity of reservoir is only $400 \times 10^3 \text{ m}^3$ [6].

3.1.2. Lithology at the Reservoir or Epicenter.

To study the influence of lithology at the reservoir area or the epicenter, we have performed statistical analysis on 48 RTE cases. It is found that the lithology in the epicenter is widely distributed, while a sufficient number of samples are essential to enhance the accuracy of SVM analysis. As a result, rocks with close lithology features are classified together. For example, limestone and dolomite are considered together as carbonatite, schist and slate are classified into one category, and pyroclastic rock and tuff are classified as volcanic rock. In this way, more than 10 types of rock are defined for lithology at the reservoir area or the epicenter based on those RTE cases. The representative

TABLE 1: Lithology classification.

Lithology at reservoir area and epicenter	Representative value
Carbonatite	1
Sandstone	2
Clay rock	3
Gneiss	4
Schist and slate	5
Quartzite	6
Marble	7
Granite	8
Basalt	9
Oceanic rock	10
Others	11

TABLE 2: Rock mass integrity.

Rock mass grade	I	II	III	IV	V
Representative value	1	2	3	4	5

values of lithology are shown in Table 1, and those of rock integrity are shown in Table 2.

3.1.3. Type and Characteristics of Fault. The condition of joint and fault has strong influence on the integrity of the rock mass and thus the magnitude of RTE. It is known that surface rock stratum with developed joints and broken rocks has high permeability, for which microearthquake or ultra-microearthquake might be triggered. In relatively deep rock stratum, areas with active or tensional faults have high permeability, and water permeating from reservoir into the deep rock mass might change the pressure distribution of pore water. In this case, the earthquakes triggered are mostly weak and moderate ones. There are normally a large number of faults existing in the reservoir area. Thus, only some main faults are selected as the analyzing subjects considering their scale and influence degree. The representative values related to the type, the development degree, the activity level, and the permeability of the fault are shown in Tables 3–6, respectively.

3.1.4. Tectonic Stress State at the Reservoir Area. It is found that RTE has no obvious relation to tectonic stress state at the reservoir area [30]. This means that RTE could take place not only at areas with active tectonic activity but also at inactive areas. However, enormous elastic strain energy can be stored within rock mass in areas with high local stress. The state of rock mass may become close to rupture or slip, which could lead to relatively strong earthquakes. The representative values of different regional stress state are listed in Table 7.

3.1.5. The Background of Seismicity at Reservoir Area. Based on statistical analysis of RTE, it is found that RTE tends to occur at weak seismic regions, and very few RTE have been reported in relatively strong seismic regions [30]. This is owing to the fact that the strain energy could be easily

TABLE 3: Fault type.

Fault type	Normal fault	Reverse fault	Strike-slip fault
Representative value	1	2	3

TABLE 4: Fault development degree.

Fault development degree	Representative value
Very developed	1
Developed	2
Slightly developed	3
Undeveloped	4
Very undeveloped	5

TABLE 5: Fault activity.

Fault activity	Active	Medium	Inactive
Representative value	1	2	3

TABLE 6: Fault permeability.

Fault permeability	Well	Medium	Bad
Representative value	1	2	3

TABLE 7: Tectonic stress state.

Tectonic stress state	Representative value
Compression and twist	1
Transtension	2
Shear	3
Unobvious	4

TABLE 8: Seismic activity background.

Seismic activity background	Representative value
No seismicity	1
Weak seismicity	2
Moderate seismicity	3
Strong seismicity	4

released via common earthquakes at relatively strong seismic regions. However, RTE is still likely to happen in those areas, under the condition of certain geological structures. The representative value related to the background of seismicity is shown in Table 8.

3.1.6. Grading for the Magnitude of RTE. It is necessary to classify the magnitude of RTE in order to obtain sufficient samples in each class. From the monitoring data on RTE, it is found that the maximum magnitude of RTE is 6.1, denoted in surface wave magnitude (M_s). Combined with the research by Wang et al. [10], the magnitude of RTE is divided into 5 grades, which are strong seismicity ($M_s \geq 6.0$), moderate seismicity ($4.5 \leq M_s < 6.0$), weak

seismicity ($3.0 \leq M_s < 4.5$), microseismicity ($0.0 < M_s < 3.0$), and no seismicity ($M_s = 0.0$). The 5 grades are denoted by 5 growing natural numbers, from 1 to 5, respectively.

It should be noted that, since the number of RTE cases is very limited as 48 ones in China at present, the number of grades should thus be controlled to obtain sufficient number of samples in each class, so that the accuracy of the developed models could be guaranteed.

Some samples (out of the database on 100 reservoirs) on the quantified data of the geological factors and reservoir parameters are shown in Table 9, including earthquake grade classified.

3.2. Numerical Experiments and Results. Since the magnitude of RTE is classified into 5 grades as mentioned above, the SVM and FSVM models are involved in a 5-class classification problem. Considering the limited number of samples, the “one-against-one” method, which can reduce the effect of strongly uneven samples on classifier, is selected to construct the SVM and FSVM models. Consequently, 10 binary classifiers, that is, 1-2 BC, 1-3 BC, 1-4 BC, 1-5 BC, 2-3 BC, 2-4 BC, 2-5 BC, 3-4 BC, 3-5 BC, and 4-5 BC, need to be constructed.

To verify and validate the numerical models established with SVM and FSVM, three numerical tests are carried out. In each test, 90 reservoirs are randomly selected out of the total dataset to provide the data for the training set in the construction of the SVM and FSVM models. Data of the remaining 10 reservoirs are then used as the prediction set to test the accuracy of SVM and FSVM models.

The first training set is used to train the established SVM and FSVM models first. Through analysis and comparison of output results, it is found that the RBF works well as the kernel function. The corresponding values of parameter σ^2 in (3) and regularization parameter C in (2), which could effectively affect the performance of the SVM and FSVM models, are determined through a threefold cross validation method [15].

After the optimization of the parameters, the optimal parameters values of σ^2 and C for all BCs, obtained from the first training set for the SVM and FSVM models, are shown in Table 10. During the process of the threefold cross validation, the accuracy of SVM and FSVM models for the training set and the validation set is shown and compared in Table 11. The accuracy is defined as the matching percentage of the predicted set of labels and the true set of labels in each BC. Based on the optimal parameters, the output of the SVM model is completely identical with the sample data for the training set. While for the validation set, the accuracy remains above 85%, except for 3-4 BC with the accuracy of 61.97%. At the same time, the accuracy of FSVM with the optimal parameters remains above 85% for training set and above 87.59% for the validation set except for 3-4 BC with accuracy of 64.74%. Consequently, under this dataset, the performance of SVM model is slightly superior to that of FSVM model.

Based on the optimal parameters above, the SVM and FSVM models should be trained again by the training set; results show that no error happens out of the 10 BCs for training set, while for the FSVM model, only 2 BCs do not

exist error, the maximum error rate reaching 14.29% and the average error rate amounting to 4.67%.

We then use the trained SVM and FSVM models to predict the magnitude of RTE of the 10 reservoirs in the prediction set as described above. The results obtained from the FSVM and SVM models are shown in Table 12.

It can be seen from Table 12 that, except for Ankang reservoir and Liujiashan reservoir, the prediction results of the 10 reservoirs are identical for SVM and FSVM models in the current experiment. Among them, Chongbahu reservoir, Fengjiashan reservoir, and Bikou reservoir have no reported cases of RTE (denoted by 5 according to the quantification of RTE magnitude as described before). Both of the two models provide accurate results of RTE predictions for those reservoirs. In addition, the two models also provide accurate predictions on the grade of RTE for some of reservoirs with previous RTE cases, including Wujiangdu reservoir, Three Gorges reservoir, and Zipingpu reservoir, as shown in Table 12. While for Ankang reservoir and Liujixia reservoir, where no RTE has been monitored, the prediction grade is grade 3 ($3.0 \leq M_s < 4.5$) by FSVM model for the former and by SVM model for the latter, higher than the actual data, which suggests weak seismicity; whereas the prediction is grade 5 by SVM model for the former and by FSVM model for the latter, identical with monitored data. For Danjiangkou reservoir, previous RTE with the magnitude of $M_s 4.7$ has happened. The prediction grade is grade 3, corresponding to RTE with $3.0 \leq M_s < 4.5$. The prediction grade by the two models is also 3 for Wuxijiang reservoir, where actual RTE with the magnitude of $M_s 2.8$ was observed. To sum up, 7 samples have been predicted accurately by SVM and FSVM models among the 10 prediction samples in the current experiment. The errors for the left 3 samples are within small range.

In order to test the robustness of the models, another two training processes are undertaken with data selected in the same way as mentioned above, called group 2&3. The values of σ^2 and C remain the same as those obtained from the optimization processes with the first training set for the SVM and FSVM models, respectively.

After applying the second training set on the SVM model, the results of training process show that no BC exists error out of the total 10 BCs. The accuracy for the training set reaches to 100%. The trained SVM model is then employed to predict the magnitude of RTE of the 10 reservoirs in the prediction set, for which the results are shown in Table 13. It can be seen that 8 results have been predicted correctly, with the accuracy rate reaching 80%. The error of the results is kept in a small range of 1 grade.

While for the FSVM model, in the training process, only 2 BCs do not exist misjudgment among the 10 BCs. The maximum and average error rates are 14.29% and 4.74%, respectively. The prediction results with FSVM in the prediction set are also shown in Table 13, where it can be seen that only two of the results (on Shenwo reservoir and Dongfeng reservoir) are different from that of SVM by one grade, on which FSVM has provided accurate prediction with the accuracy of 80% as well, in spite of a little bit lower accuracy in the training set than that with the SVM model.

TABLE 9: Quantitative index on reservoirs.

Reservoir name	Height (m)	Storage capacity ($\times 10^8 \text{ m}^3$)	Water depth before dam (m)	Lithology at reservoir area and epicenter	Rock mass grade	Fault type	Fault development degree	Fault activity	Fault permeability	Tectonic stress state	Seismic activity background	Seismicity grade
Fouziling	74	4.7	65	1	3	2	1	1	1	2	3	2
Xinfengjiang	105	115	97	8	3	2	1	1	1	3	2	1
Nanshui	81.5	10.5	75	1	3	3	2	1	1	3	1	4
Nanchong	45	0.15	35	1	3	1	2	1	1	1	1	4
Huangshi	40.5	6.12	36	1	3	1	2	1	1	1	1	4
Qianjiang	50	0.168	44	1	3	1	2	1	1	1	1	3
Tuoling	63.5	71.7	47	1	3	2	1	1	1	2	1	3
Shenwo	50	5	36	1	3	3	2	1	1	3	2	2
Shiquan	65	4.7	53	5	3	1	2	1	1	1	2	3
Tongjiezi	82	2	70	9	3	2	2	1	1	2	2	4
Manwan	126	10.06	100	8	3	3	2	1	1	3	2	3
Shuikou	100	23.4	85	8	3	1	2	1	1	1	2	3
Geheyan	151	34	145	1	3	1	2	3	1	1	2	4
Dongfeng	168	10.3	160	1	3	1	2	3	1	1	2	4
Ertan	240	58	220	9	3	2	2	3	2	2	2	3
Tianshengqiao	178	102.6	155	1	3	1	2	3	3	1	2	3
Daqiao	92.2	6.58	85	8	3	2	2	1	1	2	4	2
Xiaolangdi	154	126.5	148	2	3	3	2	1	2	3	2	3
...					
Xudong	87	3.64	50	2	3	2	2	3	1	2	2	5

TABLE 10: Optimal parameters for respective BCs.

BCs	SVM		FSVM	
	σ^2	C	σ^2	C
1-2 BC	0.0039	2.0	0.0039	2.0
1-3 BC	0.0039	2.0	0.0039	2.0
1-4 BC	0.0039	2.0	0.0039	2.0
1-5 BC	0.0039	2.0	0.0039	2.0
2-3 BC	4.0000	1024	0.3536	128
2-4 BC	1.0000	4.0	1.4142	16.0
2-5 BC	0.5000	2.0	0.7071	2.0
3-4 BC	0.0156	2.0	0.0884	128.0
3-5 BC	0.7071	16.0	0.5000	8.0
4-5 BC	0.5000	16.0	0.5000	4.0

TABLE 11: Optimal output for respective BCs in cross validation test.

BCs	SVM		FSVM	
	Training set Accuracy (%)	Validation set Accuracy (%)	Training set Accuracy (%)	Validation set Accuracy (%)
1-2 BC	100	100	85	100
1-3 BC	100	100	95.69	100
1-4 BC	100	100	93.64	100
1-5 BC	100	100	97.70	100
2-3 BC	100	88.89	96.39	88.89
2-4 BC	100	85.71	97.62	90.48
2-5 BC	100	97.92	97.92	97.92
3-4 BC	100	61.97	100	64.74
3-5 BC	100	93.79	98.25	87.59
4-5 BC	100	94.74	98.25	94.74

TABLE 12: Prediction results of RTE with group 1.

Reservoir name	Observed seismicity grade	Predicted seismicity grade	
		SVM	FSVM
Chongbahu	5	5	5
Ankang	5	5	3
Fengjiashan	5	5	5
Liujiashan	5	3	5
Bikou	5	5	5
Danjiangkou	2	3	3
Wuxijiang	4	3	3
Wujiangdu	3	3	3
Three Gorges	3	3	3
Zipingpu	3	3	3

Based on the training set in group 3, no error is found for the 10 BCs in the trained SVM model. While in the FSVM model, no error is found in 2 BCs. The maximum and average error rates are 14.29% and 4.97%, respectively, for the left

TABLE 13: Prediction results of RTE with group 2.

Reservoir name	Observed seismicity grade	Predicted seismicity grade	
		SVM	FSVM
Xinanjiang	5	5	5
Jinshuitan	5	5	5
Gezhouba	5	5	5
Fengshuba	5	5	5
Tuoling	3	3	3
Shenwo	2	2	3
Dongfeng	4	3	4
Ertan	3	3	3
Panjiakou	4	5	5
Lijiaxia	3	3	3

TABLE 14: Prediction results of RTE with group 3.

Reservoir name	Observed seismicity grade	Predicted seismicity grade	
		SVM	FSVM
Gongzui	5	5	5
Jiangkou	5	5	5
Huanglongtan	5	5	5
Shizitan	5	5	5
Qianjin	3	3	4
Huangshi	4	4	4
Kezier	2	2	2
Yunpeng	3	3	3
Xiaowan	3	3	3
Nanshui	4	2	2

8 BCs. The results of the trained SVM and FSVM models applied in the prediction of the magnitude of 10 RTE in the prediction set are shown in Table 14. It can be seen that the accuracy is 90% for SVM and 80% for FSVM, with errors limited within 1~2 grades for the two models.

Through the results obtained with the SVM and FSVM models in the three tests, it can be seen that both the two models are proved to be effective in the prediction of the magnitude of RTE with high accuracy. Through comparison, it is further found that SVM model is slightly superior to FSVM in the current study on the prediction of the magnitude of RTE with dataset on RTE in China. This is in contradiction with the findings in [24]. In theory, FSVM model is superior to SVM in minimizing the effect of noises or outliers by introducing the fuzzy membership. However, it cannot be concluded that the FSVM model is superior to SVM in general. If the training samples are of high reliability and high discreteness, the introduction of the fuzzy membership might decrease the discreteness of the data and underestimate the effect of samples far away from the optimal hyperplane and thus reduce the accuracy of classifier. This might explain the reason of the findings in the current study in terms of the comparison between the two models. Therefore,

the choice between SVM and FSVM models in the application of classifying problems should be made according to the discreteness of data. For samples with high discreteness, SVM is likely to be superior to FSVM and vice versa.

4. Conclusions

In this study, we have applied the SVM and FSVM methods as an effective approach in the prediction of the magnitude of RTE. Dataset on 100 reservoirs in China, covering the 48 RTE cases, are collected and used as the database. The main 11 influence factors of RTE, including lithology, rock mass integrity, fault characteristic, tectonic stress state, and seismic activity background, are summarized and quantified with the analytical hierarchy process (AHP). Multiple numerical tests are carried out to test the performance of the established SVM and FSVM models in the prediction of the magnitude of RTE. The main conclusions obtained are summarized below.

Both the SVM and FSVM models are found to be effective in the prediction of the magnitude of RTE with high accuracy. Moreover, the results obtained by SVM are of slight higher accuracy than those by FSVM for the current analysis of RTE cases in concern. This might be caused by the introduction of the fuzzy membership in FSVM. It is effective in the minimization of the influence of the noises or outliers in some problems and yet is not favorable when dealing with samples with high discreteness. Therefore, the discreteness of samples needs to be considered when choosing between the SVM and FSVM models in practical problems. Based on the current study, the SVM model is found to be slightly superior to FSVM in the field of RTE prediction.

It should be noted that the magnitude of RTE is analyzed in terms of its maximum value in the present study. In general, besides the main shock, there could be multiple foreshocks and aftershocks in the process of RTE, whose magnitude and epicenter are quite different with those of the main shock. To predict the magnitude of foreshocks and aftershocks, the SVM and FSVM could also be employed; however, the data has to be established based on new data collected on the features of faults and locations of foreshocks and aftershocks. Furthermore, the SVM and FSVM models for the prediction of the magnitude of RTE are constructed based on the mass data of engineering geology and hydrology geology conditions. Thus, in order to obtain prediction results with high accuracy, it is essential to collect reliable data of sufficient samples on the main factors related to RTE, including the geological conditions at the reservoir area and the features of water permeation from reservoir to deep rock.

In addition, the current work is not on the prediction of the time history of RTE, which plays an important role in engineering practices. However, in theory, if we could collect and monitor sufficient data on the evolution of hydrology parameters and geological properties of rocks in a certain reservoir area, obtain their trends via means of numerical analysis, and make the analogy with the time series of RTE cases, the evolution and prediction of RTE in this area could be

studied intensively. Therefore, the current study also provide a creative idea for the prediction of the time history of RTE, while the main challenge lies in the access to reliable evolution trends of the multiple hydrological and geological factors involved.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

The authors are grateful to the reviewers for their very useful comments and suggestions. This research was financially supported by the National Natural Science Foundation of China (Grant no. 51669008).

References

- [1] L. X. Yi, G. Wang, and L. Li, "Hydrogeological structure and reservoir induced seismicity," *Hydrogeology & Engineering Geology*, vol. 31, no. 2, pp. 29–32, 2004.
- [2] T. Chang, "Forming conditions of karst collapsing induced earthquake," *Hydrogeology & Engineering Geology*, vol. 5, pp. 42–45, 2006.
- [3] H. K. Gupta, "A review of recent studies of triggered earthquakes by artificial water reservoirs with special emphasis on earthquakes in Koyna, India," *Earth-Science Reviews*, vol. 58, no. 3–4, pp. 279–310, 2002.
- [4] D. S. Carder, "Seismic investigations in the boulder dam area, 1940 – 1944, and the influence of reservoir loading on earthquake activity," *Bulletin of the Seismological Society of America*, vol. 35, pp. 175–192, 1945.
- [5] D. Chen, Y. Wang, and X. Zeng, "A study of reservoir-induced earthquake of three gorges project," *Chinese Journal of Rock Mechanics and Engineering*, vol. 27, no. 8, pp. 1513–1524, 2008.
- [6] Y. Ding, *The Reservoir Induced Earthquake*, Geological press, Beijing, 1989.
- [7] W. Ding, *Mechanical Problems in Earthquake Prediction*, Qinghua University Press, Beijing, 2012.
- [8] S. J. Kenner and P. Segall, "A mechanical model for intraplate earthquakes: application to the New Madrid Seismic Zone," *Science*, vol. 289, no. 5488, pp. 2329–2332, 2000.
- [9] J. Su, Q. Xia, Y. Wang, X. Wang, and Z. Ju, "Preliminary prediction on Datengxia reservoir induced seismicity," *The Chinese Journal of Geological Hazard and Control*, vol. 4, no. 4, pp. 42–51, 2009.
- [10] Q. Wang, Y. Yao, J. Xia, J. Li, and D. Wang, "Application of statistical forecasting models to the prediction of the three gorges reservoir induced seismicity," *Seismology and Geology*, vol. 31, no. 2, pp. 287–194, 2009.
- [11] J. Xia, C. Li, and X. Wang, "Application of neural network theory in prediction of earthquake induced by TGP reservoir," *Yangtze River*, vol. 38, no. 2, pp. 8–10, 2007.

- [12] G. Habibagahi, "Reservoir induced earthquakes analyzed via radial basis function networks," *Soil Dynamics and Earthquake Engineering*, vol. 17, no. 1, pp. 53–56, 1998.
- [13] P. Samui and D. Kim, "Applicability of artificial intelligence to reservoir induced earthquakes," *Acta Geophysica*, vol. 62, no. 3, pp. 608–619, 2014.
- [14] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York, NY, USA, 1995.
- [15] V. N. Vapnik, *Statistical Learning Theory*, Wiley, New York, NY, USA, 1998.
- [16] N. Deng and Y. Tian, *Support Vector Machine-Theory, Algorithm and Expansion*, Science Press, Beijing, 2009.
- [17] S. K. Das, P. Samui, and A. K. Sabat, "Prediction of field hydraulic conductivity of clay liners using an artificial neural network and support vector machine," *International Journal of Geomechanics*, vol. 12, no. 5, pp. 606–611, 2012.
- [18] M. Pal and S. Deswal, "Modeling pile capacity using support vector machines and generalized regression neural network," *Journal of Geotechnical and Geoenvironmental Engineering*, vol. 134, no. 7, pp. 1021–1024, 2008.
- [19] J. Lin and C. Cheng, "Application of support vector machine method to long-term runoff forecast," *Journal of Hydraulic Engineering*, vol. 37, no. 6, pp. 681–686, 2006.
- [20] H. Zhao, *Support Vector Machine Analysis in Rock and Soil Mechanics and Engineering*, Coal Industrial Press, Beijing, 2008.
- [21] K. Gopalakrishnan and S. Kim, "Support vector machines approach. to HMA stiffness prediction," *Journal of Engineering Mechanics*, vol. 137, no. 2, pp. 138–146, 2011.
- [22] C.-F. Lin and S. Wang, "Fuzzy support vector machines," *IEEE Transactions on Neural Networks*, vol. 13, no. 2, pp. 464–471, 2002.
- [23] H. Tang, Y. Liao, F. Sun, and H. Xie, "Fuzzy support vector machine with a new fuzzy membership function," *Journal of Xi'an Jiao Tong University*, vol. 43, no. 7, pp. 40–43, 2009.
- [24] Z. Du, S. Liu, and X. Qi, "Fuzzy support vector machine with new membership function," *Journal of System Simulation*, vol. 21, no. 7, pp. 1901–1903, 2009.
- [25] S. Ding and Y. Gu, "A Fuzzy Support Vector machine algorithm with dual membership based on hypersphere," *Journal of Computational Information Systems*, vol. 7, no. 6, pp. 2028–2034, 2011.
- [26] E. C. C. Tsang, D. S. Yeung, and P. P. K. Chan, "Fuzzy support vector machines for solving two-class problems," in *Proceedings of the 2003 International Conference on Machine Learning and Cybernetics (IEEE Cat. No.03EX693)*, pp. 1080–1083, Xi'an, China, November 2003.
- [27] S. Abe, "Fuzzy support vector machines for multilabel classification," *Pattern Recognition*, vol. 48, no. 6, pp. 2110–2117, 2015.
- [28] T. L. Satty, *The Analytic Hierarchy Process*, McGraw Hill, Inc., New York, NY, USA, 1980.
- [29] J. Valyon and G. Horvath, "Selection methods for extended least squares support vector machines," *International Journal of Intelligent Computing and Cybernetics*, vol. 1, no. 1, pp. 69–93, 2008.
- [30] H. Jiang, X. Zhang, and X. Shan, *Research on Prediction Method and Statistic Characteristics of Reservoir Induced Earthquake in China Mainland*, Seismic Press, Beijing, 2014.




Hindawi

Submit your manuscripts at
www.hindawi.com

