Hindawi Publishing Corporation
Abstract and Applied Analysis
Volume 2014, Article ID 459137, 8 pages
http://dx.doi.org/10.1155/2014/459137

Research Article

Nonlinear Methodologies for Identifying Seismic Event and Nuclear Explosion Using Random Forest, Support Vector Machine, and Naive Bayes Classification

Longjun Dong,[1] Xibing Li,[1] and Gongnan Xie[2]

[1] *School of Resources and Safety Engineering, Central South University, Changsha 410083, China*
[2] *School of Mechanical Engineering, Northwestern Polytechnical University, Xi'an 710129, China*

Correspondence should be addressed to Longjun Dong; rydong001@csu.edu.cn

Received 26 December 2013; Accepted 16 January 2014; Published 26 February 2014

Academic Editor: Carlo Cattani

The discrimination of seismic event and nuclear explosion is a complex and nonlinear system. The nonlinear methodologies including Random Forests (RF), Support Vector Machines (SVM), and Naïve Bayes Classifier (NBC) were applied to discriminant seismic events. Twenty earthquakes and twenty-seven explosions with nine ratios of the energies contained within predetermined "velocity windows" and calculated distance are used in discriminators. Based on the one out cross-validation, ROC curve, calculated accuracy of training and test samples, and discriminating performances of RF, SVM, and NBC were discussed and compared. The result of RF method clearly shows the best predictive power with a maximum area of 0.975 under the ROC among RF, SVM, and NBC. The discriminant accuracies of RF, SVM, and NBC for test samples are 92.86%, 85.71%, and 92.86%, respectively. It has been demonstrated that the presented RF model can not only identify seismic event automatically with high accuracy, but also can sort the discriminant indicators according to calculated values of weights.

1. Introduction

The problems of seismic source locations and identifications are two of the most important and fundamental issues in earthquake monitoring, microseismic monitoring, analyses of active tectonics, and assessment of seismic hazards [1–4].

Seismic analysts identify seismic signals from those of explosions or blasts by visual inspection and by calculating some characteristics of seismogram. As recorded quarry blasts or nuclear explosions can mislead scientists interpreting the active tectonics and lead to erroneous results in the analysis of seismic hazards in the area; an event classification task is an important step in seismic signal processing. Such task analyses data in order to find to which class each recorded event belongs.

Such work supposes a great deal of workload for seismic analysts. Therefore, an automatic classification tool is necessary to be developed for reducing dramatically this arduous task, turning classification as reliable, as well as removing errors associated with tedious evaluations and change of personnel.

Most discrimination methods are designed for a particular source region and a particular distance of the recording station from the epicenter [5]. Some of them heavily depend on the heterogeneity of the uppermost crust in the sense that they might be effective only for a given region.

The widely used methods for discriminators include simulating explosion spectra in order to predict spectral details indicative of explosions and not of earthquakes or single-event explosions [6, 7]; examining compressional and shear-wave ratios (amplitude and spectral) between all types of explosions and earthquakes, in an attempt to apply the basic physical conclusion that explosions excite more compressional waves than earthquakes relative to shear waves [8–11]; differences in high frequency S-to-P ratios between all types of explosions and earthquakes [12–14];

analyzing observed spectra of ripple-fired explosions, instantaneous explosions, and earthquakes and contrasting time-independent modulations, path-independent modulations, spectral ratios, spectral slopes, and spectral maxima and minima [15–17]; and examining differences in energy ratios of various wave in velocity windows [18, 19].

However, most of developed methods above are based on single index or liner discriminant methods. And the methods seem to fail to capture the discontinuities, the nonlinearities, and the high complexity of wave series.

Random Forests (RFs), Support Vector Machines (SVMs), and Naive Bayes Classifier (NBC) provide enough learning capacity and are more likely to capture the complex nonlinear models, which are widely used in natural and science areas, including medicine, agriculture, and geotechnics.

So far, as to our knowledge, the RFs and SVMs were not used for seismic classification. The performance of RFs, SVMs, and NBC in this type of application has not been thoroughly compared.

In present work, RF, SVM, and NBC were applied to discriminate between earthquakes and nuclear explosions. And based on the one out cross-validation, ROC curve, and test accuracy, their discriminating performances were discussed and compared.

2. Materials and Methods

2.1. Materials. The measurements or parameters consist of ratios of the "high energies" contained within predetermined "velocity windows" on the seismograms [18]. The choice of velocity windows is guided by the assumption that earthquake source mechanism is extended both in time and space and generates a larger fraction of energy in shear waves as compared to explosion source mechanism.

The different waves of "velocity windows" are listed as follows:

(i) P_1: first arrival to 4.6 km/s;

(ii) S_1: arrival to 4.6 to 2.5 km/s;

(iii) S_2: first arrival to 4.9 km/s;

(iv) S_2: arrival to 4.9 to 2.0 km/s;

(v) P_g: arrival to 6.2 to 4.9 km/s;

(vi) B: arrival to 4.9 to 3.6 km/s;

(vii) L_{g1}: arrival to 3.6 to 3.2 km/s;

(viii) R_{g1}: arrival to 3.2 to 2.8 km/s; and

(ix) R: arrival to 2.8 to 2.5 km/s.

The factors, including ratios P_1/S_1, P_2/S_2, P_g/B, P_g/L_{g1}, $P_g/(R_{g1} + R)$, $P_g/(L_{g1} + R_{g1})$, R/R_{g1}, R_{g1}/L_{g1}, and R_{g2}/L_{g2}, as well as Average Distance, were expressed as $Ratio_1$, $Ratio_2$, $Ratio_3$, $Ratio_4$, $Ratio_5$, $Ratio_6$, $Ratio_7$, $Ratio_8$, $Ratio_9$, $Ratio_{10}$, and AD, respectively.

Nine ratios of energies included within certain velocity windows have been computed for 20 earthquakes and 27 nuclear explosions by Booker and Mitronovas [18]. All seismograms were recorded by the VELA UNIFORM LRSM Network on short-period Benioff instruments [18]. $Ratio_1$, $Ratio_2$, $Ratio_3$, $Ratio_4$, $Ratio_5$, $Ratio_6$, $Ratio_7$, $Ratio_8$, $Ratio_9$, and AD were selected as discriminant indicators. z-score is used to standardize variables in this work. First, the mean is subtracted from the value for each case, resulting in a mean of zero. Then, the difference between the individual's score and the mean is divided by the standard deviation, which results in a standard deviation of one. If we start with a variable x and generate a variable x^*, the process is

$$x^* = \frac{(x - m)}{\text{sd}}, \tag{1}$$

where m is the mean of x and sd is the standard deviation of x. z-score of each ratio and distance for seismic event and nuclear earthquake were listed in Tables 1 and 2, respectively.

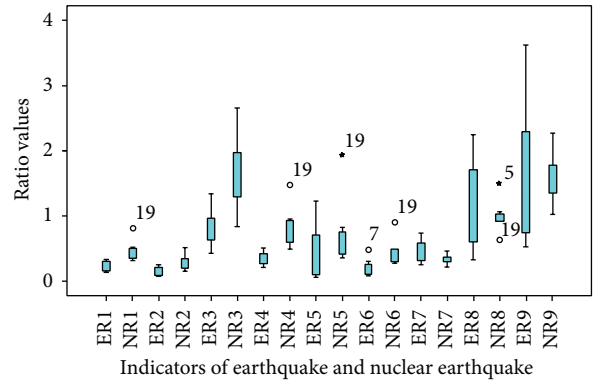

Figure 1: Box plot graph showing the distribution of values of ratios of energies (ER1, ER2, ER3, ER4, ER5, ER6, ER7, ER8, and ER9 represented $Ratio_1$, $Ratio_2$, $Ratio_3$, $Ratio_4$, $Ratio_5$, $Ratio_6$, $Ratio_7$, $Ratio_8$, and $Ratio_9$ for earthquake, respectively; NR1, NR2, NR3, NR4, NR5, NR6, NR7, NR8, and NR9 represented $Ratio_1$, $Ratio_2$, $Ratio_3$, $Ratio_4$, $Ratio_5$, $Ratio_6$, $Ratio_7$, $Ratio_8$, and $Ratio_9$ for nuclear earthquake, resp.).

Box plot graphs of energy ratios and distance were plotted in Figures 1 and 2, respectively. Each group is represented as a box whose top and bottom are drawn at the lower and upper quartiles, with a small square at the median. Thus, the box contains the middle half of the scores in the distribution. Vertical lines outside the box extend to the largest and the smallest observations within 1.5 interquartile ranges. We conclude that $Ratio_1$, $Ratio_2$, $Ratio_3$, $Ratio_4$, $Ratio_5$, $Ratio_6$, $Ratio_7$, $Ratio_8$, $Ratio_9$, $Ratio_{10}$, and AD for earthquake and nuclear earthquake are obviously different. Such it is reasonable to select the ten factors as discriminating indicator.

2.2. Methodologies. The first 70% dataset of earthquake and nuclear earthquake were used to establish discriminating models and the other 30% dataset were used to test the model.

2.2.1. Overview of Random Forest. Random Forest (RF), a metalearner comprising many individual trees, was first developed by Tin Kam Ho in 1995 and later improved by Breiman in 2001. It was developed to operate quickly over large datasets and to be diverse by using random samples to build each tree in the forest. Each tree depends on the values

Table 1: Earthquakes dataset.

No.	Earthquakes	X_1 $Ratio_1$	X_2 $Ratio_2$	X_3 $Ratio_3$	X_4 $Ratio_4$	X_5 $Ratio_5$	X_6 $Ratio_6$	X_7 $Ratio_7$	X_8 $Ratio_8$	X_9 $Ratio_9$	X_{10} AD
1	Baja Calif	0.123	0.495	−0.570	−0.145		−0.498		1.023	0.547	1.023
2	Baja Calif	−0.274	0.130	−0.207	−0.291		−0.824		2.904	1.407	2.031
3	Boxelder Creek	0.515	0.741	0.624	0.471		0.531		−0.652	0.057	0.576
4	Bridgeport	−0.371	−0.101	−0.681	−0.994	0.040	−0.686	−0.531	−1.543	−0.833	−0.614
5	Cache Creek	−0.803	−1.188	−1.082	−1.038	−0.860	−1.001	−0.750	−0.393	−0.875	2.114
6	Cache Creek AS	−0.349	−0.157	−0.866	−0.771	−0.426	−0.578	−0.876	−0.910	−0.356	0.259
7	Colona	−0.309	0.011	−0.430	−0.420	0.710	0.416	0.308	−1.453	−1.047	0.108
8	Mont.-Wyo Border	−0.446	0.243	−0.873	−0.462		−0.007		−0.661	−0.368	0.825
9	Pierre, S.Dakota	−1.354	−1.300	−1.574	−1.284		−1.131		−1.075	−0.719	2.498
10	Red Rock River	0.383	0.628	−0.231	0.120		0.315		−0.720	−0.579	0.945
11	Sierra De Juarez	−1.090	−0.992	0.624	−0.931		−1.004		1.102	4.306	1.002
12	Teton County	0.343	0.502	0.541	−0.164		0.044		−0.700	−0.442	−0.001
13	Western Mary land	−1.575	−1.546	−0.828	−1.365		−1.228		0.867	−0.459	−0.963
14	Western Vermont	−1.672	−1.602	−2.047	−1.438		−1.282		−1.472	−1.048	2.956
15	Western Vermont	−1.138	−1.034	−1.323	−0.779	−1.246	−0.896	2.417	1.004	0.238	1.184
16	Western Vermont	−1.010	−0.929	−1.538	−0.727	−1.204	−0.932	2.919	2.277	0.719	0.612
17	Western Vermont	−0.975	−0.844	−0.779	−1.023	−0.036	−0.751	0.018	−0.810	−0.842	−0.765
18	Western Vermont	0.013	0.046	0.116	−0.416	0.955	0.018	−0.758	−1.298	−0.902	−0.329
19	Western Vermont	−1.164	−1.202	−0.823	−0.754	−1.164	−0.925	1.084	1.480	1.614	0.950
20	Western Vermont	−0.494	−0.290	0.027	−0.462	−0.312	−0.245	0.567	−0.894	−0.476	1.023

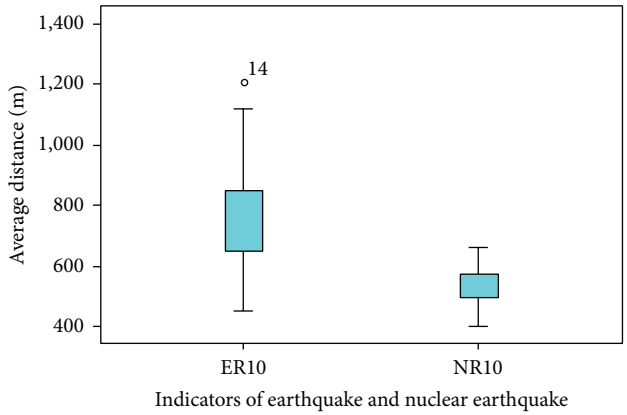

Figure 2: Box plot graph showing the distribution of values of the distance between hypocenter and sensors (ER10 and NR10 represented distance between hypocenter and sensors for earthquake and nuclear earthquake, resp.).

of a random vector sampled independently and with the same distribution for all trees in the forest. The generalization error of a forest of tree classifiers depends on the strength of the individual trees in the forest and the correlation between them [20]. Comprehensive review of applications of Random Forest have been provided by Rodriguez-Galiano et al., [21], Granitto et al. [22], and by Genuer et al. [23]. Also, a number of researches have compared the performance of other data mining technique and Random Forest in different kinds of problems [23–26]. The theory of RF is summarized as follows [20].

A Random Forest is a classifier consisting of a collection of tree-structured classifiers $\{h(x, \Theta_k), k = 1, \ldots\}$, where the $\{\Theta_k\}$ are independent identically distributed random vectors and each tree casts a unit vote for the most popular class at input x [18].

Given an ensemble of classifiers $h_1(X), h_2(X), \ldots, h_K(X)$ and with the training set drawn at random from the distribution of the random vector Y, X, define the margin function as

$$\mathrm{mg}(X, Y) = av_K I\left(h_K(X) = Y\right) - \max_{j \neq Y} av_K I\left(h_K(X) = j\right), \tag{2}$$

where $I(\cdot)$ is the indicator function. The margin measures the extent to which the average number of votes at X, Y for the right class exceeds the average vote for any other class. The larger the margin, the more confidence in the classification. The generalization error is given by

$$PE^* = P_{X,Y}\left(\mathrm{mg}(X, Y) < 0\right), \tag{3}$$

where the subscripts X, Y indicate that the probability is over the X, Y space. In Random Forests, $h_K(X) = h(X, k)$. For a large number of trees, it follows from the Strong Law of Large Numbers and the tree structure the following.

As the number of trees increases, for almost surely all sequences $\Theta_1, \ldots, PE^*$ converges to

$$P_{X,Y}\left(P_\Theta\left(h(X, \Theta) = Y\right) - \max_{j \neq Y} P_\Theta\left(h(X, \Theta) = j\right) < 0\right). \tag{4}$$

Table 2: Nuclear explosion dataset.

No.	Nuclear explosion	X_1 $Ratio_1$	X_2 $Ratio_2$	X_3 $Ratio_3$	X_4 $Ratio_4$	X_5 $Ratio_5$	X_6 $Ratio_6$	X_7 $Ratio_7$	X_8 $Ratio_8$	X_9 $Ratio_9$	X_{10} AD
1	Aardvark	1.027	1.575	2.096	1.015	0.342	0.903	−0.523	−0.334	−0.321	−0.313
2	Agouti	−0.067	−0.101	0.984	−0.124		−0.047		−0.054	−0.192	−0.547
3	Armadillo	1.164	1.042	0.761	2.079	0.202	1.315	0.535	1.102	0.324	−0.516
4	Chinchilla II	0.317	0.194	0.197	1.222		0.575		0.338	0.384	−0.562
5	Cimarron	0.335	0.095	−0.009	0.185	−0.579	−0.065	−0.515	0.827	0.093	−0.329
6	Codsaw	−0.402	−0.676	−0.830	−0.378	−0.731	−0.372	0.755	−0.152	0.453	−0.859
7	Danny Boy	−0.388	−0.480	−0.096	−0.458	−0.736	−0.426	0.120	−0.207	−0.209	−1.202
8	Des Moines	1.133	0.874	0.590	0.954		0.759		−0.074	−0.149	−0.646
9	Dormouse II	0.171	−0.290	0.094	−0.182		−0.278		1.160	−0.252	−0.640
10	Fisher	−0.314	−0.417	0.077	−0.136	−0.565	−0.332	−0.782	0.886	0.117	−0.802
11	Gnome	2.178	2.452	1.087	2.319	1.431	2.688	−1.393	−0.914	0.100	−0.656
12	Hardhat	3.293	3.321	2.319	3.245	2.420	3.505	−0.625	−0.740	−0.562	−0.396
13	Haymaker	2.495	0.965	2.235	2.208	1.889	2.276	−0.782	−0.399	−0.583	−0.074
14	Mad	−0.675	−0.746	−0.282	−0.527	−0.906	−0.545	0.206	−0.362	−0.934	−0.833
15	Madison	0.361	0.635	1.172	0.414	0.023	0.470	−0.452	−0.319	−0.020	−0.370
16	Marshmallow	−0.031	−0.150	0.094	−0.099	−0.530	−0.144	−0.170	−0.015	−0.356	−0.375
17	Mississippi	0.546	0.762	1.052	0.508	−0.178	0.427	−0.225	−0.221	−0.307	−0.084
18	Packrat	−0.177	−0.683	−0.657	−0.339	−1.351	−0.361		−0.230	−0.313	−1.041
19	Pampas	1.768	1.898	2.267	1.597	1.900	1.915	−1.119	−0.879	−0.614	0.134
20	Passaic	0.352	0.586	0.659	0.131	−0.257	0.156	0.002	−0.283	−0.227	−0.090
21	Platte	−0.181	−0.311	0.008	−0.197		−0.206		−0.211	0.014	−1.171
22	Scaramento	−0.221	−0.431	−0.436	0.145	−0.183	−0.202		0.063	−0.023	−0.599
23	Small Boy	−0.018	0.264	0.111	0.370		0.231	0.567	0.181	−0.356	−0.765
24	Stillwater	−0.622	−0.929	−0.916	−0.420		−0.643		1.170	0.926	−0.931
25	Stoat	−0.014	−0.627	−0.334	−0.199		−0.408		1.944	3.308	−1.311
26	Witchita	−0.468	−0.571	−0.300	−0.443		−0.441		−0.279	−0.330	−0.583
27	York	0.083	0.137	−0.026	−0.018		−0.086		−0.074	0.083	0.129

The margin function for a Random Forest is

$$\mathrm{mr}(X,Y) = P_\Theta(h(X,\Theta) = Y) - \max_{j \neq Y} P_\Theta\left(h_K(X,\Theta) = j\right), \tag{5}$$

and the strength of the set of classifiers $\{(h(X,\Theta)\}$ is

$$s = E_{X,Y}\mathrm{mr}(X,Y). \tag{6}$$

Assume that $s \geq 0$, Chebychev's inequality gives

$$PE^* \leq \frac{\mathrm{var}(\mathrm{mr})}{s^2}. \tag{7}$$

A more revealing expression for the variance of mr is derived in the following. Let

$$\hat{j}(X,Y) = \arg\max_{j \neq Y} P_\Theta\left(h(X,\Theta) = j\right) \tag{8}$$

so

$$\begin{aligned}\mathrm{mr}(X,Y) &= P_\Theta(h(X,\Theta) = Y) - P_\Theta\left(h(X,\Theta) = \hat{j}(X,Y)\right) \\ &= E_\Theta\left[(I(X,\Theta) = Y)\right. \\ &\quad \left. - I\left(h(X,\Theta) = \hat{j}(X,Y)\right)\right].\end{aligned} \tag{9}$$

The raw margin function is

$$\mathrm{rmg}(\Theta,X,Y) = I(h(X,\Theta) = Y) - I\left(h(X,\Theta) = \hat{j}(X,Y)\right). \tag{10}$$

Thus, $\mathrm{mr}(X,Y)$ is the expectation of $\mathrm{rmg}(\Theta,X,Y)$ with respect to Θ. For any function f the identity

$$\left[E_\Theta f(\Theta)\right]^2 = E_{\Theta,\Theta'} f(\Theta) f\left(\Theta'\right) \tag{11}$$

holds where Θ, Θ' are independent with the same distribution, implying that

$$\mathrm{mr}(X,Y)^2 = E_{\Theta,\Theta'}\mathrm{rmg}(\Theta,X,Y)\,\mathrm{rmg}\left(\Theta',X,Y\right). \tag{12}$$

Using (12) gives

$$\begin{aligned}\mathrm{var}(\mathrm{mr}) &= E_{\Theta,\Theta'}\left(\mathrm{cov}_{X,Y}\,\mathrm{rmg}(\Theta,X,Y)\,\mathrm{rmg}\left(\Theta',X,Y\right)\right) \\ &= E_{\Theta,\Theta'}\left(\rho\left(\Theta,\Theta'\right)\mathrm{sd}(\Theta)\,\mathrm{sd}\left(\Theta'\right)\right),\end{aligned} \tag{13}$$

where $\rho(\Theta, \Theta')$ is the correlation between $\mathrm{rmg}(\Theta, X, Y)$ and $\mathrm{rmg}(\Theta', X, Y)$ holding Θ, Θ' fixed and $\mathrm{sd}(\Theta)$ is the standard deviation of $\mathrm{rmg}(\Theta, X, Y)$ holding Θ fixed. Then,

$$\begin{aligned} \mathrm{var}\,(\mathrm{mr}) &= \overline{\rho}\left(E_{\Theta}\,\mathrm{sd}\,(\Theta)\right)^2 \\ &\leq \overline{\rho} E_{\Theta} \nu\,(\Theta)\,, \end{aligned} \tag{14}$$

where $\overline{\rho}$ is the mean value of the correlation; that is,

$$\overline{\rho} = \frac{E_{\Theta,\Theta'}\left(\rho\left(\Theta, \Theta'\right)\mathrm{sd}\,(\Theta)\,\mathrm{sd}\left(\Theta'\right)\right)}{E_{\Theta,\Theta'}\left(\mathrm{sd}\,(\Theta)\,\mathrm{sd}\left(\Theta'\right)\right)}. \tag{15}$$

Write

$$E_{\Theta}\mathrm{var}\,(\Theta) \leq E_{\Theta}\left(E_{X,Y}\mathrm{rmg}\,(\Theta, X, Y)\right)^2 - S^2 \leq 1 - S^2. \tag{16}$$

In this work, A RF model of discriminating between natural earthquake and nuclear earthquake is established with optimal 5000 NT trees and 8 variables in rode. In the developed RF model, the calculated weighted values of the $Ratio_1$, $Ratio_2$, $Ratio_3$, $Ratio_4$, $Ratio_5$, $Ratio_6$, $Ratio_7$, $Ratio_8$, $Ratio_9$, and AD are 1.2713, 0.1034, 0.0759, 0.3093, 0.3432, 0.1782, 0.2536, 0.0943, 0.2463, and 0.1512, respectively.

2.2.2. SVM Algorithm. The original SVM algorithm was invented by Vladimir N. Vapnik and the current standard incarnation (soft margin) was proposed by Cortes and Vapnik in 1995 [27].

SVM models were originally defined for the classification of linearly separable classes of objects. For any original separable set of two-class objects SVM are able to find the optimal hyperplanes that separates them providing the bigger margin area between the two hyperplanes. Furthermore they can also be used to separate classes that cannot be separated with a linear classifier.

The feature space in which every object is projected is a high dimensional space in which the two classes can be separated with the linear classifier. The effectiveness of SVM depends on the selection of kernel, the kernel's parameters, and soft margin parameter C.

In the present work we used the Radial Basis Function (RBF) as Kernel functions for the SVM models because of its efficiency in providing very high performance classification results. The optimal RBF parameters C and gamma were found to be 9 and 0.6, respectively, reassuring that the model does not over fit.

2.2.3. Naive Bayes Classier. The Naive Bayes Classier produces a very efficient probability estimation based on a simply structure, requiring a small amount of training data to estimate the parameters necessary for the classification. Its construction relies on two main assumptions: independency of features and absence of hidden or latent attributes.

An advantage of Naive Bayes is that it only requires a small amount of training data to estimate the parameters (means and variances of the variables) necessary for classification. Because independent variables are assumed, only the variances of the variables for each class need to be determined and not the entire covariance matrix.

The aim of the NBC, as with other classifiers, is to assign an object I to one of a discrete set of categories $C_1, C_2, \ldots, C_m$ based on its observable attributes $X_1, X_2, \ldots, X_n$. The NBC calculates the probability that I belongs to each category, conditioning on the observed attributes; I is typically assigned to the category with the greatest probability. This classifier is naive in the sense that it makes the strong assumption that the attributes are mutually conditionally independent; that is, the conditional probability that I belongs to a particular class given the value of some attribute is independent of the values of all other attributes. Despite this unrealistic assumption, empirical studies demonstrate that this assumption does not need to significantly compromise the accuracy of the prediction, and NBCs are used in a variety of applications, including document classification [28], medical diagnosis [29, 30], systems performance management [31], probability classification of rockburst [32], and other fields. Domingos and Pazzani [33] prove optimality of the NBC under certain conditions even when the conditional independence assumption is violated.

In this paper, the prior probability of natural earthquake and nuclear earthquake is calculated according to the size of data. The prior probabilities of earthquake and nuclear earthquake are 0.424 and 0.576, respectively.

The discriminate functions for the earthquake and nuclear are

$$\begin{aligned} Y_1 = {} & 0.45X_1 + 2.72X_2 - 1.28X_3 - 4.6X_4 + 1.61X_5 + 0.34X_6 \\ & + 0.03X_7 - 0.17X_8 + 0.68X_9 + 1.65X_{10} - 1.92, \\ Y_2 = {} & -0.87X_1 - 2.28X_2 + 1.44X_3 + 4.34X_4 - 2.69X_5 \\ & + 0.31X_6 - 2.04X_7 + 0.11X_8 + 0.68X_9 - 0.76X_{10} \\ & - 1.73. \end{aligned} \tag{17}$$

If $Y_1 > Y_2$, the record is an earthquake, otherwise a nuclear event.

2.2.4. Classification Performance. ROC is a graphical plot which illustrates the performance of a binary classifier system as its discrimination threshold is varied [34]. It is created by plotting the fraction of true positives out of the positives (TPR = true positive rate) versus the fraction of false positives out of the negatives (FPR = false positive rate), at various threshold settings.

ROC analysis provides tools to select possibly optimal models and to discard suboptimal ones independently from (and prior to specifying) the cost context or the class distribution. ROC analysis is related in a direct and natural way to cost/benefit analysis of diagnostic decision making.

In this study, seismic event and nuclear explosion were considered as a two-class prediction problem (binary classification), in which the outcomes were labeled either as positive (p, events) or negative (n, blasts). There are four possible outcomes from a binary classifier. If the outcome from a prediction is p' and the actual value is also p, then it is called a true positive (TP); however, if the actual value is n then it is said to be a false positive (FP). Conversely, a true negative

Table 3: Contingency matrix for two class prediction problem.

	Actual value		Total
	P	N	
Prediction outcome			
p'	True Positive (TP)	False Positive (FP)	P
n'	False Negative (FN)	True Negative (TN)	N
Total	P	N	

(TN) has occurred when both the prediction outcome and the actual value is n and false negative (FN) is when the prediction outcome is n', while the actual value is p.

An experiment from P positive and N negative was defined, for instance. The four outcomes can be formulated in a 2×2 contingency table or confusion matrix, as follows in Table 3.

The specificity or true negative rate (TNR) is defined as the percentage of seismic record which is correctly identified as being blast:

$$\text{TNR} = \frac{\text{TN}}{\text{TN} + \text{FP}}. \tag{18}$$

The quantity 1-specificity is the false positive rate (FPR) and is the percentage of seismic records that are incorrectly identified as being blasts. The sensitivity or true positive rate (TPR) is defined as the percentage of seismic records which is correctly identified as being events:

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{19}$$

The accuracy (ACC) can be expressed as

$$\text{ACC} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{FP} + \text{FN}}. \tag{20}$$

3. Results and Discussions

The back-test classification for training samples is calculated using established models. The back-test accuracies of RF, SVM, and NBC are 100%, 100%, and 96.97% for training samples, respectively. The one out cross-validation method was used to validate the methods. Results show that accuracies of RF, SVM (RBF), SVM (liner), and NBC are 100%, 96.97%, and 84.88%, respectively.

The ROC curve is also used to verify and compare the discriminating performance of established models. The established RF model, SVM model, and NBC model were applied to both the training and test samples. The ROC curve is shown in Figure 3. The area under the curve is listed in Table 4. The classification results of test samples using all developed models are presented in Table 5.

In Figure 3, the closer a result from a contingency table is to the upper left corner, the better it predicts, but the distance from the random guess line in either direction or area under curve is the best indicator of how much predictive power a method has.

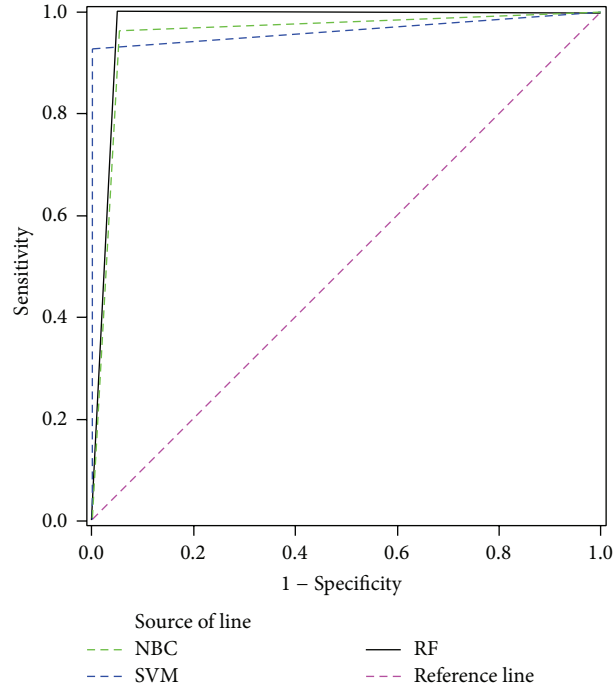

Figure 3: ROC of established RF, SVM (RBF), SVM (liner), and NBC models.

As shown in Figure 3 and Table 4, the result of RF method clearly shows the best predictive power with a maximum area of 0.975 among RF, SVM, and NBC. The result of SVM (area: 0.963) is better than that of NBC (area: 0.956).

According to Table 5, we can get the discriminant accuracy of RF, SVM, and NBC for test samples; their accuracy are 92.86%, 85.71%, and 92.86%, respectively. From back test results, one out cross-validation, ROC, and test results, we get the conclusion that RF discriminant model has the best accuracy and discriminant ability. Also, according to weighted values of RF, the most important factors are $Ratio_1$, followed by $Ratio_5$, $Ratio_4$, $Ratio_7$, $Ratio_9$, $Ratio_6$, AD, $Ratio_2$, $Ratio_8$, and $Ratio_3$.

4. Conclusions

RF, SVM, and NBC were applied to seismic event identification. A thorough investigation of the discrimination capabilities of the techniques were undertaken using seismograms from 20 earthquakes and 27 nuclear explosions. Ratios P_1/S_1, P_2/S_2, P_g/B, P_g/L_{g1}, $P_g/(R_{g1} + R)$, $P_g/(L_{g1} + R_{g1})$, R/R_{g1}, R_{g1}/L_{g1}, and R_{g2}/L_{g2} within certain velocity windows, as well as average distance, were selected as discriminant indicators.

The classification performance of RF, SVM, and NBC was analyzed and compared based on back test of training samples, one out cross-validation, and ROC curve. The result of RF method clearly shows the best predictive power with a maximum ROC area of 0.975 among RF, SVM, and NBC. The result of SVM (area: 0.963) is better than that of NBC (area: 0.956). Test results show the discriminant accuracies

Table 4: Area under the curve.

Test result variable(s)	Area	Std. error[a]	Asymptotic Sig.[b]	Asymptotic 95% confidence interval	
				Lower bound	Upper bound
NBC	0.956	0.035	0.000	0.887	1.026
SVM	0.963	0.030	0.000	0.904	1.022
RF	0.975	0.029	0.000	0.919	1.031

[a]Under the nonparametric assumption.
[b]Null hypothesis: true area = 0.5.

Table 5: Discriminant results of test samples.

No.	Event	Measured results	NBC			SVM result	RF		
			Prob (N)	Prob (NE)	Result		Prob (N)	Prob (NE)	Result
1	NE20	2	0.1646	0.8354	2	2	0.2354	0.7646	2
2	NE21	2	0.0276	0.9724	2	2	0.018	0.982	2
3	NE22	2	0.0041	0.9959	2	2	0.1572	0.8428	2
4	NE23	2	0.0144	0.9856	2	2	0.0178	0.9822	2
5	NE24	2	0.2324	0.7676	2	2	0.3342	0.6658	2
6	NE25	2	0.4416	0.5584	2	2	0.0744	0.9256	2
7	NE26	2	0.2447	0.7553	2	1*	0.1576	0.8424	2
8	NE27	2	0.7336	0.2664	1*	1*	0.4752	0.5248	2
9	E15	1	0.998	0.002	1	1	0.8934	0.1066	1
10	E16	1	0.9991	0.0009	1	1	0.8214	0.1786	1
11	E17	1	0.8956	0.1044	1	1	0.7562	0.2438	1
12	E18	1	0.9807	0.0193	1	1	0.4076	0.5924	2*
13	E19	1	0.9277	0.0723	1	1	0.8904	0.1096	1
14	E20	1	0.9725	0.0275	1	1	0.9216	0.0784	1

Note: result with "*" is incorrect.

of RF, SVM, and NBC are 92.86%, 85.71% and 92.86%, respectively.

From back-test results, one out cross-validation, ROC curve, and test results, we get the conclusion that RF discriminant model has the best accuracy and discriminant ability. Not only can RF discriminant method be applied to seismic identification with high accuracy, but also it can give the weighted sorts of discriminant indicators. In this study, the most important factors are $Ratio_1$, followed by $Ratio_5$, $Ratio_4$, $Ratio_7$, $Ratio_9$, $Ratio_6$, AD, $Ratio_2$, $Ratio_8$, and $Ratio_3$.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The authors gratefully acknowledge the financial support of the National Natural Science Foundation of China (50934006 and 41272304), National Basic Research (973) Program of China (2010CB732004), China Scholarship Council (CSC), Scholarship Award for Excellent Doctoral Student from Ministry of Education of China (105501010), and Support Program for Cultivating Excellent Ph.D. Thesis of Central South University.

References

[1] L. Dong and X. Li, "A microseismic/acoustic emission source location method using arrival times of PS waves for unknown velocity system," *International Journal of Distributed Sensor Networks*, vol. 2013, Article ID 307489, 8 pages, 2013.

[2] L. Dong and X. Li, "Three-dimensional analytical solution of acoustic emission or microseismic source location under cube monitoring network," *Transactions of Nonferrous Metals Society of China*, vol. 22, no. 12, pp. 3087–3094, 2012.

[3] X. B. Li and L. J. Dong, "Comparison of two methods in acoustic emission source location using four sensors without measuring sonic speed," *Sensor Letters*, vol. 9, no. 5, pp. 2025–2029, 2011.

[4] L. Dong and X. Li, "Hypocenter relocation for Wenchuan Ms 8. 0 and Lushan Ms 7. 0 earthquakes using TT and TD methods," *Disaster Advances*, vol. 6, no. 13, pp. 304–313, 2013.

[5] Y. Gitterman, V. Pinsky, and A. Shapira, "Spectral discrimination analysis of Eurasian nuclear tests and earthquakes recorded by the Israel Seismic Network and the NORESS array," *Physics of the Earth and Planetary Interiors*, vol. 113, no. 1–4, pp. 111–129, 1999.

[6] S. J. Arrowsmith, M. D. Arrowsmith, M. A. H. Hedlin, and B. Stump, "Discrimination of delay-fired mine blasts in Wyoming using an automatic time-frequency discriminant," *Bulletin of the Seismological Society of America*, vol. 96, no. 6, pp. 2368–2382, 2006.

[7] A. J. Mendecki, *Seismic Monitoring in Mines*, Chapman & Hall, 1996.

[8] J. Wuster, "Discrimination of chemical explosions and earthquakes in central Europe—a case study," *Bulletin of the Seismological Society of America*, vol. 83, no. 4, pp. 1184–1212, 1993.

[9] R. Blandford, "Discrimination between earthquakes and underground explosions," *Annual Review of Earth and Planetary Sciences*, vol. 5, p. 111, 1977.

[10] A. T. Smith, "Discrimination of explosions from simultaneous mining blasts," *Bulletin of the Seismological Society of America*, vol. 83, no. 1, pp. 160–179, 1993.

[11] D. R. Baumgardt and G. B. Young, "Regional seismic waveform discriminants and case-based event identification using regional arrays," *Bulletin of the Seismological Society of America*, vol. 80, no. 6, pp. 1874–1892, 1990.

[12] S. R. Taylor, M. D. Denny, E. S. Vergino, and R. E. Glaser, "Regional discrimination between NTS explosions and western US earthquakes," *Bulletin of the Seismological Society of America*, vol. 79, no. 4, pp. 1142–1176, 1989.

[13] S. Taylor, M. Denny, and E. Vergino, "Regional m/sub b: M/sub s/discrimination of NTS explosions and western United States earthquakes," Progress Report, Lawrence Livermore National Laboratory, Livermore, Calif, USA, 1986.

[14] S. G. Kim, Y. Park, and W. Kim, "Discrimination of small earthquakes and artificial explosions in the Korean Peninsula using Pg/Lg ratios," *Geophysical Journal International*, vol. 134, no. 1, pp. 267–276, 1998.

[15] D. R. Baumgardt and K. A. Ziegler, "Spectral evidence for source multiplicity in explosions: application to regional discrimination of earthquakes and explosions," *Bulletin of the Seismological Society of America*, vol. 78, no. 5, pp. 1773–1795, 1988.

[16] M. A. H. Hedlin, J. B. Minster, and J. A. Orcutt, "An automatic means to discriminate between earthquakes and quarry blasts," *Bulletin of the Seismological Society of America*, vol. 80, no. 6, pp. 2143–2160, 1990.

[17] Y. Gitterman and T. van Eck, "Spectra of quarry blasts and microearthquakes recorded at local distances in Israel," *Bulletin of the Seismological Society of America*, vol. 83, no. 6, pp. 1799–1812, 1993.

[18] A. Booker and W. Mitronovas, "An application of statistical discrimination to classify seismic events," *Bulletin of the Seismological Society of America*, vol. 54, no. 3, pp. 961–971, 1964.

[19] L. Dong, X. Li, C. Ma, and W. Zhu, "Comparisons of Logistic regression and Fisher discriminant classifier to seismic event identification," *Disaster Advances*, vol. 6, supplement 4, pp. 1–8, 2013.

[20] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[21] V. F. Rodriguez-Galiano, B. Ghimire, J. Rogan, M. Chica-Olmo, and J. P. Rigol-Sanchez, "An assessment of the effectiveness of a random forest classifier for land-cover classification," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 67, no. 1, pp. 93–104, 2012.

[22] P. M. Granitto, F. Gasperi, F. Biasioli, E. Trainotti, and C. Furlanello, "Modern data mining tools in descriptive sensory analysis: a case study with a Random forest approach," *Food Quality and Preference*, vol. 18, no. 4, pp. 681–689, 2007.

[23] R. Genuer, J. Poggi, and C. Tuleau-Malot, "Variable selection using random forests," *Pattern Recognition Letters*, vol. 31, no. 14, pp. 2225–2236, 2010.

[24] L. J. Dong, X. B. Li, and K. Peng, "Prediction of rockburst classification using Random Forest," *Transactions of Nonferrous Metals Society of China*, vol. 23, no. 2, pp. 472–477, 2013.

[25] L. Dong and X. Li, "Comprehensive models for evaluating rockmass stability based on statistical comparisons of multiple classifiers," *Mathematical Problems in Engineering*, vol. 2013, Article ID 395096, 10 pages, 2013.

[26] L. J. Dong, X. B. Li, M. Xu, and Q. Li, "Comparisons of random forest and Support Vector Machine for predicting blasting vibration characteristic parameters," *Procedia Engineering*, vol. 26, pp. 1772–1781, 2011.

[27] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[28] M. E. Maron, "Automatic indexing: an experimental inquiry," *Journal of the ACM*, vol. 8, no. 3, pp. 404–417, 1961.

[29] I. Kononenko, "Inductive and Bayesian learning in medical diagnosis," *Applied Artificial Intelligence*, vol. 7, no. 4, pp. 317–337, 1993.

[30] P. Berchialla, F. Foltran, and D. Gregori, "Naïve Bayes classifiers with feature selection to predict hospitalization and complications due to objects swallowing and ingestion among European children," *Safety Science*, vol. 51, no. 1, pp. 1–5, 2013.

[31] R. Powers, M. Goldszmidt, and I. Cohen, "Short term performance forecasting in enterprise systems," in *Proceedings of the 11th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '05)*, pp. 801–807, ACM, August 2005.

[32] Y. Fu and L. Dong, "Bayes discriminant analysis model and its application to the prediction and classification of rockburst," *Journal of China University of Mining and Technology*, vol. 38, no. 4, pp. 56–64, 2009.

[33] P. Domingos and M. Pazzani, "On the optimality of the simple Bayesian classifier under zero-one loss," *Machine Learning*, vol. 29, no. 2-3, pp. 103–130, 1997.

[34] A. P. Bradley, "The use of the area under the ROC curve in the evaluation of machine learning algorithms," *Pattern Recognition*, vol. 30, no. 7, pp. 1145–1159, 1997.