

Application and Prospect of Machine Learning in Predicting the Performance of Grouted Sleeve Connections

Jiaqing Zong and Liwei Wu

School of Civil Engineering and Architecture, North China University of Science and Technology, Tangshan 063210, Hebei, China

Abstract: As a key connecting component in prefabricated buildings, the mechanical properties of steel pipe grouted sleeves, such as tensile and compressive properties, are directly related to the safety and reliability of the overall structure. However, traditional research methods rely on physical tests and theoretical models, which have limitations such as long test cycles, high costs, and insufficient prediction accuracy for nonlinear mechanical behaviors. It is difficult to efficiently meet the performance evaluation requirements under complex working conditions. In recent years, machine learning technology has provided an innovative solution for predicting the performance of grouted sleeve connections. By mining the nonlinear relationships between parameters through a data - driven model, it significantly improves the prediction efficiency and accuracy and effectively overcomes the shortcomings of traditional methods. Research shows that there are significant differences in the prediction accuracy of different machine learning models (such as support vector machines, logistic regression, decision trees, etc.). Among them, ensemble learning algorithms (such as random forests, gradient boosting, etc.), with their high robustness and generalization ability, exhibit better prediction performance and have become a hot research direction. However, the application of machine learning in the field of grouted sleeves is still in the development stage. In the future, it is necessary to further optimize the model architecture, expand high - quality datasets, and explore more interpretable hybrid prediction models by combining mechanical mechanisms to promote the in - depth transformation of this technology into engineering practice.

Keywords: Steel pipe grouted sleeve, connection performance, machine learning, prediction.

1. Introduction

With the rapid development of prefabricated building technology, the grouted steel tube sleeve, as the core technology for connecting prefabricated components, has become an indispensable part of modern industrial building systems due to its high load - bearing capacity, construction convenience, and good seismic performance. It forms a reliable connection by filling the gap between the sleeve and the steel bar with grouting material, and its mechanical behavior under tensile, compressive, and cyclic loads directly affects the integrity and safety of the structure. However, traditional research methods mainly rely on physical tests and empirical theoretical models, which have problems such as long test cycles, high costs, and insufficient prediction accuracy for complex nonlinear mechanical responses. It is difficult to meet the efficient performance evaluation requirements under multi - parameter coupling working conditions. In recent years, with the breakthrough of artificial intelligence technology, machine learning has provided new ideas for predicting and optimizing the performance of grouted sleeve connections. By mining the complex nonlinear relationships between factors such as material properties, geometric parameters, and load conditions through data - driven models, machine learning technology has significantly improved the prediction efficiency and accuracy, becoming an important way to break through the bottleneck of traditional research. However, existing research still lacks a systematic discussion on the applicability of different algorithms and the interpretability of models, especially the potential of ensemble learning algorithms remains to be further explored.

Domestic scholars have carried out a large number of

experiments and theoretical analyses on the mechanical properties of grouted steel tube sleeves. Early research focused on the influence laws of parameters such as sleeve material properties, grouting material strength, and steel bar anchorage length on the connection performance. Zhang Yanxiang et al. the basic requirements for the composition of sleeve connections were elaborated. Meanwhile, the tensile performance of grouted steel tube sleeve connections was specifically analyzed [1]. Wu Liwei, Su Youpo, Chen Haibin, et al. twenty-seven axial compression specimens were designed to investigate the axial compression performance of grouted steel tube sleeve connections, and static load tests were carried out. Taking the grouting material strength and connection length as two variable parameters, a calculation formula for the bearing capacity of these connections was proposed. The test results from relevant domestic and foreign literatures were compared with the formula's calculation results, and the calculated values were conservative [2]. Wu Liwei, Su Youpo, et al. in the experiment, 16 groups consisting of 48 grouted steel tube sleeve connection specimens were designed. The aim was to research the axial tensile failure process and mechanism of these connections. Instead of round steel tubes, steel plates were utilized for the specimens, and static load tests were carried out [3]. Zhang Rui, Chen Jianwei, et al. to improve the connection performance of concrete - filled steel tube composite column joints, sleeves were installed for grouting connection between the upper and lower concrete - filled steel tubes. Subsequently, eccentric compression tests were performed on 6 concrete - filled steel tube composite columns equipped with grouted sleeve connections [4].

In recent years, with the promotion of prefabricated buildings, the research has further expanded to the durability

assessment under complex loads (such as fatigue, impact) and seismic performance optimization. However, domestic research still mainly focuses on experiments, and numerical simulations mostly rely on finite element methods, with limited prediction accuracy for multi - variable coupling effects. Research on grouted sleeves abroad started earlier, especially in regions with mature prefabricated technologies such as Europe, America, and Japan. The Precast/Prestressed Concrete Institute (PCI) in the United States developed design guidelines for sleeve connections at an early stage, emphasizing the control standards for the fluidity and compactness of grouting materials. European scholars analyzed the damage evolution process of the sleeve - steel bar interface through refined finite element models and proposed a life - prediction method combined with fracture mechanics theory. Japan systematically studied the stiffness degradation law of sleeve connections under repeated loads in response to the needs of high - intensity earthquake - prone areas. In recent years, foreign research has gradually introduced machine learning technology. For example, a team from the University of California in the United States used neural networks to predict the impact of grouting defects on bearing capacity, but its data scale and model generalization ability still need to be improved.

Overall, domestic and foreign research has accumulated rich results in experimental analysis and theoretical modeling, but the limitations of traditional methods in terms of efficiency and accuracy have become increasingly prominent. The introduction of machine learning technology has opened up a new direction for predicting the performance of grouted sleeves, especially ensemble learning algorithms, which show greater potential by integrating the advantages of multiple models. In the future, it is necessary to further combine mechanical mechanisms with data - driven methods to construct hybrid models with strong interpretability and wide adaptability, so as to promote the leap - forward development of this technology from theoretical research to engineering practice.

2. Data Sources and Data Processing

2.1. Data Sources

In the research on predicting the performance of grouted steel tube sleeve connections driven by machine learning, the acquisition of high - quality data is the basis for model training and verification. Currently, data sources include experimental databases and numerical simulation data, which complement each other in terms of data scale, cost, and applicable scenarios.

Experimental data is obtained through standardized mechanical tests and has high credibility. However, due to limitations in test costs and cycles, the sample size is usually small. The concrete joint performance database released by the National Institute of Standards and Technology (NIST) in the United States collects tensile and shear test data of sleeve connections from multiple research institutions around the world, covering load - displacement curves, failure modes, and material parameters (such as compressive strength of grouting materials, yield strength of steel bars). For example, Henin et al. shared 120 sets of sleeve tensile test data, which are widely used for benchmark testing of machine learning models. The test results indicate that the designed steel bar splicing sleeve has adequate capacity to effectively connect the steel bars. Moreover, compared with existing proprietary

splicing sleeves, it is simpler in structure and more cost - effective [5].

Numerical simulations generate high - fidelity virtual data through finite element analysis (FEA), which can cover a wide range of parameter spaces at low cost. Typical tools include Abaqus, ANSYS, etc. For example, Junbin Lou et al. in this research, a parameter - sharing residual multi - fidelity neural network (PsRMFNN) was proposed to predict the ultimate bond strength of GSC. For this purpose, a database was constructed, which encompassed 209 existing tensile experimental data of GSC. This database was then used to train and assess the PsRMFNN [6]. Wang Ning, Che Wenpeng, et al. the static tensile simulation analysis of joints was carried out using the ABAQUS finite element software. Through comparison, a sleeve joint modeling method featuring good calculation accuracy and an appropriate number of elements was derived [7].

To make up for the deficiencies of a single data source, researchers often adopt the following strategies: fuse experimental data and simulation data, and use transfer learning to improve the model's generalization ability. For example, Gao Ma et al. a database consisting of 418 existing test data for grouted sleeve connections was utilized, along with 11 machine learning algorithms (among which was the random forest algorithm), to construct models for predicting the tensile bearing capacity and failure mode of grouted sleeve connections. The results demonstrate that machine learning methods are capable of effectively forecasting the tensile bearing capacity and failure mode of grouted sleeve connections with diverse common construction defects [8].

2.2. Data Processing

The rational selection of input features is the key to constructing a high - precision machine learning model. For the problem of predicting the performance of grouted steel tube sleeve connections, feature selection needs to be closely combined with its mechanical mechanism and engineering practice. For example, Chen Haibin, Wu Liwei, and Su Youpo et al. through the axial tensile performance test of 27 full - scale grouted steel tube sleeve connection specimens of concrete - filled steel tube columns, the influencing factors of grouting material strength and connection length on the ultimate bearing capacity and bond strength of grouted steel tube sleeve connection joints were analyzed [9]. It mainly covers the following three categories of parameters: material properties, including the compressive strength of grouting materials, the yield strength of steel bars, and the elastic modulus of sleeve steel; geometric parameters, including the ratio of sleeve wall thickness to inner diameter, anchorage length, and steel bar diameter; environmental variables: temperature (related to the curing rate and final strength of grouting materials) and humidity (the risk of corrosion affects the long - term performance of the interface). These features together construct a multi - dimensional input space, laying the foundation for machine learning to capture complex parameter coupling effects.

To improve the robustness and generalization ability of machine learning models, it is necessary to systematically preprocess the original data. Normalization: Min - Max normalization is used to eliminate the difference in dimensions. For example, the sleeve wall thickness (range 5 - 15mm) and grouting material strength (20 - 80MPa) are mapped to the [0,1] interval to avoid the interference of numerical scale differences on model training; Outlier

removal: Based on the 3σ criterion for normal distribution parameters (such as the yield strength of steel bars), or the Isolation Forest algorithm is used to identify abnormal samples in high - dimensional non - linear data (such as abnormal load values caused by sensor failures in experiments); Imbalanced data enhancement: For the imbalance of failure mode samples (such as "steel bar pull - out" samples accounting for 70% and "sleeve buckling" only accounting for 10%), the SMOTE (Synthetic Minority Over - sampling Technique) oversampling technology is used to generate synthetic samples of minority classes through interpolation. At the same time, cross - validation is needed to avoid overfitting [10].

3. Application of Single Machine Learning Algorithms in Predicting the Performance of Grouted Sleeve Connections

3.1. Logistic Regression (LR)

Logistic regression is a probability - based classification method. Its core goal is to predict the probability that a sample belongs to a specific category through the combination of a linear model and a non - linear transformation. In binary classification problems, logistic regression assumes a linear relationship between features and class probabilities, but its output is mapped to the $[0,1]$ interval through the Sigmoid function, thus transforming the linear combination into a probability value. Specifically, the input features of the model generate an intermediate value through the linear combination of weights and biases (i.e., $w^T x + b$), and then the probability that the sample belongs to the positive class is calculated through the Sigmoid function. When the probability exceeds 0.5, the model determines that the sample is a positive class; otherwise, it is a negative class. Its decision boundary is essentially a hyperplane in the feature space. For the training of the model, the cross - entropy loss function serves as the basis. The gradient descent method is adopted to carry out the optimization of parameters. This is done to reduce the gap between the predicted probability and the real label to the minimum. The advantage of logistic regression is that its output has a probabilistic meaning, and the weight coefficients can directly reflect the influence degree of each feature on the classification result, which is suitable for scenarios with high requirements for interpretability.

Regarding the prediction of the performance of grouted sleeve connections, a prediction model founded on logistic regression can be established by means of historical experimental data or engineering monitoring data. The input features include grouting material strength, sleeve geometric dimensions, environmental temperature and humidity, and construction parameters (such as grouting pressure and curing time), and the output is a binary classification label (qualified/unqualified). When training the model, the data needs to be standardized, and regularization techniques (such as L2 regularization) are used to avoid overfitting. In engineering applications, the prediction results of the model can quickly screen potential unqualified connections and guide key re - inspections on - site, thereby reducing the risk of missed inspections. For example, if the weight coefficient of grouting compactness is significantly higher than other features, it indicates that this parameter is a key factor affecting the connection performance and needs to be strictly

controlled during construction. However, the linear assumption of logistic regression may not be able to capture complex non - linear relationships. If there are high - order interaction effects in the actual data, it is necessary to combine feature engineering or switch to non - linear models (such as support vector machines or neural networks) to improve the prediction accuracy. Nevertheless, its efficiency and interpretability still make it a practical tool for engineering quality control, especially suitable for scenarios with limited data and the need for quick decision - making.

3.2. Support Vector Machine (SVM)

The Support Vector Machine (SVM), a supervised learning approach grounded in statistical learning theory, functions with a central concept. This concept involves constructing an optimal classification hyperplane to accomplish efficient data separation. SVM demonstrates particular proficiency in dealing with high - dimensional data and non - linear issues. In classification tasks, the goal of SVM is to find a decision boundary that can separate samples of different classes and maximize the margin between the two classes of samples to this boundary, thereby enhancing the model's generalization ability. For linearly separable data, SVM determines the hyperplane equation ($w^T x + b = 0$) by solving a convex quadratic programming problem. The optimization process of the weight vector w and the bias term b focuses on the support vectors (the sample points closest to the hyperplane), and only these vectors play a decisive role in the decision boundary. When the data is linearly inseparable, SVM maps the original features to a high - dimensional space through the kernel trick. With the help of nonlinear kernel functions (such as the Gaussian radial basis kernel RBF and the polynomial kernel), it implicitly realizes linear separability and avoids the complexity brought by explicit high - dimensional calculations. In addition, the concept of a soft margin and the regularization parameter C are introduced, allowing some samples to violate the margin constraint to balance the classification accuracy and model complexity and enhance the robustness to noisy data. The global optimization characteristics of SVM make it perform well in small - sample and high - dimensional scenarios, but it is sensitive to the selection of kernel functions and parameter tuning.

In the prediction of the performance of steel pipe grouted sleeve connections, SVM can effectively capture the nonlinear relationships under the coupling of multiple factors. The mechanical properties of grouted sleeve connections (such as tensile strength and shear bearing capacity) are affected by multi - dimensional features such as grouting density, sleeve geometric parameters (diameter, wall thickness), grouting material strength, environmental temperature and humidity, and construction technology (grouting pressure, curing time). There may be complex interaction effects among these features. For example, when using the RBF kernel, the model can identify the implicit laws between grouting defects (such as local insufficient density) and performance degradation through local similarity measurement (based on the distance in the feature space). In actual modeling, it is necessary to screen features combined with domain knowledge and determine the optimal kernel function and parameters (such as C and the kernel coefficient γ) through grid search and cross - validation to avoid overfitting. In engineering applications, the prediction results of SVM can be used for grading evaluation (such as "qualified/critical/unqualified") to guide the priority of on -

site spot checks; the physical meaning of its support vectors can also assist in identifying key quality control indicators. If most of the support vectors are concentrated in a specific combination interval of grouting pressure and density, it indicates that this parameter combination is a performance - sensitive area and needs to be strictly controlled during construction. However, the computational complexity of SVM increases significantly with the growth of the sample size. If the scale of on - site monitoring data is large, it is necessary to combine feature dimension reduction or sampling strategies to improve efficiency. Nevertheless, its advantages in complex nonlinear pattern recognition make it an effective tool to supplement traditional detection methods, especially suitable for scenarios of fusion analysis of multi - source heterogeneous data (such as sensor data and material parameters).

3.3. Decision Tree (DT)

The decision tree is a supervised learning algorithm based on a tree - shaped structure. It realizes classification or regression tasks by recursively dividing the feature space. Its core idea is to simulate the human decision - making process. In classification tasks, the decision tree starts from the root node, selects the optimal feature for data division according to the feature selection criterion (such as information gain, Gini coefficient, or chi - square test), and generates child nodes; each child node represents a data subset under a certain feature value condition. Through iterative splitting until the termination condition is reached (such as the sample purity of the node is high enough or the depth reaches the preset threshold). Finally, the leaf nodes are assigned category labels as the prediction results.

In the prediction of the performance of steel pipe grouted sleeve connections, the decision tree can reveal the quality thresholds under the coupling of multiple factors by constructing a hierarchical rule system. The performance qualification of grouted sleeves is affected by multi - dimensional features such as grouting density, sleeve geometric dimensions, material strength, and construction parameters (such as grouting pressure, curing temperature). There may be nonlinear or conditional dependence relationships among these features. For example, the decision tree may first use "grouting density" as the basis for root - node division. If the density is lower than the critical value (such as 90%), it is directly determined as "unqualified"; if it is higher than this value, it is further divided according to "grouting material compressive strength", forming a tree - shaped judgment chain. This characteristic of explicit rules enables engineers to directly identify key control parameters and their thresholds. For example, it is found that when the combination of "curing time < 24 hours" and "winter construction" occurs, the risk of unqualified products increases significantly, so as to optimize the process accordingly. In addition, through feature importance ranking (such as based on the information gain amount when the node is split), the influence weights of each parameter on the performance can be quantitatively evaluated to assist in construction priority management. However, the decision tree is sensitive to data noise. If there are abnormal measurement values or missing values in the training set, redundant branches may be generated. It is necessary to combine feature engineering and ensemble methods (such as gradient - boosting trees) to improve robustness. Nevertheless, its low computational cost and rule interpretability make it an

effective tool for on - site rapid diagnosis and quality control, especially suitable for engineering scenarios that require transparent decision - making bases.

4. Advantages and Applications of Ensemble Learning Models in Predicting the Performance of Grouted Sleeve Connections

Although single machine learning algorithms have achieved certain results in predicting the performance of grouted sleeve connections, due to the complexity and uncertainty of the influencing factors of grouted sleeve connection performance, the prediction accuracy of single algorithms still needs to be improved. Ensemble learning models can effectively make up for the deficiencies of single models and enhance the overall prediction performance by combining the prediction results of multiple single models. Ensemble learning improves the overall performance of the model by combining the prediction results of multiple base learners, using the "wisdom of the crowd". The core principle can be summarized as "Three humble cobblers with their wits combined equal Zhuge Liang the mastermind". Common ensemble learning methods include Bagging, Boosting, and Stacking, etc. For example, Zhang Y et al. an evolutionary - based selective ensemble learning framework was proposed to address classification problems. In this proposed ensemble learning framework, the extreme learning machine (ELM) was chosen as the basic learner, and an evolutionary algorithm was employed to optimize the weights of the basic learners within the ensemble [11]. Webb I G et al. it has been revealed that ensemble learning strategies, with a particular emphasis on boosting and bagging decision trees, have manifested remarkable capabilities in enhancing the prediction accuracy of basic learning algorithms [12]

The Bagging technique creates numerous sub - datasets through the process of sampling the original data with replacement. Subsequently, it trains a model for each individual sub - dataset. Ultimately, it combines the prediction outcomes of these models (for instance, through voting or averaging operations) to yield the final prediction. Take the random forest as an example. It is an ensemble learning algorithm based on Bagging and consists of multiple decision trees. During the prediction of the performance of grouted sleeve connections, every decision tree within the random forest undergoes training on a distinct sub - dataset. By synthesizing the prediction results of all decision trees, it can reduce the variance of the model and improve the stability and accuracy of the prediction.

The Boosting method iteratively trains a series of weak learners, and each new learner is dedicated to correcting the errors of the previous learner. Adaboost and the Gradient Boosting Tree are typical Boosting algorithms. When predicting the performance of grouted sleeve connections, these algorithms continuously adjust the weights of the model step by step, enabling the model to pay more attention to the data points that are difficult to predict, thereby continuously improving the prediction performance.

The Stacking method first uses multiple base models for prediction, and then takes the prediction results of these base models as new features and inputs them into another model (meta - model) for retraining and prediction. For example, first use an artificial neural network, a support vector machine,

and a decision tree to make a preliminary prediction of the connection strength of the grouted sleeve. Take the predicted values of these three models as new features, and then use a logistic regression model as the meta - model for training. Finally, the meta - model gives the prediction results of the ensemble model.

In the prediction of the performance of steel pipe grouted sleeve connections, ensemble algorithms can effectively deal with problems such as multi - scale, nonlinearity, and noise interference in complex engineering data. The mechanical properties of grouted sleeves are affected by the coupling of multiple factors such as grouting density, material parameters, and construction technology. A single model may only be able to capture local feature correlations, while an ensemble model, through the collaboration of multiple base learners, can more comprehensively map the complex relationships between input features and performance. For example, the random forest reduces the sensitivity of decision trees to specific noise features through random feature subset selection and sample resampling, and improves the robustness to grouting defects (such as local cavities). The Gradient Boosting Decision Tree (GBDT) accurately identifies the marginal contribution of construction parameters (such as the combined effect of grouting pressure and curing time) to performance by iteratively optimizing the residuals. Heterogeneous integration (such as Stacking) can integrate the linear discrimination ability of logistic regression and the nonlinear classification advantage of SVM to adapt to the joint analysis of multi - source heterogeneous data (such as numerical sensor data and categorical process records). In practical applications, the feature importance assessment of ensemble models (such as based on the Gini index or the number of splits) can quantify the influence weights of various parameters (such as grouting material strength, sleeve wall thickness) on the connection performance and guide construction quality control. Their high - precision prediction ability can reduce the dependence on traditional destructive testing and achieve low - cost and rapid screening of high - risk samples. However, the "black - box" characteristic of ensemble models may weaken their interpretability, and it is necessary to combine interpretability tools such as SHAP (Shapley Additive Explanations) to assist in decision - making. In addition, their high computational complexity requires a trade - off between hardware resources and engineering timeliness requirements. Nevertheless, the robust performance of ensemble learning in complex engineering scenarios makes it a key technical path to improve the reliability of predictions.

5. Conclusions and Prospects

In conclusion, machine learning technology provides an innovative and effective solution for predicting the performance of grouted sleeve connections. Single machine learning algorithms, such as logistic regression, support vector machines, and decision trees, each based on unique principles, play important roles in predicting the performance of grouted sleeve connections through steps like data collection, preprocessing, model training, and evaluation. However, due to the diversity and complexity of the factors affecting the performance of grouted sleeve connections, single algorithms have certain limitations.

Ensemble learning models significantly improve the accuracy and stability of predictions by combining the

advantages of multiple single models, showing greater potential in practical engineering applications. Currently, though, the prediction of grouted sleeve connection performance based on machine learning still faces some challenges, such as limitations in data quality and quantity, difficulties in optimizing model parameters, and insufficient adaptability to complex working conditions.

In the future, with the continuous development of sensor technology and the Internet of Things technology, it is expected to obtain more high - quality and multi - dimensional data related to the performance of grouted sleeve connections, providing richer training materials for machine learning models. At the same time, the further development and application of emerging machine learning technologies such as deep learning, as well as the continuous optimization of ensemble learning models, will help improve the accuracy and generalization ability of prediction models. In addition, interdisciplinary research cooperation, which deeply integrates machine learning with disciplines like materials science and structural mechanics, is expected to deeply reveal the internal mechanism of the performance of grouted sleeve connections, providing a more solid theoretical foundation for model construction and promoting the wide application and continuous development of machine - learning - based prediction technology for grouted sleeve connection performance in the field of construction engineering.

References

- [1] ZHANG, Y.X., LIN, F., MA, G.Y., et al.: Discussion on Tensile Behavior of Grouted Sleeve Connection in Steel Pipe, *Jiangxi Building Materials*, Vol. (2020) No.10, p.193-194.
- [2] WU, L.W., SU, Y.P., CHEN, H.B., et al.: Experimental Study on Axial Compression Behavior of Grouted Sleeve Connection in Steel Pipe, *Journal of Building Structures*, Vol. 40 (2019) No.11, p.191-199.
- [3] WU, L.W., SU, Y.P., and ZHAO, J.S.: Simulation Experimental Study on Axial Tensile Behavior of Grouted Sleeve Connection in Steel Pipe, *Journal of Building Structures*, Vol. 40 (2019) No.2, p.169-177.
- [4] ZHANG, R., CHEN, J.W., WU, S., et al.: Experimental Study on Eccentric Compression Behavior of Concrete-Filled Steel Tube Composite Columns with Grouted Sleeve Connections, *Journal of Building Structures*, Vol. 44 (2023) No.S1, p.239-247.
- [5] Henin E ,Morcous G .Non-proprietary bar splice sleeve for precast concrete construction[J].*Engineering Structures*, 2015, 83154-162.
- [6] Lou J ,Li Y ,Feng Q , et al.Machine learning models for predicting ultimate bond strength of grouted sleeve connections[J].*Structures*,2025,72108186-108186.
- [7] WANG, N., CHE, W.P., WU, L.W., et al.: Experimental Study and Finite Element Analysis on Mechanical Properties of New 12mm Rebar Grouted Sleeve Connection, *Industrial Construction*, Vol. 49 (2019) No.4, p.81-87+94.
- [8] Gao M ,Chunxiong Q ,Hyeon-Jong H , et al.Data-driven models for predicting tensile load capacity and failure mode of grouted splice sleeve connection[J].*Engineering Structures*, 2023, 289.
- [9] CHEN, H.B., WU, L.W., and SU, Y.P.: Experimental Study on Tensile Behavior of Grouted Sleeve Connection in Steel Pipe, *World Earthquake Engineering*, Vol. 32 (2016) No.2, p.18-24.
- [10] Sáez A J ,Luengo J ,Stefanowski J , et al.SMOTE-IPF: Addressing the noisy and borderline examples problem in

- imbalanced classification by a re-sampling method with filtering[J].Information Sciences,2015,291:184-203.
- [11] Zhang Y ,Liu B ,Yu J .A selective ensemble learning approach based on evolutionary algorithm[J].journal-titleJournal of Intelligent & Fuzzy Systems,2017,32(3):2365-2373.
- [12] Webb I G ,Zheng Z .Multistrategy Ensemble Learning: Reducing Error by Combining Ensemble Learning Techniques[J].IEEE Trans. Knowl. Data Eng.,2004,16(8):980-991.