IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received June 25, 2019, accepted July 1, 2019, date of publication July 18, 2019, date of current version August 5, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2929533

Evaluation of Goaf Stability Based on Transfer Learning Theory of Artificial Intelligence

YAGUANG QIN, ZHOUQUAN LUO, JIE WANG, SHAOWEI MA, AND CHUNDI FENG
School of Resources and Safety Engineering, Central South University, Changsha 410083, China

Corresponding authors: Shaowei Ma (mashaowei@csu.edu.cn) and Chundi Feng (cdfeng@csu.edu.cn)

This work was supported in part by the National Key Research and Development Program of China during the Thirteenth Five Year Plan Period: The Continuous Mining Theory and Technology on Spatiotemporal Synergism of Multi-Mining Areas Within a Large Ore Block for Deep Metal Deposit under Grant 2017YFC0602901, in part by the National Natural Science Foundation of China under Grant 51274250, and in part by the Fundamental Research Funds for the Central Universities of Central South University under Grant 2017zzts204.

ABSTRACT Current artificial intelligence models for evaluating goaf stability in underground metal mines need a large amount of sample data for training, and their accuracy declines with small sample size. With the aim of solving this problem, this paper proposes an improved TrAdaBoost algorithm based on transfer learning theory. The scope of the TrAdaBoost algorithm is extended from the two-level classification to multi-level classification problems, which makes it suitable for evaluating goaf stability. The isolated forest method is used to filter the bad points of the auxiliary training set, thereby eliminating the interference of abnormal data. The dynamic factor concept is introduced to solve the problem that the weight of source domain data decreases too quickly and irreversibly, and this enhances the generalization performance of the algorithm for different goaf samples. To test the accuracy of the proposed model in predicting goaf stability, an evaluation model is constructed and the performance compared with other algorithms in current use. The prediction accuracy and generalization ability of the model are evaluated by mean square error and F1 measurements, which prove that the performance of the model is excellent. The most obvious finding to emerge from this paper is that, with suitable improvements, goaf evaluation models can still maintain a high level of accuracy with small sample size.

INDEX TERMS Artificial intelligence, data mining, safety, transfer functions.

I. INTRODUCTION

Goaf stability in underground mines is one of the major hazards affecting safety in metal production. The evaluation of goaf stability has important technical and economic significance for ensuring mine safety. A reasonable analysis of goaf stability is helpful for enterprises to understand correctly the safety status of their mines. The main methods in current use for analyzing goaf stability are the prediction model evaluation method, theoretical analysis method, physical model test method, and numerical simulation method [1].

There are many studies on the evaluation of goaf stability. Xie *et al.* [2] used catastrophe theory to evaluate and predict the stability of a goaf group. Wang *et al.* [3] evaluated goaf stability levels by using support vector machine and compared the results with a traditional neural network, which proved that the support vector machine has better generalization performance under small data samples. Song *et al.* [4] monitored the failure of surrounding rock in the process of stope mining by means of a physical model test method, and explained the deformation and failure mechanism of surrounding rock. Hu *et al.* [5] built the RS–TOPSIS model based on the results of expert investigation to predict precisely the degree of goaf hazard. Through theoretical analysis, Wang *et al.* [6] found that the stability of overlying strata and grouting entities are the key factors affecting goaf stability. Zhang *et al.* [7] proposed the idea of analyzing and evaluating goaf stability by multiscale decomposition, which can help us evaluate the deformation characteristics reflected in raw data more accurately and effectively. On the basis of variable rock mechanics parameters from experimental data, Xiao *et al.* [8] found that the mechanical properties of rock are greatly influenced by rheological and dynamic disturbance, which can greatly reduce goaf stability. All of the above studies have made positive progress.

Among the various methods for evaluating goaf stability, the prediction model evaluation method takes a comprehensive account of the factors affecting goaf stability.

The associate editor coordinating the review of this manuscript and approving it for publication was Jiankang Zhang.

The evaluation speed is fast and the method is easy to understand. However, current prediction models need a large amount of sample data to train the model. With a small sample size, their evaluation accuracy will be greatly reduced. Transfer learning uses data from a similar task to the target task in order to solve the problem of small sample size [9]. However, current methods of transfer learning cannot be used directly in the evaluation of goaf stability. Therefore, transfer learning needs to be able to find the components of the target task from the source task, and combine them to form target predictive classifiers. It could then be used to evaluate goaf stability. Improved transfer learning could improve the accuracy of the model and overcomes the shortcomings of a small sample size. It could also overcome the impact of ignoring differences in attributes in the same area due to similar attributes in the samples.

II. TRANSFER LEARNING THEORY

A. SELECTION OF TRANSFER LEARNING ALGORITHMS

From the perspective of the Transfer scenario, transfer learning can be divided into three categories based on the differences of the domain and task between source and target [10]: Inductive TL, Transductive TL, and Unsupervised TL.

Inductive TL is used mainly for the same source domain and target domain. However, for a different source task and target task, the interface is determined by the category information of the target domain data, so the target data must have a label [11]. In Transductive TL, the source domain data and the target domain data may be different but should have a certain correlation, and the final tasks between the data are the same [12]. This method relies heavily on the source domain data category information. The source domain data and the target domain data required for Unsupervised TL are different, but they need to be correlated to extract the common features, and to deal with clustering and dimensionality reduction [13].

As regards goaf stability in underground metal mines, the information collected by field measurement and literature review is of different mines. When the source data is used to evaluate the stability of the target mine goaf, the domain feature space of the source data and the target data is the same, and the probability difference is weak. However, due to the limited amount of data, some information in the target domain is not easily captured by the model. Source data is the main source of information for the target data. Therefore, it is appropriate to choose Inductive TL.

Inductive TL algorithms include TrAdaBoost (an extension of AdaBoost), MT–IVM, Statistical Relational Learning (SRL), Supervised Feature Construction (SFC), and Unsupervised Feature Construction (UFC) [14]. Among them, MT–IVM needs to split the parameters; SRL emphasizes the logical relationship between parameters; and SFC and UFC are used mainly to solve the problem of feature selection, which is not a concern in this paper. The TrAdaBoost method assumes that the source and target domain data have the same characteristics and labels, but the distribution is quite different. Some source domain data will help the target domain learning, but some may be detrimental to the target domain learning [15]. Therefore, it is necessary to extract instances from the source data to improve the model. With TrAdaBoost, a portion of the beneficial source data is tagged and combined with new target data. The weight of the target data with a poor classification effect in the model training is continuously increased, while the weight of the source data with a weak classification effect is reduced. There are similarities and inconsistencies between the evaluation indexes of different goafs. However, goaf data in the source domain has great reference value, which can help to extract goaf features in the target domain. Therefore, this paper selects the TrAdaBoost method to evaluate goaf stability. A schematic diagram of the TrAdaBoost method is shown in Fig. 1.

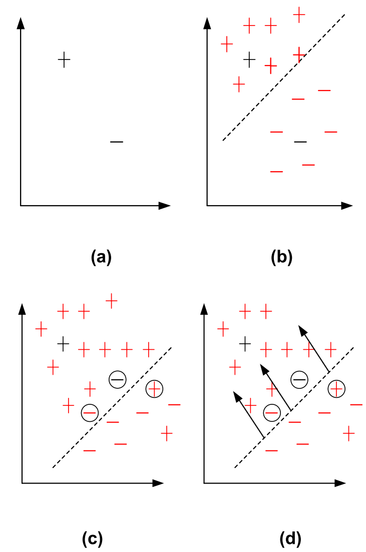

FIGURE 1. TrAdaBoost schematic: (a) data to be classified; (b) adding source data; (c) interference of source data; (d) adjusting the interference of source data.

In Fig. 1(a), when the number of labeled training samples is small, classification learning is difficult, and the interface is determined to be random and blind. If a large number of auxiliary training samples is obtained at this time, such as the red '+' and '−' samples in Fig. 1(b), the classification surface of the training samples can be estimated. However, according to the principle of equivalence exchange in machine learning, auxiliary data and target data are not exactly the same. Therefore, there may be abnormal auxiliary training samples that mislead the classification results. For example, the black '−' sample in Fig. 1(c) is misclassified.

In transfer learning, the abnormal auxiliary training samples do not help the training sample classification, but decrease the accuracy of the interface. By increasing the weight of misclassified target data and reducing the weight of the source data, the TrAdaBoost method can direct classification training along the correct path, as shown in Fig. 1(d). By continuously adjusting the weight of the source domain and target domain, and maximizing the common features of the auxiliary training samples and training samples, a more realistic model can be trained.

B. IMPROVEMENT OF TRADABOOST ALGORITHM

1) INTRODUCING THE TRADABOOST ALGORITHM

The basis of TrAdaBoost is the AdaBoost algorithm, which is a typically integrated algorithm [16]. For a classification problem, it is much easier to find weak classifiers than strong classifiers under a given training set. The AdaBoost algorithm first obtains a series of weak classifiers, and then combines them to form a strong classifier.

The idea of assigning weights to source and target data in the TrAdaBoost algorithm is consistent with the idea of assigning values to weak classifiers in the AdaBoost algorithm. The principle of the AdaBoost algorithm can be used to solve the problem of how to reasonably increase the weight of target data and reduce the weight of source data in TrAdaBoost. The source data and target data are regarded as weak classifiers in the AdaBoost algorithm. At first, each sample in the training data is granted a weight. When a sample in the target domain is misclassified, it is considered that the target sample is difficult to classify. Therefore, the weight of the training sample is increased, so that the proportion of the sample in the next training sample will be larger. If a sample in the source dataset is misclassified, the sample is considered to be very different from the target data. Hence the proportion of the sample in the classifier is decreased, which is basically the same idea behind the AdaBoost algorithm.

The TrAdaBoost algorithm works well when the source data and the auxiliary data have numerous similarities. However, training classifiers becomes more difficult if the sample noise in auxiliary data is large, or the number of iterations is not well controlled.

2) THE MULTI-CLASSIFICATION PROBLEM

The TrAdaBoost algorithm is a two-factor classification algorithm whose classification result is of type (0,1). Following previous research [17], [18], this paper divides goaf stability into four levels. Therefore, the TrAdaBoost algorithm needs to be adjusted so that it can be applied to the solution of multi-level classification problems. In machine learning, there are three main ways to extend a two-level classification algorithm to a multi-level classification algorithm: one-to-the-other; one-to-one; and multi-to-multi classification.

Although these methods can be used to extend the TrAdaBoost algorithm, they do not consider the specific characteristics of goaf stability. The ordering relationship of the labels in the sample tag set is not considered in the multi-classification extensions, but there is a logical order among the goaf stability indicators. Therefore, if the characteristics of the research object are taken into account, the TrAdaBoost algorithm can be extended to play an effective role in classification.

Since the evaluation of goaf stability is divided into four levels, the evaluation requires four-level classifications. In model training, Level 1 and Level 2 indicators are merged into Class A, and Level 3 and Level 4 indicators are merged into Class B. Thus, the four-level classification problem is reconstructed as a two-level classification problem. The TrAdaBoost algorithm is used to train the newly constructed two-level classification problem. The trained model can be used to classify goaf stability into Class A and Class B initially, and the model is named Model 1.

The above process is repeated, and Class A (Levels 1 and 2) and Class B (Levels 3 and 4) are trained respectively to obtain two new classifiers, Model 2 and Model 3. When using the model to test the goaf data, Model 1 is first called for a preliminary classification of the goaf stability. The next classification model is selected according to the results of Model 1. If the goaf is classified as Class A, then Model 2 is called to further classify and determine its stability level to be Level 1 or Level 2. Otherwise, Model 3 is called for further classification, and its stability category is determined to be Level 3 or Level 4. The goaf evaluation is then completed. The algorithm steps of the multi-classification model are shown in Fig. 2.

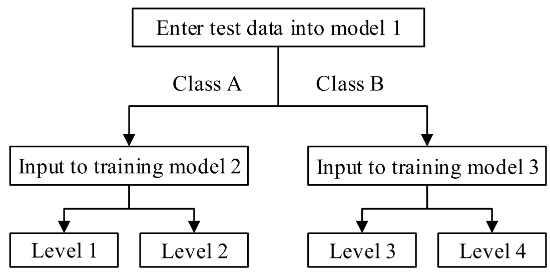

FIGURE 2. Algorithmic steps of the multi-classification model.

3) ALGORITHMIC STRATEGY OF MULTI-CLASSIFICATION MODEL

Given an s-classification problem, if the total number of samples is z, the learning sample is:

$$(x_1, y_1), \ldots (x_l, y_l), \ldots, x_i \in R, \qquad y_j \in (1, 2, \ldots, s), \quad j = 1, 2, \ldots, z \tag{1}$$

TrAdaBoost = (TrAdaBoos t_1, TrAdaBoos t_2, …, TrAdaBoos t_{n-1}) is a collection of classifiers designed according to the training sample classification. The i-th TrAdaBoost solves the following problem:

$$\min e_{tj} = \sum_{i=n+1}^{n_j+m_j} \frac{\omega_{ij}^t \left| h_{tj}(x_i) - c(x_i) \right|}{\sum_{i=n+!}^{n_j+m_j} \omega_i^t}, \quad j = (1, 2, \ldots, s-1) \tag{2}$$

After $s - 1$ training sessions, $s - 1$ training models are obtained. When evaluating new verification data, evaluation values of the data can be obtained by searching the binary tree from high to low. It is necessary to construct $s - 1$ classifiers for multi-classification evaluation using this method.

Increasing the number of classifiers increases the number of model operations and the computational cost. The number of samples included in each classifier also affects the duration of the model operation. Therefore, the amount of repetitive sample training needs to be taken into account when evaluating the performance of the model. For a one-to-the-other strategy, the amount of repetitive training is:

$$M_1 = s \times z \tag{3}$$

For a one-to-one strategy, the amount of repetitive training is:

$$M_2 = (s-1) \times z \tag{4}$$

For a multi-to-multi strategy, the amount of repetitive training is:

$$M_{3\,\max} = \log_2 s \times z \tag{5}$$

Comparing the above three expansion strategies, the time complexity of the multi-to-multi strategy is linear, and the maximum amount of repetitive training is smaller than the other two strategies. Using this classification strategy not only greatly reduces the number of classifiers, but also speeds up the model operation and reduces memory usage. At the same time, as the algorithm is hierarchically structured by constructing binary trees, the data imbalance of a one-to-the-other strategy is avoided, and the risk of overfitting and underfitting in the model is also largely avoided.

4) FILTERING OF SOURCE DATA

The training effect of transfer learning may be positive in improving data classification, or negative in decreasing the classification accuracy, which is called the negative transfer phenomenon. Although goaf characteristics in different underground metal mines are not identical, they have a certain reference value. Using TrAdaBoost methods to process data can help to build models with better generalization and performance. Analysis of the source domain data shows that the normal data distribution should conform to a specified probability distribution in the feature space. Even if the aggregation of data is sparse or dense, there will be no isolated or outlier points (such points are called bad points). Bad points will interfere with the training of the target domain model and reduce its accuracy. Therefore, an indispensable step is to process the source domain data and screen the bad points to prevent the negative migration phenomenon in the model training.

Isolated forest is an unsupervised anomaly detection method suitable for continuous numerical data, which is used to detect data that is inconsistent with other data rules [19]. To detect which points are easy to isolate, isolated forests use a very efficient strategy. When constructing an isolated tree, some data is extracted from the training data under the process of non-repeated sampling. A binary partition method is used to partition the selected samples, and the dataset is divided randomly and recursively until all the sample points are isolated. Under this random segmentation strategy, the bad points usually have shorter paths, and therefore can be separated from the normal points and eliminated effectively.

Given the sample $X = \{x_1, \ldots, x_n\}$, the dimension of the feature is d. In order to construct an isolated tree, a feature q and its segmentation value p need to be randomly selected. We split dataset X recursively until one of the following conditions is satisfied: (1) the tree reaches the limit height; (2) there is only one sample on the node; or (3) all the features of the sample on the node are the same. The average path length of the tree is:

$$c(n) = 2\mathrm{H}(n-1) - \frac{2(n-1)}{n} \tag{6}$$

In Eq. (6), $\mathrm{H}(n)$ is a harmonic number, which can be estimated as $\ln(n)+0.5772156649$.

As shown in Fig. 3, the average path length of abnormal points is less than that of normal points. The abnormal score of sample x is defined as:

$$s(x, n) = 2^{-\frac{E(h(x))}{c(n)}} \tag{7}$$

In Eq. (7), $E(h(x))$ is the expected value of the path length of sample x in a batch of isolated trees, $s(x, n)$ is the abnormal index of x in the i-th tree of n samples, and $s(x, n)$ has a range of [0, 1]. The judgment of an abnormal situation is divided into three cases: (1) the closer to 1, the higher the probability of there being an abnormal point; (2) the closer to 0, the higher the probability of there being a normal point; and (3) if the $s(x, n)$ values of most training samples are close to 0.5, there is no obvious outlier for the entire dataset.

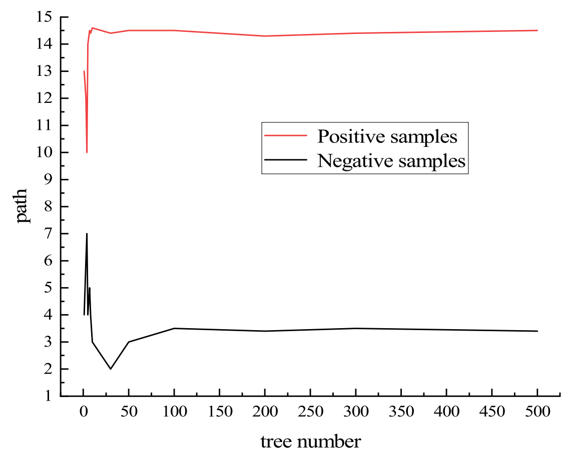

FIGURE 3. **Comparison of point paths in isolated forests.**

Anomaly detection based on isolated forests involves two steps. (1) In the training phase, an isolated tree is established based on the subsamples of the training set. (2) In the test phase, the isolated tree is used to calculate the abnormal score for each test sample. The characteristic of abnormal data is that the amount of data is small, which is different from normal data. Therefore, the isolated trees only need to consider the data points whose path is below the average

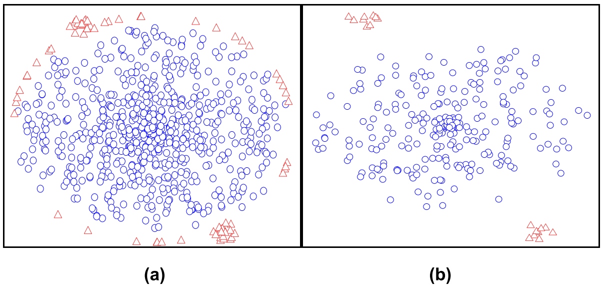

FIGURE 4. Comparison of the isolated forest before and after sampling: (a) abnormal point identification before sampling; (b) abnormal point identification after sampling.

length, without the need to span the whole tree. Experiments on the number of isolated trees show that when the number reaches 100, the path length can be covered.

In Fig. 4, Fig. 4(a) represents the isolated forest model of the original data, and Fig. 4(b) the isolated forest model of sampled data. Among them, the blue points are normal samples, and the red points are abnormal samples. As shown in Fig. 4(a), the normal and abnormal samples are densely overlapping, and they are difficult to separate effectively. To maintain a high level of recognition accuracy, the number of abnormal points that cannot be identified correctly needs to be reduced. After sampling, the amount of data is reduced for each tree. Each isolated tree can identify a specific subsample, and its performance is therefore enhanced. The abnormal samples and normal samples then become easy to segment. Therefore, using isolated forest to preprocess the source domain data and eliminate the bad points will greatly improve the classification effect.

5) ADJUSTMENT OF SOURCE DOMAIN WEIGHT

The most serious problem with the TrAdaBoost method is that the weight of the source domain decreases too fast. Even if the source samples express the target concept, their weights will drop rapidly. With the strategy of updating weights using the TrAdaBoost method, the weight difference between source sample and target sample is expanded continuously as the number of iterations increases, and the expansion is irreversible. When the model achieves the best effect within the number of iterations, the final training model may not be the best model due to the persistence of weight reduction of the source data. Therefore, it is necessary to improve the weight transfer phenomenon of the TrAdaBoost method. This phenomenon can be categorized as a single-source dynamic transfer problem. Following multi-source dynamic theory [20], the concept of dynamic factors is used to modify the TrAdaBoost method, so that the dynamic factor plays an enhanced role in the single-source dynamic migration problem.

Before updating the sample weight vector, we add the dynamic factor C_t:

$$C_t = 2(1 - e_t) \tag{8}$$

Dynamic factors prevent the weight of source data falling too fast in the iteration and adjust the update of the weight.

$$\omega_i^{t+1} = \begin{cases} C_t \omega_i^t \beta^{|h_t(x_i) - c(x_i)|} & i = 1, \ldots, n \\ \omega_i^t \beta_t^{-|h_t(x_i) - c(x_i)|} & i = n+1, \ldots, n+m \end{cases} \tag{9}$$

The following proves that in single-source dynamic migration, the introduction of dynamic factors can slow down the decline of source data weight. In the iteration of step $t+1$, let P and Q be the sums of the weights of the correct classification and incorrect classification targets, respectively. The weight vector of the source data is updated to:

$$\omega_{ai}^{t+1} = \omega_i^t \beta^{|h_t(x_i) - c(x_i)|} = \frac{\omega_{ai}^t}{m\omega_{ai}^t + P + Q} \tag{10}$$

To prevent the source data weight from decreasing, a dynamic factor is introduced in each iteration step:

$$\omega_{ai}^{t+1} = C_t \omega_i^t \beta^{|h_t(x_i) - c(x_i)|} \tag{11}$$

After adding dynamic factors, the iterator passes through t iterations, and the weight of the source domain data tends to become stable. Therefore, the equality of $\omega_a^{t+1} = \omega_a^t$ holds, and the value of the dynamic factor can be obtained as follows:

$$\omega_{ai}^t = \frac{C_t \omega_{ai}^t}{C_t \omega_{ai}^t + P + Q} = \frac{C_t \omega_{ai}^t}{C_t \omega_{ai}^t + 2n\omega_b^t (1 - e_b^t)} \tag{12}$$

$$C_t = 2(1 - e_t) \tag{13}$$

Using the TrAdaBoost method, the formula for updating the weight vector after adding dynamic factors can now be expressed as follows:

$$\omega_i^{t+1} = \begin{cases} C_t \omega_i^t \beta^{|h_t(x_i) - c(x_i)|}, & i = 1, \ldots, n \\ \omega_i^t \beta_t^{-|h_t(x_i) - c(x_i)|}, & i = n+1, \ldots, n+m \end{cases} \tag{14}$$

Thus, when the model tends to become stable after a sufficient number of iterations, the weight of the source domain data will no longer decrease.

III. CONSTRUCTION AND PERFORMANCE EVALUATION OF GOAF STABILITY MODEL

A. CONSTRUCTION OF GOAF STABILITY MODEL

In engineering practice, the factors impacting goaf stability fall into four categories: geological, hydrological, environmental, and geometric factors. The evaluation indexes can be selected synthetically bearing in mind three objectives: to reduce overfitting, improve accuracy, and reduce training time. The indexes can be divided into qualitative indicators and quantitative indicators. The qualitative indicators are: rock mass structure (X_1); geological structure (X_2); groundwater (X_4); goaf layout (X_5); mining disturbance (X_{10}); and condition of adjacent goaf (X_{11}). Qualitative indicators take values between 0 and 1, according to actual conditions. Quantitative goaf indexes include: rock compressive strength (X_3); goaf volume (X_6); exposed area of goaf roof (X_7); buried depth (X_8); and span-depth ratio (X_9). Drawing on previous research into goaf stability evaluation and practical engineering experience [17], [18], the goaf stability of underground mines is divided into four levels, as shown in Table 1.

TABLE 1. Grading standard for goaf stability.

Stability level	Stability status or risk of instability
1	The goaf has very good stability and no risk of instability
2	The goaf has good stability and low risk of instability
3	The goaf is less stable and the risk of instability is greater
4	Stability of the goaf is extremely poor and the risk of instability is very high

Drawing on engineering practice, 87 groups of goaf data from two different mines were collected and used in the evaluation of the proposed model. These are shown in Table 2.

Since the training concentration includes data from different mines, the values of qualitative indicators are given by relevant experts. Different experts often adopt different scoring systems, which lead to differences in values between the qualitative indicators of different mines. However, the degree of influence of the different factors is reflected in the scores. Therefore, it is necessary to normalize these qualitative indicators to eliminate the errors caused by the same index occupying different ranges [21]. The min-max standard was used to normalize the data, and the results were mapped between 0 and 1. The conversion function is as follows:

$$X_{inew} = (X_i - Min) \,/\, (Max - Min) \tag{15}$$

In Eq. (15), X_{inew} is the normalized data, X_i is the original data, and Min and Max are the minimum and maximum values, respectively, in the original data. The results of normalization are shown in Table 3.

The goaf evaluation data were collected from two different underground metal mines. As noted above, 87 datasets were used as source domain data. For the evaluation process, 20 of the 50 datasets from one mine were used as target domain data, and the other 30 datasets were used as verification datasets. The isolated forest algorithm was used to process the source data and eliminate the bad points.

While training the sample, the sample data is predicted. The normal data is marked as +1 and the abnormal outlier data is marked as −1. Source domain data were substituted into the isolated forest model, and the evaluation results showed that groups 6, 15, and 31 were bad points, so these groups of data were eliminated. The model data is distributed in eleven-dimensional data space. Selecting any two dimensions and observing the distribution of training data in two-dimensional space, it can be found that the positions of bad points in different two-dimensional spaces are at the edge of the data distribution, which is shown in Fig. 5.

Since TrAdaBoost is a two-level classification algorithm, it cannot be used directly for multi-level classification problems. According to the multi-classification expansion process as shown in Fig. 2, the training data needs to be subdivided into three datasets I, II, and III, which are used for Models 1, 2, and 3, respectively. Model 1 is used to classify

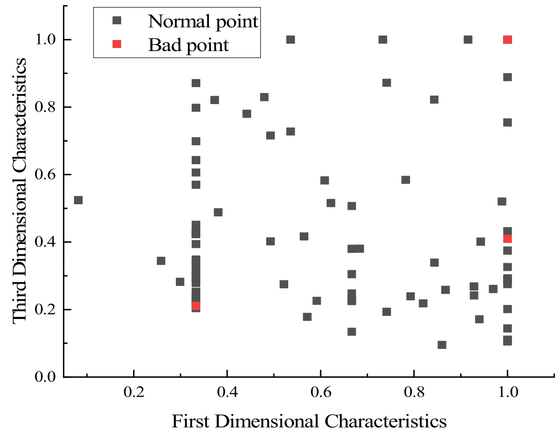

FIGURE 5. Distribution of training data.

the goafs. Therefore, it is necessary to redefine the stability of Levels 1 and 2 goafs as Level A, and Levels 3 and 4 goafs as Level B. The processed dataset is shown in Table 4.

Model 2 is used to classify Level 1 and Level 2 goafs. All goaf samples at these stability levels were selected from the training set, obtaining a total of 55 groups as shown in Table 5. Model 3 is used to classify Level 3 and Level 4 goafs. All goaf samples at these stability levels were selected from the training set, obtaining a total of 29 groups, as shown in Table 6.

B. PERFORMANCE EVALUATION OF GOAF STABILITY MODEL

1) ACCURACY ANALYSIS

As noted above, 50 datasets from one underground metal mine were selected for the evaluation process; 20 were used as target domain data for training the model, and 30 datasets were used as verification datasets to test the model's accuracy and stability. For the source domain data, 87 datasets from Table 2 were used to assist training, which were reduced to 84 after the elimination of the bad points. The total number of training samples was therefore 104 sets of goaf data. The data used for testing is shown in Table 7.

The validation datasets were substituted into the model to obtain the prediction results, as shown in Table 8. Among the 30 groups of data, 27 groups were exactly the same as the actual results, and the accuracy rate reached 90%. As for the remaining three datasets, the prediction level was also very close to the actual level.

In traditional evaluation models, AdaBoost is a widely used classification algorithm, which enhances the classification effect by integrating weak classifiers. The AdaBoost algorithm provides the basis for the TrAdaBoost algorithm. The TrSVM and TrBys algorithms are also widely used multi-classification methods in transfer learning, which can handle directly the multi-classification of training objects. By comparing the accuracy and generalization performance of the four algorithms above, the performance of the TrAd-

TABLE 2. Initial dataset.

Serial number	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	Stability level
1	3	1	39	2	2	46680	4330	370	4	4	4	4
2	2	2	49	2	1	11700	1300	250	2	4	4	2
3	3	3	37	2	2	32240	2480	300	3	4	4	3
4	3	2	50	2	3	33600	1680	270	4	4	4	3
5	3	1	28	3	2	179400	5100	332	3	4	4	4
6	3	3	52	2	3	137760	5740	295	4	4	4	3
7	2	1	53	1	1	46560	2910	305	1	4	4	1
8	2	1	59	2	1	43600	2180	320	2	1	1	2
9	1	1	66	2	1	12000	1200	335	1	4	4	1
10	1	2	50	3	2	232580	4010	380	2	4	4	3
11	1	2	59	1	1	33900	2260	305	2	4	4	1
12	1	1	61	2	1	18850	1450	290	1	1	1	1
13	1	1	52	2	1	18130	2590	201	1	3	3	1
14	1	1	55	1	1	29160	2430	208	1	3	3	1
15	3	3	56	2	2	10575	2350	208	1	3	3	2
16	1	1	54	1	1	18000	1800	208	2	3	3	1
17	1	2	57	2	2	25600	1600	411	3	4	4	2
18	1	1	54	2	1	16000	2000	300	2	1	1	2
19	1	1	55	1	1	44850	1950	201	2	3	3	2
20	1	1	52	1	1	37500	5000	195	1	3	3	1
21	1	1	51	1	1	42510	3270	180	2	3	3	2
22	1	2	54	1	2	9750	1300	180	2	4	3	2
23	1	2	55	1	1	10726	1730	180	2	4	4	2
24	1	1	53	1	1	11220	1870	230	1	3	3	1
25	1	1	53	1	1	16380	1170	230	2	3	3	2
26	1	1	54	1	1	39840	2490	230	2	1	1	1
27	1	1	53	1	1	31720	2440	230	2	1	1	1
28	1	1	43	2	3	80040	3480	230	3	4	4	3
29	1	1	55	1	1	11280	1410	230	2	1	1	1
30	2	1	44	3	2	148750	1750	350	3	3	3	3
31	1	1	52	1	2	12100	1210	350	1	1	1	1
32	1	1	56	1	2	17160	1320	375	2	3	3	2
33	1	1	57	1	1	27675	3690	210	2	3	3	2
34	2	2	41	3	3	25560	1420	210	3	3	3	3
35	1	2	40	3	2	43680	1680	290	3	4	3	3
36	2	2	43	4	2	81270	1290	290	3	4	4	3
37	1	2	42	2	2	128240	4580	400	3	4	4	4
38	83	9	129	37	1	55	33	185	0.67	0.25	0.8	1
39	68	8	129	62	6	108.3	65	185	0.45	0.63	0.76	2
40	70	7	82.8	81	8	408.3	245	182	0.29	0.85	0.65	3
41	82	8	82.8	65	5	258.3	155	174	0.41	0.68	0.79	2
42	72	9	82.8	40	4	128.3	77	176.5	0.55	0.35	0.6	2
43	76	6	59.6	83	7	496.7	298	143	0.25	0.78	0.43	3
44	78	9	93.6	42	3	85	51	129	0.56	0.43	0.75	2
45	1.18	1.3	53.33	2.2	1.12	21333.33	2058.82	295.83	2.08	1.12	1.04	2
46	1.12	1.06	52.67	1.08	1.03	39843.75	4942.53	190.85	1.08	3.08	3.04	1
47	1.09	2.09	54.67	0.92	2.03	9750	1359.09	184.19	2.2	4	2.96	2
48	1.03	0.97	54.99	0.96	1.18	11220	1986.88	225.82	1.08	3.04	2.92	1
49	1.3	1	54	1	0.97	39840	2547.91	225.82	2.2	1.08	0.88	1
50	1.03	1.06	43.66	2.24	2.79	75331.76	3538	234.18	3.04	3.6	3.8	3
51	1.97	1.06	42	3.04	2.06	151111.11	1983.33	358.24	2.88	2.84	3.04	3
52	2.09	1.94	41.66	3.04	2.97	25560	1360.83	205.88	2.88	2.92	3.12	3
53	1.94	1.94	44.32	4.21	2.06	86811.14	1326.86	278.99	3.08	4	4	4
54	67.5	6.39	80.92	78.49	8	398.34	242.01	176.43	0.3	0.85	0.66	4
55	78.51	6	55.71	83	6.84	496.7	268.2	143	0.27	0.78	0.52	3
56	1.12	1.18	41.68	2.28	3	82394.12	3596	221.64	3	3.64	3.88	3
57	76	9	93.6	45	4	118.3	71	124	0.56	0.28	0.72	2
58	2.01	1.61	49.58	2.72	0.15	70477.66	2540.27	202.7	0.6	0.24	0.92	3
59	0.12	0.52	52.31	0.84	0.94	74880	2186.36	191.67	1.16	0.28	1.36	2
60	1.41	2.04	48.23	2.24	0.58	176080	5536.74	300	1.6	3.6	1.32	3
61	1.74	0.82	44	2.84	2.49	81600	1506.21	303.23	2.84	3.48	2.84	3
62	1.68	4.24	73.21	2.03	2.55	26360.53	1191.52	261	1.34	3.36	3.92	3
63	0.06	2.73	52.66	2.56	1.95	115374	6674.42	259.02	1.87	3.2	2.24	3
64	2.18	1.97	45.05	2.52	1.48	204864	5820	292.64	1.12	3.08	0.84	1
65	1.79	0.52	44.42	0.84	3	125955.56	3888.65	415.58	0.84	0.6	3.52	2
66	2.3	1.36	56.1	1.48	0.55	129600	3900	392.9	1.16	2.68	2.92	1
67	2.94	2.7	59.33	1.92	1.79	104661	2498.99	165.97	1.24	0.66	0.72	3
68	2.85	2.3	66.29	3.04	2	92014.29	1738.46	263.78	1.56	1.08	1.72	1
69	2.48	0.91	63.65	2.44	2.09	82468.75	2262	261	0.96	0.96	1.36	1
70	2.73	2.45	60.67	3.76	0.88	82880	1611.56	263.81	0.88	2.08	1.52	1
71	1.3	2.21	35.12	1.4	1.67	243000	5207.14	407.68	1.2	1.6	1.16	1
72	0.24	2.64	50.75	2.52	2.06	6128.69	3498.89	232.27	0.7	3.4	2.4	2
73	0.03	0.67	43.33	0.16	0.58	115714.29	4122.58	386.88	1.2	1.12	1.8	1
74	1.12	2.82	42.42	1.72	2.55	109381.82	3259.26	374.01	1.52	1.56	1.56	2
75	0.76	1.64	49.33	1.24	1.06	152343.75	2298.85	402.45	0.88	3.12	1.36	1
76	2.33	0.3	46.39	0.68	1.03	70815.79	1595.45	192.63	0.96	0.16	0.28	2
77	0.88	1.42	45.33	1.8	2.39	117333.33	1882.35	404.17	0.72	3.48	3.92	2
78	2.18	2.09	43	4.16	2.06	77575.9	1290	282.6	3	3.8	4	4
79	75	9	93.6	47	2	88.3	53	120	0.71	0.48	0.86	2
80	55	6	46.4	35	5	88.3	53	106	0.45	0.73	0.82	2
81	1.12	1.15	41.68	2.24	3	75331.7	3596	230	3	3.76	3.88	3
82	3	2	94.91	2	2	41580	1260	148	0.91	0.4	2	2
83	2	3	75.98	3	3	14112	588	146	0.58	0.43	3	2
84	3	3	98.87	2	2	29988	1428	148	1.62	0.41	2	2
85	3	3	83.26	2	3	26460	882	145	0.7	0.4	2	2
86	3	3	77.36	2	2	13230	630	146	0.71	0.38	2	2
87	3	1	39	2	2	65950	4380	259	1.8	3	3	4

TABLE 3. Normalized datasets.

Serial number	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	Stability level
1	1.00	0.33	0.59	0.50	0.67	0.20	0.75	0.90	1.00	1.00	1.00	4
2	0.67	0.67	0.74	0.50	0.33	0.05	0.23	0.61	0.50	1.00	1.00	2
3	1.00	1.00	0.56	0.50	0.67	0.14	0.43	0.73	0.75	1.00	1.00	3
4	1.00	0.67	0.76	0.50	1.00	0.14	0.29	0.66	1.00	1.00	1.00	3
5	1.00	0.33	0.42	0.75	0.67	0.77	0.89	0.81	0.75	1.00	1.00	4
6	1.00	1.00	0.79	0.50	1.00	0.59	1.00	0.72	1.00	1.00	1.00	3
7	0.67	0.33	0.80	0.25	0.33	0.20	0.51	0.74	0.25	1.00	1.00	1
8	0.67	0.33	0.89	0.50	0.33	0.19	0.38	0.78	0.50	0.25	0.25	2
9	0.33	0.33	1.00	0.50	0.33	0.05	0.21	0.82	0.25	1.00	1.00	1
10	0.33	0.67	0.76	0.75	0.67	1.00	0.70	0.92	0.50	1.00	1.00	3
11	0.33	0.67	0.89	0.25	0.33	0.15	0.39	0.74	0.50	1.00	1.00	1
12	0.33	0.33	0.92	0.50	0.33	0.08	0.25	0.71	0.25	0.25	0.25	1
13	0.33	0.33	0.79	0.50	0.33	0.08	0.45	0.49	0.25	0.75	0.75	1
14	0.33	0.33	0.83	0.25	0.33	0.13	0.42	0.51	0.25	0.75	0.75	1
15	1.00	1.00	0.85	0.50	0.67	0.05	0.41	0.51	0.25	0.75	0.75	2
16	0.33	0.33	0.82	0.25	0.33	0.08	0.31	0.51	0.50	0.75	0.75	1
17	0.33	0.67	0.86	0.50	0.67	0.11	0.28	1.00	0.75	1.00	1.00	2
18	0.33	0.33	0.82	0.50	0.33	0.07	0.35	0.73	0.50	0.25	0.25	2
19	0.33	0.33	0.83	0.25	0.33	0.19	0.34	0.49	0.50	0.75	0.75	2
20	0.33	0.33	0.79	0.25	0.33	0.16	0.87	0.47	0.25	0.75	0.75	1
21	0.33	0.33	0.77	0.25	0.33	0.18	0.57	0.44	0.50	0.75	0.75	2
22	0.33	0.67	0.82	0.25	0.67	0.04	0.23	0.44	0.50	1.00	0.75	2
23	0.33	0.67	0.83	0.25	0.33	0.05	0.30	0.44	0.50	1.00	1.00	2
24	0.33	0.33	0.80	0.25	0.33	0.05	0.33	0.56	0.25	0.75	0.75	1
25	0.33	0.33	0.80	0.25	0.33	0.07	0.20	0.56	0.50	0.75	0.75	2
26	0.33	0.33	0.82	0.25	0.33	0.17	0.43	0.56	0.50	0.25	0.25	1
27	0.33	0.33	0.80	0.25	0.33	0.14	0.43	0.56	0.50	0.25	0.25	1
28	0.33	0.33	0.65	0.50	1.00	0.34	0.61	0.56	0.75	1.00	1.00	3
29	0.33	0.33	0.83	0.25	0.33	0.05	0.25	0.56	0.50	0.25	0.25	1
30	0.67	0.33	0.67	0.75	0.67	0.64	0.30	0.85	0.75	0.75	0.75	3
31	0.33	0.33	0.79	0.25	0.67	0.05	0.21	0.85	0.25	0.25	0.25	1
32	0.33	0.33	0.85	0.25	0.67	0.07	0.23	0.91	0.50	0.75	0.75	2
33	0.33	0.33	0.86	0.25	0.33	0.12	0.64	0.51	0.50	0.75	0.75	2
34	0.67	0.67	0.62	0.75	1.00	0.11	0.25	0.51	0.75	0.75	0.75	3
35	0.33	0.67	0.61	0.75	0.67	0.19	0.29	0.71	0.75	1.00	0.75	3
36	0.67	0.67	0.65	1.00	0.67	0.35	0.22	0.71	0.75	1.00	1.00	3
37	0.33	0.67	0.64	0.50	0.67	0.55	0.80	0.97	0.75	1.00	1.00	4
38	1.00	1.00	1.00	0.45	0.13	0.11	0.11	1.00	1.00	0.29	1.00	1
39	0.82	0.89	1.00	0.75	0.75	0.22	0.22	1.00	0.67	0.74	0.95	2
40	0.84	0.78	0.64	0.98	1.00	0.82	0.82	0.98	0.43	1.00	0.81	3
41	0.99	0.89	0.64	0.78	0.63	0.52	0.52	0.94	0.61	0.80	0.99	2
42	0.87	1.00	0.64	0.48	0.50	0.26	0.26	0.95	0.82	0.41	0.75	2
43	0.92	0.67	0.46	1.00	0.88	1.00	1.00	0.77	0.37	0.92	0.54	3
44	0.94	1.00	0.73	0.51	0.38	0.17	0.17	0.70	0.84	0.51	0.94	2
45	0.56	0.62	0.97	0.52	0.37	0.14	0.42	0.83	0.68	0.28	0.26	2
46	0.54	0.51	0.96	0.26	0.34	0.26	1.00	0.53	0.35	0.77	0.76	1
47	0.52	1.00	0.99	0.22	0.68	0.06	0.27	0.51	0.71	1.00	0.74	2
48	0.49	0.46	1.00	0.23	0.39	0.07	0.40	0.63	0.35	0.76	0.73	1
49	0.62	0.48	0.98	0.24	0.32	0.26	0.52	0.63	0.71	0.27	0.22	1
50	0.49	0.51	0.79	0.53	0.93	0.50	0.72	0.65	0.99	0.90	0.95	3
51	0.94	0.51	0.76	0.72	0.69	1.00	0.40	1.00	0.94	0.71	0.76	3
52	1.00	0.93	0.76	0.72	0.99	0.17	0.28	0.57	0.94	0.73	0.78	3
53	0.93	0.93	0.81	1.00	0.69	0.57	0.27	0.78	1.00	1.00	1.00	4
54	0.86	1.00	1.00	0.95	1.00	0.01	0.10	0.87	0.50	1.00	0.72	4
55	1.00	0.94	0.69	1.00	0.86	0.01	0.11	0.71	0.45	0.92	0.57	3
56	0.54	0.56	0.76	0.54	1.00	0.55	0.73	0.62	0.97	0.91	0.97	3
57	0.03	0.25	0.61	0.03	0.02	1.00	1.00	1.00	1.00	0.28	1.00	2
58	0.68	0.38	0.68	0.65	0.05	0.29	0.38	0.49	0.20	0.06	0.23	3
59	0.04	0.12	0.71	0.20	0.31	0.31	0.33	0.46	0.39	0.07	0.34	2
60	0.48	0.48	0.66	0.54	0.19	0.72	0.83	0.72	0.53	0.95	0.33	3
61	0.59	0.19	0.60	0.68	0.83	0.34	0.23	0.73	0.95	0.92	0.71	3
62	0.57	1.00	1.00	0.49	0.85	0.11	0.18	0.63	0.45	0.88	0.98	3
63	0.02	0.64	0.72	0.62	0.65	0.47	1.00	0.62	0.62	0.84	0.56	3
64	0.74	0.46	0.62	0.61	0.49	0.84	0.87	0.70	0.37	0.81	0.21	1
65	0.61	0.12	0.61	0.20	1.00	0.52	0.58	1.00	0.28	0.16	0.88	2
66	0.78	0.32	0.77	0.36	0.18	0.53	0.58	0.95	0.39	0.71	0.73	1
67	1.00	0.64	0.81	0.46	0.60	0.43	0.37	0.40	0.41	0.17	0.18	3
68	0.97	0.54	0.91	0.73	0.67	0.38	0.26	0.63	0.52	0.28	0.43	1
69	0.84	0.21	0.87	0.59	0.70	0.34	0.34	0.63	0.32	0.25	0.34	1
70	0.93	0.58	0.83	0.90	0.29	0.34	0.24	0.63	0.29	0.55	0.38	1
71	0.44	0.52	0.48	0.34	0.56	1.00	0.78	0.98	0.40	0.42	0.29	1
72	0.08	0.62	0.69	0.61	0.69	0.03	0.52	0.56	0.23	0.89	0.60	2
73	0.01	0.16	0.59	0.04	0.19	0.48	0.62	0.93	0.40	0.29	0.45	1
74	0.38	0.67	0.58	0.41	0.85	0.45	0.49	0.90	0.51	0.41	0.39	2
75	0.26	0.39	0.67	0.30	0.35	0.63	0.34	0.97	0.29	0.82	0.34	1
76	0.79	0.07	0.63	0.16	0.34	0.29	0.24	0.46	0.32	0.04	0.07	2
77	0.30	0.33	0.62	0.43	0.80	0.48	0.28	0.97	0.24	0.92	0.98	2
78	0.74	0.49	0.59	1.00	0.69	0.32	0.19	0.68	1.00	1.00	1.00	4
79	1.00	1.00	1.00	1.00	0.40	1.00	1.00	1.00	1.00	0.66	1.00	2
80	0.73	0.67	0.50	0.74	1.00	1.00	1.00	0.88	0.63	1.00	0.95	2
81	0.37	0.38	0.42	0.75	1.00	1.00	0.82	0.89	1.00	1.00	1.00	3
82	1.00	0.67	0.96	0.67	0.67	0.55	0.29	0.57	0.30	0.11	0.52	2
83	0.67	1.00	0.77	1.00	1.00	0.19	0.13	0.56	0.19	0.11	0.77	2
84	1.00	1.00	1.00	0.67	0.67	0.40	0.33	0.57	0.54	0.11	0.52	2
85	1.00	1.00	0.84	0.67	1.00	0.35	0.20	0.56	0.23	0.11	0.52	2
86	1.00	1.00	0.78	0.67	0.67	0.18	0.14	0.56	0.24	0.10	0.52	2
87	1.00	0.33	0.39	0.67	0.67	0.88	1.00	1.00	0.60	0.80	0.77	4

TABLE 4. Processed dataset.

Serial number	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	Stability level
1	1.00	0.33	0.59	0.50	0.67	0.20	0.75	0.90	1.00	1.00	1.00	B
2	0.67	0.67	0.74	0.50	0.33	0.05	0.23	0.61	0.50	1.00	1.00	A
3	1.00	1.00	0.56	0.50	0.67	0.14	0.43	0.73	0.75	1.00	1.00	B
4	1.00	0.67	0.76	0.50	1.00	0.14	0.29	0.66	1.00	1.00	1.00	B
5	1.00	0.33	0.42	0.75	0.67	0.77	0.89	0.81	0.75	1.00	1.00	B
7	0.67	0.33	0.80	0.25	0.33	0.20	0.51	0.74	0.25	1.00	1.00	A
8	0.67	0.33	0.89	0.50	0.33	0.19	0.38	0.78	0.50	0.25	0.25	A
9	0.33	0.33	1.00	0.50	0.33	0.05	0.21	0.82	0.25	1.00	1.00	A
10	0.33	0.67	0.76	0.75	0.67	1.00	0.70	0.92	0.50	1.00	1.00	B
11	0.33	0.67	0.89	0.25	0.33	0.15	0.39	0.74	0.50	1.00	1.00	A
12	0.33	0.33	0.92	0.50	0.33	0.08	0.25	0.71	0.25	0.25	0.25	A
13	0.33	0.33	0.79	0.50	0.33	0.08	0.45	0.49	0.25	0.75	0.75	A
14	0.33	0.33	0.83	0.25	0.33	0.13	0.42	0.51	0.25	0.75	0.75	A
16	0.33	0.33	0.82	0.25	0.33	0.08	0.31	0.51	0.50	0.75	0.75	A
17	0.33	0.67	0.86	0.50	0.67	0.11	0.28	1.00	0.75	1.00	1.00	A
18	0.33	0.33	0.82	0.50	0.33	0.07	0.35	0.73	0.50	0.25	0.25	A
19	0.33	0.33	0.83	0.25	0.33	0.19	0.34	0.49	0.50	0.75	0.75	A
20	0.33	0.33	0.79	0.25	0.33	0.16	0.87	0.47	0.25	0.75	0.75	A
21	0.33	0.33	0.77	0.25	0.33	0.18	0.57	0.44	0.50	0.75	0.75	A
22	0.33	0.67	0.82	0.25	0.67	0.04	0.23	0.44	0.50	1.00	0.75	A
23	0.33	0.67	0.83	0.25	0.33	0.05	0.30	0.44	0.50	1.00	1.00	A
24	0.33	0.33	0.80	0.25	0.33	0.05	0.33	0.56	0.25	0.75	0.75	A
25	0.33	0.33	0.80	0.25	0.33	0.07	0.20	0.56	0.50	0.75	0.75	A
26	0.33	0.33	0.82	0.25	0.33	0.17	0.43	0.56	0.50	0.25	0.25	A
27	0.33	0.33	0.80	0.25	0.33	0.14	0.43	0.56	0.50	0.25	0.25	A
28	0.33	0.33	0.65	0.50	1.00	0.34	0.61	0.56	0.75	1.00	1.00	B
29	0.33	0.33	0.83	0.25	0.33	0.05	0.25	0.56	0.50	0.25	0.25	A
30	0.67	0.33	0.67	0.75	0.67	0.64	0.30	0.85	0.75	0.75	0.75	B
32	0.33	0.33	0.85	0.25	0.67	0.07	0.23	0.91	0.50	0.75	0.75	A
33	0.33	0.33	0.86	0.25	0.33	0.12	0.64	0.51	0.50	0.75	0.75	A
34	0.67	0.67	0.62	0.75	1.00	0.11	0.25	0.51	0.75	0.75	0.75	B
35	0.33	0.67	0.61	0.75	0.67	0.19	0.29	0.71	0.75	1.00	0.75	B
36	0.67	0.67	0.65	1.00	0.67	0.35	0.22	0.71	0.75	1.00	1.00	B
37	0.33	0.67	0.64	0.50	0.67	0.55	0.80	0.97	0.75	1.00	1.00	B
38	1.00	1.00	1.00	0.45	0.13	0.11	0.11	1.00	1.00	0.29	1.00	A
39	0.82	0.89	1.00	0.75	0.75	0.22	0.22	1.00	0.67	0.74	0.95	A
40	0.84	0.78	0.64	0.98	1.00	0.82	0.82	0.98	0.43	1.00	0.81	B
41	0.99	0.89	0.64	0.78	0.63	0.52	0.52	0.94	0.61	0.80	0.99	A
42	0.87	1.00	0.64	0.48	0.50	0.26	0.26	0.95	0.82	0.41	0.75	A
43	0.92	0.67	0.46	1.00	0.88	1.00	1.00	0.77	0.37	0.92	0.54	B
44	0.94	1.00	0.73	0.51	0.38	0.17	0.17	0.70	0.84	0.51	0.94	A
45	0.56	0.62	0.97	0.52	0.37	0.14	0.42	0.83	0.68	0.28	0.26	A
46	0.54	0.51	0.96	0.26	0.34	0.26	1.00	0.53	0.35	0.77	0.76	A
47	0.52	1.00	0.99	0.22	0.68	0.06	0.27	0.51	0.71	1.00	0.74	A
48	0.49	0.46	1.00	0.23	0.39	0.07	0.40	0.63	0.35	0.76	0.73	A
49	0.62	0.48	0.98	0.24	0.32	0.26	0.52	0.63	0.71	0.27	0.22	A
50	0.49	0.51	0.79	0.53	0.93	0.50	0.72	0.65	0.99	0.90	0.95	B
51	0.94	0.51	0.76	0.72	0.69	1.00	0.40	1.00	0.94	0.71	0.76	B
52	1.00	0.93	0.76	0.72	0.99	0.17	0.28	0.57	0.94	0.73	0.78	B
53	0.93	0.93	0.81	1.00	0.69	0.57	0.27	0.78	1.00	1.00	1.00	B
54	0.86	1.00	1.00	0.95	1.00	0.01	0.10	0.87	0.50	1.00	0.72	B
55	1.00	0.94	0.69	1.00	0.86	0.01	0.11	0.71	0.45	0.92	0.57	B
56	0.54	0.56	0.76	0.54	1.00	0.55	0.73	0.62	0.97	0.91	0.97	B
57	0.03	0.25	0.61	0.03	0.02	1.00	1.00	1.00	1.00	0.28	1.00	A
58	0.68	0.38	0.68	0.65	0.05	0.29	0.38	0.49	0.20	0.06	0.23	B
59	0.04	0.12	0.71	0.20	0.31	0.31	0.33	0.46	0.39	0.07	0.34	A
60	0.48	0.48	0.66	0.54	0.19	0.72	0.83	0.72	0.53	0.95	0.33	B
61	0.59	0.19	0.60	0.68	0.83	0.34	0.23	0.73	0.95	0.92	0.71	B
62	0.57	1.00	1.00	0.49	0.85	0.11	0.18	0.63	0.45	0.88	0.98	B
63	0.02	0.64	0.72	0.62	0.65	0.47	1.00	0.62	0.62	0.84	0.56	B
64	0.74	0.46	0.62	0.61	0.49	0.84	0.87	0.70	0.37	0.81	0.21	A
65	0.61	0.12	0.61	0.20	1.00	0.52	0.58	1.00	0.28	0.16	0.88	A
66	0.78	0.32	0.77	0.36	0.18	0.53	0.58	0.95	0.39	0.71	0.73	A
67	1.00	0.64	0.81	0.46	0.60	0.43	0.37	0.40	0.41	0.17	0.18	B
68	0.97	0.54	0.91	0.73	0.67	0.38	0.26	0.63	0.52	0.28	0.43	A
69	0.84	0.21	0.87	0.59	0.70	0.34	0.34	0.63	0.32	0.25	0.34	A
70	0.93	0.58	0.83	0.90	0.29	0.34	0.24	0.63	0.29	0.55	0.38	A
71	0.44	0.52	0.48	0.34	0.56	1.00	0.78	0.98	0.40	0.42	0.29	A
72	0.08	0.62	0.69	0.61	0.69	0.03	0.52	0.56	0.23	0.89	0.60	A
73	0.01	0.16	0.59	0.04	0.19	0.48	0.62	0.93	0.40	0.29	0.45	A
74	0.38	0.67	0.58	0.41	0.85	0.45	0.49	0.90	0.51	0.41	0.39	A
75	0.26	0.39	0.67	0.30	0.35	0.63	0.34	0.97	0.29	0.82	0.34	A
76	0.79	0.07	0.63	0.16	0.34	0.29	0.24	0.46	0.32	0.04	0.07	A
77	0.30	0.33	0.62	0.43	0.80	0.48	0.28	0.97	0.24	0.92	0.98	A
78	0.74	0.49	0.59	1.00	0.69	0.32	0.19	0.68	1.00	1.00	1.00	B
79	1.00	1.00	1.00	1.00	0.40	1.00	1.00	1.00	1.00	0.66	1.00	A
80	0.73	0.67	0.50	0.74	1.00	1.00	1.00	0.88	0.63	1.00	0.95	A
81	0.37	0.38	0.42	0.75	1.00	1.00	0.82	0.89	1.00	1.00	1.00	B
82	1.00	0.67	0.96	0.67	0.67	0.55	0.29	0.57	0.30	0.11	0.52	A
83	0.67	1.00	0.77	1.00	1.00	0.19	0.13	0.56	0.19	0.11	0.77	A
84	1.00	1.00	1.00	0.67	0.67	0.40	0.33	0.57	0.54	0.11	0.52	A
85	1.00	1.00	0.84	0.67	1.00	0.35	0.20	0.56	0.23	0.11	0.52	A
86	1.00	1.00	0.78	0.67	0.67	0.18	0.14	0.56	0.24	0.10	0.52	A
87	1.00	0.33	0.39	0.67	0.67	0.88	1.00	1.00	0.60	0.80	0.77	B

TABLE 5. Dataset of Model 2.

Serial number	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	Stability level
2	0.67	0.67	0.74	0.50	0.33	0.05	0.23	0.61	0.50	1.00	1.00	2
7	0.67	0.33	0.80	0.25	0.33	0.20	0.51	0.74	0.25	1.00	1.00	1
8	0.67	0.33	0.89	0.50	0.33	0.19	0.38	0.78	0.50	0.25	0.25	2
9	0.33	0.33	1.00	0.50	0.33	0.05	0.21	0.82	0.25	1.00	1.00	1
11	0.33	0.67	0.89	0.25	0.33	0.15	0.39	0.74	0.50	1.00	1.00	1
12	0.33	0.33	0.92	0.50	0.33	0.08	0.25	0.71	0.25	0.25	0.25	1
13	0.33	0.33	0.79	0.50	0.33	0.08	0.45	0.49	0.25	0.75	0.75	1
14	0.33	0.33	0.83	0.25	0.33	0.13	0.42	0.51	0.25	0.75	0.75	1
16	0.33	0.33	0.82	0.25	0.33	0.08	0.31	0.51	0.50	0.75	0.75	1
17	0.33	0.67	0.86	0.50	0.67	0.11	0.28	1.00	0.75	1.00	1.00	2
18	0.33	0.33	0.82	0.50	0.33	0.07	0.35	0.73	0.50	0.25	0.25	2
19	0.33	0.33	0.83	0.25	0.33	0.19	0.34	0.49	0.50	0.75	0.75	2
20	0.33	0.33	0.79	0.25	0.33	0.16	0.87	0.47	0.25	0.75	0.75	1
21	0.33	0.33	0.77	0.25	0.33	0.18	0.57	0.44	0.50	0.75	0.75	2
22	0.33	0.67	0.82	0.25	0.67	0.04	0.23	0.44	0.50	1.00	0.75	2
23	0.33	0.67	0.83	0.25	0.33	0.05	0.30	0.44	0.50	1.00	1.00	2
24	0.33	0.33	0.80	0.25	0.33	0.05	0.33	0.56	0.25	0.75	0.75	1
25	0.33	0.33	0.80	0.25	0.33	0.07	0.20	0.56	0.50	0.75	0.75	2
26	0.33	0.33	0.82	0.25	0.33	0.17	0.43	0.56	0.50	0.25	0.25	1
27	0.33	0.33	0.80	0.25	0.33	0.14	0.43	0.56	0.50	0.25	0.25	1
29	0.33	0.33	0.83	0.25	0.33	0.05	0.25	0.56	0.50	0.25	0.25	1
32	0.33	0.33	0.85	0.25	0.67	0.07	0.23	0.91	0.50	0.75	0.75	2
33	0.33	0.33	0.86	0.25	0.33	0.12	0.64	0.51	0.50	0.75	0.75	2
38	1.00	1.00	1.00	0.45	0.13	0.11	0.11	1.00	1.00	0.29	1.00	1
39	0.82	0.89	1.00	0.75	0.75	0.22	0.22	1.00	0.67	0.74	0.95	2
41	0.99	0.89	0.64	0.78	0.63	0.52	0.52	0.94	0.61	0.80	0.99	2
42	0.87	1.00	0.64	0.48	0.50	0.26	0.26	0.95	0.82	0.41	0.75	2
44	0.94	1.00	0.73	0.51	0.38	0.17	0.17	0.70	0.84	0.51	0.94	2
45	0.56	0.62	0.97	0.52	0.37	0.14	0.42	0.83	0.68	0.28	0.26	2
46	0.54	0.51	0.96	0.26	0.34	0.26	1.00	0.53	0.35	0.77	0.76	1
47	0.52	1.00	0.99	0.22	0.68	0.06	0.27	0.51	0.71	1.00	0.74	2
48	0.49	0.46	1.00	0.23	0.39	0.07	0.40	0.63	0.35	0.76	0.73	1
49	0.62	0.48	0.98	0.24	0.32	0.26	0.52	0.63	0.71	0.27	0.22	1
57	0.03	0.25	0.61	0.03	0.02	1.00	1.00	1.00	1.00	0.28	1.00	2
59	0.04	0.12	0.71	0.20	0.31	0.31	0.33	0.46	0.39	0.07	0.34	2
64	0.74	0.46	0.62	0.61	0.49	0.84	0.87	0.70	0.37	0.81	0.21	1
65	0.61	0.12	0.61	0.20	1.00	0.52	0.58	1.00	0.28	0.16	0.88	2
66	0.78	0.32	0.77	0.36	0.18	0.53	0.58	0.95	0.39	0.71	0.73	1
68	0.97	0.54	0.91	0.73	0.67	0.38	0.26	0.63	0.52	0.28	0.43	1
69	0.84	0.21	0.87	0.59	0.70	0.34	0.34	0.63	0.32	0.25	0.34	1
70	0.93	0.58	0.83	0.90	0.29	0.34	0.24	0.63	0.29	0.55	0.38	1
71	0.44	0.52	0.48	0.34	0.56	1.00	0.78	0.98	0.40	0.42	0.29	1
72	0.08	0.62	0.69	0.61	0.69	0.03	0.52	0.56	0.23	0.89	0.60	2
73	0.01	0.16	0.59	0.04	0.19	0.48	0.62	0.93	0.40	0.29	0.45	1
74	0.38	0.67	0.58	0.41	0.85	0.45	0.49	0.90	0.51	0.41	0.39	2
75	0.26	0.39	0.67	0.30	0.35	0.63	0.34	0.97	0.29	0.82	0.34	1
76	0.79	0.07	0.63	0.16	0.34	0.29	0.24	0.46	0.32	0.04	0.07	2
77	0.30	0.33	0.62	0.43	0.80	0.48	0.28	0.97	0.24	0.92	0.98	2
79	1.00	1.00	1.00	1.00	0.40	1.00	1.00	1.00	1.00	0.66	1.00	2
80	0.73	0.67	0.50	0.74	1.00	1.00	1.00	0.88	0.63	1.00	0.95	2
82	1.00	0.67	0.96	0.67	0.67	0.55	0.29	0.57	0.30	0.11	0.52	2
83	0.67	1.00	0.77	1.00	1.00	0.19	0.13	0.56	0.19	0.11	0.77	2
84	1.00	1.00	1.00	0.67	0.67	0.40	0.33	0.57	0.54	0.11	0.52	2
85	1.00	1.00	0.84	0.67	1.00	0.35	0.20	0.56	0.23	0.11	0.52	2
86	1.00	1.00	0.78	0.67	0.67	0.18	0.14	0.56	0.24	0.10	0.52	2

aBoost algorithm can be evaluated visually. The evaluation results of the AdaBoost algorithm are shown in Table 9.

The TrAdaBoost algorithm, AdaBoost algorithm, TrBys and TrSVM algorithms were used to build models to evaluate the verification set, and the results are shown in Fig. 6. Among these four methods, the closer the evaluation result to the real result, the better the evaluation method.

It can be seen from Fig. 6 that the evaluation result of the TrAdaBoost model is much closer to the real result than the other three models. Among the 30 samples, there were seven key sample points (Evaluation Level 4), and the TrAdaBoost model evaluated all seven correctly. In contrast, the evaluation results of the AdaBoost model matches four points while the other two models match less than four. In the evaluation of goaf stability levels, the evaluation of unstable goaf is the most important. However, the training results of the AdaBoost, TrBys, and TrSVM models were relatively stable and insensitive to the key point information. Therefore, the TrAdaBoost model performed better and was more accurate than the other three models in predicting goaf stability.

TABLE 6. Dataset of Model 3.

Serial number	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	Stability level
1	1.00	0.33	0.59	0.50	0.67	0.20	0.75	0.90	1.00	1.00	1.00	4
3	1.00	1.00	0.56	0.50	0.67	0.14	0.43	0.73	0.75	1.00	1.00	3
4	1.00	0.67	0.76	0.50	1.00	0.14	0.29	0.66	1.00	1.00	1.00	3
5	1.00	0.33	0.42	0.75	0.67	0.77	0.89	0.81	0.75	1.00	1.00	4
10	0.33	0.67	0.76	0.75	0.67	1.00	0.70	0.92	0.50	1.00	1.00	3
28	0.33	0.33	0.65	0.50	1.00	0.34	0.61	0.56	0.75	1.00	1.00	3
30	0.67	0.33	0.67	0.75	0.67	0.64	0.30	0.85	0.75	0.75	0.75	3
34	0.67	0.67	0.62	0.75	1.00	0.11	0.25	0.51	0.75	0.75	0.75	3
35	0.33	0.67	0.61	0.75	0.67	0.19	0.29	0.71	0.75	1.00	0.75	3
36	0.67	0.67	0.65	1.00	0.67	0.35	0.22	0.71	0.75	1.00	1.00	3
37	0.33	0.67	0.64	0.50	0.67	0.55	0.80	0.97	0.75	1.00	1.00	4
40	0.84	0.78	0.64	0.98	1.00	0.82	0.82	0.98	0.43	1.00	0.81	3
43	0.92	0.67	0.46	1.00	0.88	1.00	1.00	0.77	0.37	0.92	0.54	3
50	0.49	0.51	0.79	0.53	0.93	0.50	0.72	0.65	0.99	0.90	0.95	3
51	0.94	0.51	0.76	0.72	0.69	1.00	0.40	1.00	0.94	0.71	0.76	3
52	1.00	0.93	0.76	0.72	0.99	0.17	0.28	0.57	0.94	0.73	0.78	3
53	0.93	0.93	0.81	1.00	0.69	0.57	0.27	0.78	1.00	1.00	1.00	4
54	0.86	1.00	1.00	0.95	1.00	0.01	0.10	0.87	0.50	1.00	0.72	4
55	1.00	0.94	0.69	1.00	0.86	0.01	0.11	0.71	0.45	0.92	0.57	3
56	0.54	0.56	0.76	0.54	1.00	0.55	0.73	0.62	0.97	0.91	0.97	3
58	0.68	0.38	0.68	0.65	0.05	0.29	0.38	0.49	0.20	0.06	0.23	3
60	0.48	0.48	0.66	0.54	0.19	0.72	0.83	0.72	0.53	0.95	0.33	3
61	0.59	0.19	0.60	0.68	0.83	0.34	0.23	0.73	0.95	0.92	0.71	3
62	0.57	1.00	1.00	0.49	0.85	0.11	0.18	0.63	0.45	0.88	0.98	3
63	0.02	0.64	0.72	0.62	0.65	0.47	1.00	0.62	0.62	0.84	0.56	3
67	1.00	0.64	0.81	0.46	0.60	0.43	0.37	0.40	0.41	0.17	0.18	3
78	0.74	0.49	0.59	1.00	0.69	0.32	0.19	0.68	1.00	1.00	1.00	4
81	0.37	0.38	0.42	0.75	1.00	1.00	0.82	0.89	1.00	1.00	1.00	3
87	1.00	0.33	0.39	0.67	0.67	0.88	1.00	1.00	0.60	0.80	0.77	4

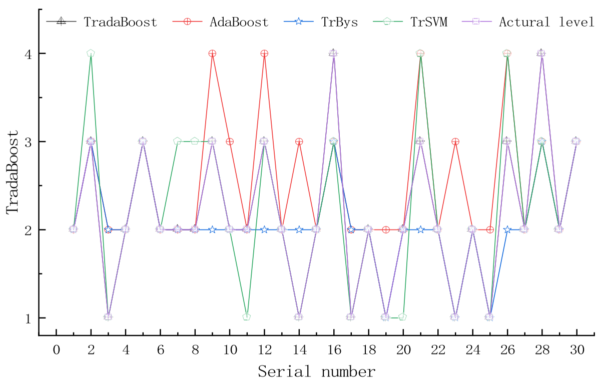

FIGURE 6. Comparison of evaluation results.

The mean square error of the evaluation results of the four models was calculated. The mean square error represents the deviation between the evaluation value and the real value of validated data. The larger the deviation, the worse the evaluation result. The mean square error MSE is given by:

$$MSE\left(y, \widehat{y}\right) = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} \left(y_i - \widehat{y}_i\right)^2 \tag{16}$$

In Eq. (16), $MSE\left(y, \hat{y}\right)$ represents the mean square error of the model, $n_{samples}$ represents the number of original data samples, y_i is the actual stability level of the i-th sample, and $\hat{y}_i$ is the evaluation result of the i-th sample. By calculating the evaluation results, the MSE value of TrAdaBoost is 0.1, AdaBoost is 0.63, TrBys is 0.3, and TrSVM is 0.3. Comparing the MSE values of the four algorithms, the TrAdaBoost model performs better than the other models.

It can be concluded from the comparison that the TrAdaBoost model has higher sensitivity and greater accuracy in the evaluation of key points, and can identify accurately the unstable goaf, which demonstrates the significance of the training model. This also proves that by establishing transfer learning between goafs of different mines and utilizing datasets from other mines, the accuracy in evaluating goaf stability can be greatly improved. In particular, when the accurate evaluation of high-risk goaf is a concern, the model provides support for mine managers to take corresponding safety measures, and provides a guarantee of mine safety.

2) F1 MEASURE

The TrAdaBoost model not only has superior accuracy in the verification dataset, but also has good generalization ability. The TrAdaBoost model was trained based on 104 training datasets and performed well on existing verification datasets. However, the accuracy of the model does not mean that the model will still perform well when applied to new objects. When the model is applied to a new underground mine, the data distribution and characteristics may not be consistent with the training data. Therefore, it is necessary to evaluate the model's generalization ability. The F1 metric was used for this purpose.

The F1 metric can reflect comprehensively the special types of evaluation results in the two-level classification problem, and the basis of the F1 metric is precision and recall. According to the combination of verification data and

TABLE 7. Test data.

Serial number	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	Stability level
1	0.68	0.55	0.75	0.72	0.05	0.36	0.44	0.39	0.37	0.07	0.23	1
2	0.04	0.18	0.79	0.22	0.31	0.23	0.37	0.25	0.36	0.08	0.35	1
3	0.47	0.71	0.73	0.60	0.19	0.51	0.96	0.40	0.50	0.99	0.34	2
4	0.59	0.28	0.66	0.76	0.84	0.24	0.26	0.40	0.89	0.96	0.72	3
5	0.57	0.73	0.76	0.81	0.85	0.38	0.58	0.39	0.48	0.92	1.00	2
6	0.02	0.63	0.80	0.68	0.43	0.48	1.00	0.39	0.44	0.88	0.57	2
7	0.73	0.68	0.68	0.67	0.49	0.62	1.00	0.39	0.35	0.85	0.21	1
8	0.60	0.18	0.67	0.22	1.00	0.37	0.66	0.55	0.26	0.16	0.90	4
9	0.77	0.47	0.85	0.39	0.18	0.39	0.65	0.52	0.36	0.74	0.74	1
10	0.98	0.93	0.89	0.51	0.60	0.32	0.43	0.34	0.39	0.73	0.73	2
11	0.95	0.79	1.00	0.81	0.67	0.27	0.30	0.35	0.49	0.30	0.44	2
12	0.83	0.31	0.96	0.65	0.70	0.25	0.39	0.35	0.30	0.26	0.35	1
13	0.91	0.84	0.91	1.00	0.29	0.23	0.28	0.35	0.27	0.57	0.39	1
14	0.43	0.76	0.53	0.37	0.56	0.71	0.90	0.53	0.37	0.44	0.30	2
15	0.08	0.92	0.58	0.88	0.34	0.36	0.67	0.37	0.29	0.93	0.61	1
16	0.01	0.23	0.65	0.04	0.19	0.32	0.71	0.51	0.37	0.31	0.46	1
17	0.37	0.97	0.64	0.46	0.85	0.33	0.55	0.50	0.48	0.43	0.40	3
18	0.25	0.56	0.74	0.33	0.35	0.46	0.40	0.53	0.27	0.86	0.35	1
19	0.78	0.74	0.70	0.18	0.34	0.21	0.27	0.25	0.30	0.04	0.07	1
20	0.29	0.49	0.68	0.48	0.80	0.31	0.32	0.53	0.22	0.96	1.00	3
21	0.63	0.89	0.23	0.90	0.33	0.16	0.37	0.71	0.63	0.88	0.41	2
22	0.29	0.44	0.27	0.02	0.94	0.28	0.35	0.73	0.27	0.89	0.80	3
23	0.02	0.95	0.29	0.31	0.26	0.27	0.46	0.77	0.48	0.97	0.08	1
24	1.00	0.77	0.13	0.82	0.33	0.25	0.39	0.77	0.64	0.74	0.83	2
25	0.48	0.27	0.23	0.04	0.78	0.19	0.29	0.96	0.51	0.32	0.63	3
26	0.15	0.94	0.16	0.40	0.15	0.36	0.68	1.00	0.48	0.41	0.17	2
27	0.93	0.38	0.21	0.85	0.37	0.71	0.93	0.96	0.49	0.71	0.91	2
28	0.94	0.34	0.17	0.07	0.35	0.62	0.83	0.92	0.60	0.82	0.93	2
29	0.75	0.72	0.28	0.47	0.65	1.00	0.81	0.98	0.52	1.00	0.65	3
30	0.19	0.13	0.28	0.60	0.17	0.57	0.44	0.98	0.52	0.13	0.48	2
31	0.98	0.91	0.20	0.54	0.47	0.41	0.62	0.65	0.26	0.30	0.90	2
32	0.15	1.00	0.15	0.38	0.66	0.46	0.47	0.69	0.46	0.59	0.85	3
33	0.88	0.64	0.16	0.66	0.30	0.26	0.41	0.73	0.43	0.96	0.34	2
34	0.29	0.40	0.22	0.85	0.47	0.57	0.74	0.78	0.35	0.12	0.12	1
35	0.90	0.53	0.20	0.69	0.59	0.40	0.55	0.98	0.52	1.00	0.62	2
36	0.79	0.28	0.25	0.72	0.99	0.31	0.31	0.97	0.33	0.33	0.82	4
37	0.03	0.89	0.17	0.23	0.17	0.19	0.29	0.96	0.51	0.38	0.10	1
38	0.09	0.72	0.22	0.52	0.56	0.22	0.38	0.58	1.00	0.88	0.61	2
39	0.90	0.51	0.24	0.22	0.18	0.25	0.48	0.65	0.25	0.95	0.57	1
40	0.59	0.10	0.22	0.33	0.41	0.31	0.57	0.73	0.48	0.76	0.12	2
41	0.45	0.88	0.17	0.31	0.71	0.54	0.78	0.67	0.37	0.03	0.59	3
42	0.53	0.85	0.25	0.96	0.48	0.55	0.89	0.68	0.37	0.49	0.44	2
43	0.01	0.97	0.28	0.85	0.53	0.37	0.53	0.67	0.36	0.23	0.19	1
44	0.58	0.06	0.23	0.87	0.19	0.33	0.51	0.77	0.54	0.86	0.36	2
45	0.18	0.27	0.29	0.15	0.18	0.37	0.56	0.67	0.36	0.66	0.26	1
46	0.27	0.19	0.15	0.21	0.65	0.25	0.35	0.73	0.54	0.57	0.15	3
47	0.68	0.39	0.25	0.86	0.10	0.23	0.33	0.77	0.51	0.96	0.52	2
48	0.51	0.89	0.27	0.32	0.77	0.30	0.40	0.96	0.81	0.51	0.47	4
49	0.66	0.68	0.24	0.28	0.11	0.21	0.68	0.94	0.92	0.76	0.84	2
50	0.08	0.73	0.27	0.53	0.76	0.26	0.54	0.95	0.52	0.53	0.48	3

model evaluation categories, the verification data is divided into four cases: TP (True Positive); FP (False Positive); TN (True Negative); and FN (False Negative), where TP + FP + TN + FN is equal to the total amount of verification data. The confusion matrix of the classification results is shown in Table 10.

According to the definition of a confusion matrix, precision P and recall R are defined as:

$$P = \frac{TP}{TP + FP} \tag{17}$$

$$R = \frac{TP}{TP + FN} \tag{18}$$

The precision and recall rates provide a set of conflicting data. For models with high precision, due to the high purity of the evaluation results, some positive examples are inevitably excluded. On the other hand, in order to classify all the positive samples as correctly as possible, the model with a high recall rate will have a low judgment threshold, which will result in some negative examples being misclassified as positive. In order to measure comprehensively the performance of

TABLE 8. Prediction results of TrAdaBoost model.

Serial number	1	2	3	4	5	6	7	8	9	10
Actual level	2	3	1	2	3	2	2	2	3	2
TrAdaBoost	2	3	1	2	4	2	2	2	3	2
Serial number	11	12	13	14	15	16	17	18	19	20
Actual level	2	3	2	1	2	4	1	2	1	2
TrAdaBoost	2	3	2	1	2	4	1	2	1	1
Serial number	21	22	23	24	25	26	27	28	29	30
Actual level	3	2	1	2	1	3	2	4	2	3
TrAdaBoost	4	2	1	2	1	3	2	4	2	3

TABLE 9. Prediction results of AdaBoost model.

Serial number	1	2	3	4	5	6	7	8	9	10
Actual level	2	3	1	2	3	2	2	2	3	2
AdaBoost	2	3	2	2	3	2	2	2	4	3
TrBys	2	3	2	2	3	2	2	2	2	2
TrSVM	2	4	1	2	3	2	3	3	3	2
Serial number	11	12	13	14	15	16	17	18	19	20
Actual level	2	3	2	1	2	4	1	2	1	2
AdaBoost	2	4	2	3	2	3	2	2	2	2
TrBys	2	2	2	2	2	3	2	2	1	2
TrSVM	1	3	2	1	2	3	1	2	1	1
Serial number	21	22	23	24	25	26	27	28	29	30
Actual level	3	2	1	2	1	3	2	4	2	3
AdaBoost	4	2	3	2	2	4	2	3	2	3
TrBys	2	2	1	2	1	2	2	3	2	3
TrSVM	4	2	1	2	1	4	2	3	2	3

TABLE 10. Confusion matrix.

Real situation	Evaluation results	
	Positive example	Counterexample
Positive example	TP	FN
Counterexample	FP	TN

TABLE 11. Confusion matrix of TrAdaBoost model.

Real situation	Evaluation results	
	Positive example	Counterexample
Positive example	2	0
Counterexample	2	26

TABLE 12. Confusion matrix of AdaBoost model.

Real situation	Evaluation results	
	Positive example	Counterexample
Positive example	0	2
Counterexample	0	8

model precision and recall, the F1 metric is introduced, and is defined as:

$$F1 = \frac{2 \times P \times R}{P + R} \tag{19}$$

F1 is actually a harmonic average of the precision and recall. The larger the F1, the better the overall performance of the model. In this paper, the goafs with Level 4 stability are the key tags to be identified in the evaluation. Therefore, these goafs are set as positive examples, and the goafs of other stability levels are set as counterexamples. Based on the information in Table 8 and Table 9, the confusion matrix of the TrAdaBoost model is shown in Table 11.

The confusion matrix of the AdaBoost model is shown in Table 12.

The F1 values of TrAdaBoost and AdaBoost are calculated separately. Firstly, the precision and recall of the TrAdaBoost model are calculated according to formulae (20) and (21).

$$P = \frac{2}{2+2} = 0.5 \tag{20}$$

$$R = \frac{2}{2+0} = 1 \tag{21}$$

The F1 value of the TrAdaBoost model can be calculated from formula (19):

$$F1 = \frac{2 \times 0.5 \times 1}{0.5 + 1} = 0.66 \tag{22}$$

The precision and recall values of the other three algorithms can be obtained similarly, and their values are all 0. That is, the precision of the AdaBoost, TrBys, and TrSVM models is non-existent; hence their F1 values are also non-existent. In contrast, the F1 value of the TrAdaBoost model is 0.66, which indicates that the improved TrAdaBoost has a strong generalization performance. However, the accuracy and recall of the other three models are poor. If we consider the mean square error and F1 metric of the improved TrAdaBoost model, the performance of the model is excellent. At present, with the development of big data and artificial intelligence, the rational and efficient use of existing data can help mine owners to evaluate goaf stability, and provides technical support for improving the efficiency of mining and ensuring safe production.

IV. CONCLUSION

Data from a large number of goafs in underground metal mines have not been fully utilized in previous research on goaf stability. However, there are differences in the characteristics of different goafs, and differences in the assignment of qualitative indicators in the evaluation system established by different experts. Combined with the results of engineering practice, a transfer learning model was used in this paper to study the evaluation of goaf stability. Accurate evaluation of goaf stability was achieved on the basis of a small sample size of the target mine and provides technical support for ensuring mine safety. The contributions of this paper can be summarized as follows:

(1) The TrAdaBoost algorithm was improved, and the scope of the TrAdaBoost model was extended from two-level classification to multi-level classification, which makes it suitable for the evaluation of goaf stability.

(2) The isolated forest method was used to filter the bad points of the auxiliary training set, and data in the auxiliary training set was tested to eliminate the influence of abnormal data on the algorithm.

(3) The dynamic factor concept was introduced to solve the problem that the weight of the source domain data decreases too quickly and irreversibly, which enhanced the generalization performance of the algorithm for different goaf samples.

(4) Based on the improved TrAdaBoost algorithm and artificial intelligence transfer learning theory, this paper constructed a model to predict goaf stability. The prediction accuracy and generalization ability of the model were evaluated by means of mean square error and F1 measurements, which prove that the performance of the model is excellent.

(5) Research has shown that the model can still maintain a high level of accuracy with a small sample size when evaluating goaf stability in underground metal mines.

REFERENCES

[1] L. Liu and Z. Chen, "Application of fuzzy set pair in stability evaluation of mining goaf," *J. Central South Univ.*, vol. 46, no. 7, pp. 2665–2672, Jul. 2015.

[2] X.-B. Xie, R.-N. Deng, X.-J. Dong, and Z.-Z. Yan, "Stability of goaf group system based on catastrophe theory and rheological theory," *Rock Soil Mech.*, vol. 39, no. 6, pp. 1963–1972, Jun. 2018.

[3] Z. Wang, J.-P Guo, and L.-G Wang, "Recognition of goaf risk based on support vector machines method," *J. Chongqing Univ.*, vol. 38, no. 4, pp. 85–90+127, Aug. 2015.
[4] W. Song, X. Wenbin, and D. U. Jianhua, "Deformation and failure mechanism of cap rock in mining a gently inclined and extremely thin iron mine by the long-wall method," *J. Univ. Sci. Technol.*, vol. 33, no. 3, pp. 264–269, Mar. 2011.
[5] J.-H. Hu, J.-L. Shang, K.-P. Zhou, Y.-K. Chen, Y.-L. Ning, L. Liu, and M. M. Aliyu, "Hazard degree identification of goafs based on scale effect of structure by RS-TOPSIS method," *J. Central South Univ.*, vol. 22, no. 2, pp. 684–692, Feb. 2015.
[6] C. Wang, Y. Lu, B. Cui, G. Hao, and X. Zhang, "Stability evaluation of old goaf treated with grouting under building load," *Geotech. Geol. Eng.*, vol. 36, no. 4, pp. 2553–2564, Aug. 2018.
[7] Z. An-Bing, L. Xin-Xia, and L. Hui, "Goaf surface stability analysis using EMD method," in *Proc. 3rd Int. Conf. Inf. Comput.*, Wuxi, China, 2010, pp. 237–240.
[8] C. Xiao, H. Zheng, X. Hou, and X. Zhang, "A stability study of goaf based on mechanical properties degradation of rock caused by rheological and disturbing loads," *Int. J. Mining Sci. Technol.*, vol. 25, no. 5, pp. 741–747, Sep. 2015.
[9] N. A. Goussies, S. Ubalde, and M. Mejail, "Transfer learning decision forests for gesture recognition," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 3667–3690, Jan. 2017.
[10] N. Patel and J. T. L. Wang, "Semi-supervised prediction of gene regulatory networks using machine learning algorithms," *J. Biosci.*, vol. 40, no. 4, pp. 731–740, Oct. 2015.
[11] Z. Deng, K.-S. Choi, Y. Jiang, and S. Wang, "Generalized hidden-mapping ridge regression, knowledge-leveraged inductive transfer learning for neural networks, fuzzy systems and kernel methods," *IEEE Trans. Cybern.*, vol. 44, no. 12, pp. 2585–2599, Dec. 2014.
[12] J. Lu, V. Behbood, P. Hao, H. Zuo, S. Xue, and G. Zhang, "Transfer learning using computational intelligence: A survey," *Knowl.-Based Syst.*, vol. 80, pp. 14–23, May 2015.
[13] H. Chang, J. Han, C. Zhong, A. M. Snijders, and J.-H. Mao, "Unsupervised transfer learning via multi-scale convolutional sparse coding for biomedical applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 5, pp. 1182–1194, May 2018.
[14] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.
[15] S. Zhou, E. Smirnov, G. Schoenmakers, K. Driessens, and R. Peeters, "Testing exchangeability for transfer decision," *Pattern Recognit. Lett.*, vol. 88, no. 1, pp. 64–71, Mar. 2017.
[16] M. Bereta, "Entropy-based regularization of AdaBoost," *Comput. Assist. Methods Eng. Sci.*, vol. 24, no. 2, pp. 89–100, Dec. 2017.
[17] F. Q. Gong, X. B. Li, L. J. Dong, and X. L. Liu, "Underground goaf risk evaluation based on uncertainty measurement theory," *Chin. J. Rock Mech. Eng.*, vol. 27, no. 2, pp. 323–330, Feb. 2008.
[18] J. Chen, Z.-Q. Luo, and Z.-S. Hou, "Stability evaluation of metal mine goaf based on improved catastrophe progression method," *J. Saf. Sci. Technol.*, vol. 3, no. 11, pp. 17–24, Nov. 2013.
[19] Y. She and A. B. Owen, "Outlier detection using nonconvex penalized regression," *J. Amer. Statist. Assoc.*, vol. 106, no. 494, pp. 626–639, 2011.
[20] L. Beghou, "Methodology for the design of multi-source transmitters dedicated to perpendicular dynamic wireless power transfer: Theoretical study," *Wireless Power Transf.*, vol. 5, no. 1, pp. 54–63, Mar. 2018.
[21] M. D. Robinson and A. A. Oshlack, "A scaling normalization method for differential expression analysis of RNA-seq data," *Genome Biol.*, vol. 11, no. 3, p. R25, Mar. 2010.