WILEY | Hindawi

Research Article

Online Data Migration Model and ID3 Algorithm in Sports Competition Action Data Mining Application

Li Ju ,[1] Lei Huang,[2] and Sang-Bing Tsai [3]

[1]*Department of Physical Education, Guangxi Medical University, Nanning, Guangxi, 530000, China*
[2]*Beijing Sport University, Beijing, 100000, China*
[3]*Regional Green Economy Development Research Center, School of Business, WUYI University, China*

Correspondence should be addressed to Li Ju; juli12342010@sina.com and Sang-Bing Tsai; sangbing@hotmail.com

The ID3 algorithm is a key and important method in existing data mining, and its rules are simple and easy to understand and have high application value. If the decision tree algorithm is applied to the online data migration of sports competition actions, it can grasp the sports competition rules in the relationship between massive data to guide sports competition. This paper analyzes the application performance of the traditional ID3 algorithm in online data migration of sports competition actions; realizes the application steps and data processing process of the traditional ID3 algorithm, including original data collection, original data preprocessing, data preparation, constructing a decision tree, data mining, and making a comprehensive evaluation of the traditional ID3 algorithm; and clarifies the problems of the traditional ID3 algorithm. Mainly, the problems of missing attributes and overfitting are clarified, which provide directions for the subsequent algorithm optimization. Then, this paper proposes a k-nearest neighbor-based ID3 optimization algorithm, which selects values similar to k-nearest neighbors to fill in the missing values for the attribute missing problem of the traditional ID3 algorithm. Based on this, the improved algorithm is applied to the online data migration of sports competition actions, and the application effect is evaluated. The results show that the performance of the k-nearest neighbor-based ID3 optimization algorithm is significantly improved, and it can also solve the overfitting problem existing in the traditional ID3 algorithm. For the overall classification problem of six types of samples of travel patterns, the experimental data samples have the characteristics of high data quality, a considerable number of samples, and obvious sample differentiation. Therefore, this paper also uses the deep factorization machine algorithm based on deep learning to classify the six classes of travel patterns of sports competition action data using the previously extracted relevant features. The research in this paper provides a more accurate method and a higher-performance online data migration model for sports competition action data mining.

1. Introduction

Sports competition is a competitive recreational activity based on certain sports rules and is a very important form of sports activity. The viewing of sports competitions plays a unique role in meeting the exercise needs of people. Since their inception, sports competitions have attracted many participants and spectators with their unique charm [1]. However, when people study sports competitions, they tend to pay more attention to the technical factors and how to promote the level of competition, but few people study sports competitions from the aesthetic level, to explore the deep-seated reasons why sports competitions are so attractive to people, and to develop a certain ideal space for sports competitions. In recent years, clustering technology has been developing, both in terms of clustering technology and clustering applications, and academics have made multifaceted research on clustering algorithms; recommendation algorithms have been applied to the Internet in recent years, and the increasingly diverse society has led to the rapid development of recommendation algorithms [2]. Applying data mining-related methods to sports competition action data mining can solve the situation in that the number of athletes is huge and the condition of the sport is complex and changeable,

which can effectively solve the problems faced by sports competition [3].

By comparing the previously learned knowledge with the new knowledge, we can find out whether there is a similarity between the two, and it is this similarity that forms the basis of transfer learning [4]. At present, it seems that such similarity exists in the vast majority of data, which allows migration learning to be widely applied to many areas of machine learning. Usually, the training of convolutional neural networks does not start from scratch, because as the size of the dataset increases, the time required to train the model is still long even with good hardware performance [5]. This avoids the problem of a long training time. The use of migration learning can mainly save the cost of model training, improve the efficiency of model training, and also optimize the final results. At the theoretical level, migration learning can be applied to any relevant domain with good results [6]. However, if the similarity between the models to be trained and the pretrained models is not good enough, the final results will not be good or even negative migration will occur, so the similarity between the models is the cornerstone for migration learning [7]. Clustering is a machine learning technique that involves the grouping of data points. Given a set of data points, we can use a clustering algorithm to divide each data point into a specific group. Theoretically, data points in the same group should have similar attributes and/or features, while data points in different groups should have highly different attributes and/or features. Clustering is an unsupervised learning method and is a common statistical data analysis technique used in many fields. In data science, we can use clustering analysis to gain some valuable insights from our data.

Currently, the amount of research on decision tree algorithms is increasing, and the focus of research is mostly on the improvement and optimization of decision tree algorithms to improve the classification accuracy of decision tree algorithms, the application effect of decision tree algorithms, and the improvement of the decision tree pruning process to improve the comprehensive application effect of decision tree algorithms. The machine learning method is a method in which a computer uses its data to derive a certain model and uses these models to predict the future [8]. This process is like the human learning process, except that machines can analyze large dimensional data and are tireless. The study is described as follows. Chapter one is an introduction, which first analyzes the background and significance of studying the application of decision tree techniques and transfer learning in sports competition action data mining, and provides an overview of the research content. Chapter two provides an in-depth analysis of the collected domestic and international research data and grasps the status of domestic and international research. Chapter three introduces the data processing and migration models, as well as the design study of data mining classifiers; evaluates the application effect of the optimized ID3 algorithm in sports competition action data mining; and proposes an application of sports competition action data mining based on the optimized decision tree algorithm according to the evaluation results. Chapter four analyzes the research results of this

paper and evaluates the overall effect of sports competition action data mining application. Chapter five is the conclusion and outlook, which summarizes the conclusions obtained by conducting this study and proposes further research directions and priorities for the shortcomings in the research conclusions.

2. Related Work

Clustering is an unsupervised machine learning approach that aggregates data items, observations, or feature vectors into groups. Currently, clustering techniques have been developed significantly, and as an important branch of data mining, clustering techniques have diverse applications in various fields [9]. Karmani et al. proposed a maximum edge clustering (MMC) method, which is based on the support vector machine model. The results of the study proved that the clustering results are better than the K-Means algorithm [10]. In web data mining, structured text processing with structural and semantic coincidences is a challenge in the field of data mining, and processing XML document data is a challenge. Ma and Tsai used the tree meta-ancestor approach to identify XML document data and use clustering methods to process these data and make experiments accordingly [11]. When the number of athletes increases, it is not possible to effectively personalize the decision based on individual circumstances, so the efficiency is low and the response strategy and results are not satisfactory. In terms of data mining, we can automate the modeling process and use a series of techniques such as clustering and recommendation algorithms to help experts make recommendations and give coping strategies for different aspects of athletes, which is rarely done in sports competitions [12].

With the in-depth study of sports competition, the study of the ideas and theories of sports competition has penetrated all fields of sports. Gao et al. analyzed basketball comprehensively from two perspectives, the core level and the auxiliary level, which involved many aspects of basketball such as the body, technique, tactics, stars, confrontation, spirit, style, and the beauty of form, costume, and field equipment, which have positive significance for the study of the characteristics of basketball sports competition [13]. When the number of layers of neural networks increases, the learning ability of neural networks does not improve but, on the contrary, may become worse, which was later proven to be due to the disappearance of gradients [14]. Xie and Ma proposed the SVM algorithm, which became the mainstream of machine learning algorithms at that time because of its excellent migration performance that demonstrated greater advantages over neural networks in many problems [15]. Fujin et al. proposed the convolutional neural network (CNN), which is the first deep learning algorithm with the practical significance of a multilayer network structure [16]. It uses the spatial relativity of data and a human visual neural structure to reduce the number of network parameters, thus effectively improving the performance of the model [17]. A good data preprocessing method not only eliminates structural defects in the existing data but also prepares the data for data mining. The period of archery movements among individuals is not equal, and there are differences in the

timing of applying movement techniques at each stage. Equal-width discretization is suitable for handling complex data with chaotic data structure and strong data continuity [18].

Data mining algorithms such as association rules, clustering, and Markov-based data mining algorithms have been intensively studied and applied in sports competition action data mining to achieve the set research objectives. Several specialized data mining tools have been developed for different domains. The diversity of data mining tasks determines that data mining faces many challenging topics [19]. For data mining researchers, designing data mining languages, adopting efficient and useful data mining methods, and developing systems to build interactive and integrated data-mining environments are the main issues. Processing the collected kinematic data of sports competitions into a standard data structure suitable for use in data mining techniques is a major challenge. Since the speed of different athletes' sports movements varies, the preprocessing of the data is necessary to ensure the integrity of the data and the structure of the data. How to adapt kinematic data to data mining techniques requires a lot of experimentation and detailed analysis. We analyze the strengths and weaknesses of the models studied in the relevant references and use them to determine our own research model.

3. Online Data Migration Study of Sports Competition Action Based on ID3 Algorithm

3.1. Data Processing. A suitable data model is built based on the dataset and the meaning it is intended to convey. The data being analyzed is the training set, and each attribute corresponds to a class label. The samples are randomly selected from the training subset samples divided by the training dataset, and the data are classified under the guidance of the class label number, which is often referred to as guided learning. The concept of guided learning is that this data analysis knows exactly which class is guiding the training samples, as opposed to not knowing which class is guiding the training samples, unguided learning, or what is called clustering. Data analysis is performed on the decision tree that has been constructed. There are many ways to perform data analysis on decision trees; generally, the accuracy of the decision tree is analyzed first, because if a decision tree cannot guarantee its accuracy, the other data analysis loses its meaning. Therefore, the accuracy rate indicates the proportion of data that can be correctly classified by the decision tree classification algorithm and is the most important measure. We generally use the known information rules as the standard and compare the results inferred from the model constructed with the decision tree with the standard [20]. If the comparison results are very different, the accuracy rate is lower, and if the difference is smaller, the accuracy rate is higher.

Data mining, also known as database file expertise discovery (KDD), is the noncommon process of obtaining reasonable, novel, potentially effective, and finally understandable ways from a lot of data. Preprocessing is an important component of the association analysis of datasets. Noise elimination,

data synthesis, and data standard unification are performed on the dataset using preprocessing to facilitate the subsequent data management analysis. In general, the specific requirements of preprocessing are the following: first, the smoothing of the processed dataset, mainly to eliminate the noise in the data, using box division, clustering, and query techniques; second, the synthesis of the dataset, using a concept hierarchy, replacing the low-level "raw" data with high-level concepts; and third, ensuring the standardization of the data—scaling attribute data into a small specific interval [21].

Let the set of datasets be D and the data samples in the dataset D bed. Assume that the dataset D has m different class attributes with different values, which are labeled as B_i $(i = 1, 2 \cdots n)$. Therefore, the amount of classification information can be expressed by equation (1), where the weight of class attribute B_i is expressed by b_i, which can be calculated by b_i/d. The logarithmic function with a base of 10 is used here because the information is encoded in binary.

$$f(D) = \sum_{i=0}^{n} b_i \ln b_i. \tag{1}$$

Suppose that one of the attributes, here denoted as A, has v different values, which can be denoted as $\{b_1, b_2 \cdots b_m\}$. Thus, the dataset D can be divided into v different subsets by the attribute A, here denoted as $\{D_1, D_2 \cdots D_m\}$, where D_j $(j = 1, 2 \cdots m)$ is denoted as the set of samples with the same value b_j $(j = 1, 2 \cdots n)$ on the attribute A set. Suppose b_{ij} denotes the total number of samples belonging to category B_i in subset D_j. The formula for calculating the information entropy of attribute A is shown in (2), and $f(x)$ denotes the proportion of the number of samples with the value b_j to the total number of samples for attribute A.

$$\begin{cases} f(x) = \dfrac{\sum_i^n b_{ij}}{d}, \\ G(D) = \sum_{j=1}^{m} f(x)_j f\left(\sum_{i=1}^{n} b_{ij}\right). \end{cases} \tag{2}$$

The information gain of attribute A can be calculated from

$$H(A) = G(D) - G(D_A). \tag{3}$$

Deep learning is proposed in response to the shallow learning phenomenon of mechanical learning, rote learning, and knowing what is right but not knowing what is wrong in practice. The "depth" here refers to the depth of learning of students. We do not require teachers to adopt a fixed model or method but rather emphasize that teachers should use appropriate methods to trigger, promote, and enhance students' deep learning. In this sense, deep learning is the opposite of shallow learning, and it is a criticism of the times.

The strategy of the decision tree classification algorithm is to calculate the information gain rate of all the test attributes in the alternative dataset and to use the attribute with

the maximum information gain rate as the current division attribute and finally to complete the construction of the decision tree by iterating the above process. The data were analyzed using divisional clustering, but the results were found to be unsatisfactory and not well differentiated [22]. Therefore, when conducting the reason analysis, the small amount of data and many data dimensions may have an impact on the clustering results, and the improved algorithm was considered according to the characteristics of sports competition action data. The improved K-Means algorithm is applied to the sports competition action data to mine the clustering results. The normalization formula is shown in equation (4), where max and min sub another and J is the maximum and minimum values of the sample data.

$$F(x) = \frac{x - \min\,(G(D))}{\max\,(G(D)) - \min\,(G(D))}. \tag{4}$$

Clustering all attributes of the data makes the clustering results reflect the relationship between metadata more fully; on the other hand, in cluster analysis, the attributes of the tuple inevitably form high attribute clusters with similarity. Due to the attributes of all data, clusters have the characteristics of clustering, and each group of data within its data cluster contains attributes; the formation of attribute clusters on each attribute will inevitably occur simultaneously. In other words, if the number of tuples in the dataset presatisfies the minimum support, then the cluster formed by all attribute clusters becomes the maximum item frequency set of the data cluster [23].

3.2. Online Data Migration Model Construction. To address the problem of changing the cost of online data, this paper proposes an adaptive cost-sensitive online migration learning method. First, we introduce the marker distribution into the traditional hinge loss function to calculate the classification cost adaptively; second, we combine the source and target domains using the combination parameters to realize the online migration from the source to the target domain; finally, we migrate to learn the classification model based on both cost and accuracy [24]. For each sample, we automatically calculate the adaptive cost of that sample based on the ratio of the current sample polarity to the polarity of all samples and then add the adaptive cost to the fusion to the loss function and use it to iteratively update the classification model to obtain the current latest classification model. The online data migration model in this paper is shown in Figure 1.

When the sample m_x is read, the positive and negative ratios in the sample set at the previous m moments will change, which will inevitably cause the final classification model to have a learning bias for some samples. To reduce the impact of this change, the current ratio of positive and negative samples and the ratio of the loss of the current sample to the total loss are multiplied together and denoted as D_m, as in equation (5). Where H_m^+ and H_m^- represent the number of positive and negative samples, respectively, $g(b_m, y_m, f(x_m))$ is the cost hinge loss function with introduced cost and $g(D_m, b_m, y_m, f(x_m))$ is the loss function with introduced marker distribution parameters.

$$Y(m) = \frac{H_m^+}{H_m^-} * \frac{g(b_m, y_m, f(x_m))}{\sum_{i=0}^{m} g(D_m, b_m, y_m, f(x_m))} * 100\%. \tag{5}$$

Combining multiple weak classifiers using weight parameters to obtain a strong classifier is the β common approach. The adaptive cost-based online migration algorithm similarly combines the initial classifier $f(0)$ and the online adaptation function $w(m_x)$ by combining the parameters β to obtain a combined classifier.

$$\text{Classifier} = \beta f(0) + (1 - \beta) w(m_x). \tag{6}$$

In the online migration learning process, we aim to update m_{x-1} to m_x using a suitable algorithm. The classification model is obtained based on both smoothing and minimizing the cost sum. The distance between m_{x-1} and m_x is smoothed using Euclidean distance as in

$$M(x) = \min\left(\frac{(m_x - m_{x-1})^2}{2}\right), \quad x \subseteq [0, m]. \tag{7}$$

The cost changes as the online data changes, so it is necessary to find a way to make the cost change adaptively. The misclassification cost is updated adaptively using the ratio of positive and negative samples, and the updated cost can be used to dynamically adjust the learning of the classifier for different samples. Since the cost of rare samples is usually high, a classifier based on adaptive cost tends to improve the classification accuracy of rare samples, which in turn improves the performance of classification.

The availability and prevalence of large amounts of data and the use of tools to properly extract knowledge information have become very frequent. This fact has changed the traditional data analysis by orienting the data to certain specialized techniques under data science. In short, data science can be considered as a discipline that discovers new and important relationships, patterns, and trends after examining large amounts of data. Thus, data science techniques pursue the automatic discovery of knowledge contained in information stored in large databases.

The structural similarity matrix is used to describe the distribution between a descriptive attribute and a categorical attribute; the number of occurrences of all attribute values of each descriptive attribute under all attribute values of the categorical attribute and the attribute values of the categorical attribute are used to obtain the values with the highest number of overlaps to participate in the next step. The structural similarity matrix model will create a matrix for each described attribute to provide the corresponding parameters for calculating the sample structural similarity.

Since the sample structural similarity of the attributes is calculated based on the structural similarity matrix, we can assume that if the structural similarity matrix is verified to be free of multivalued bias, we can conclude that there is no multivalued bias through the sample structural similarity of the attributes [24]. And if it can be verified that there is no definite size relationship between the structural similarity matrix and the increase of attribute values and since the
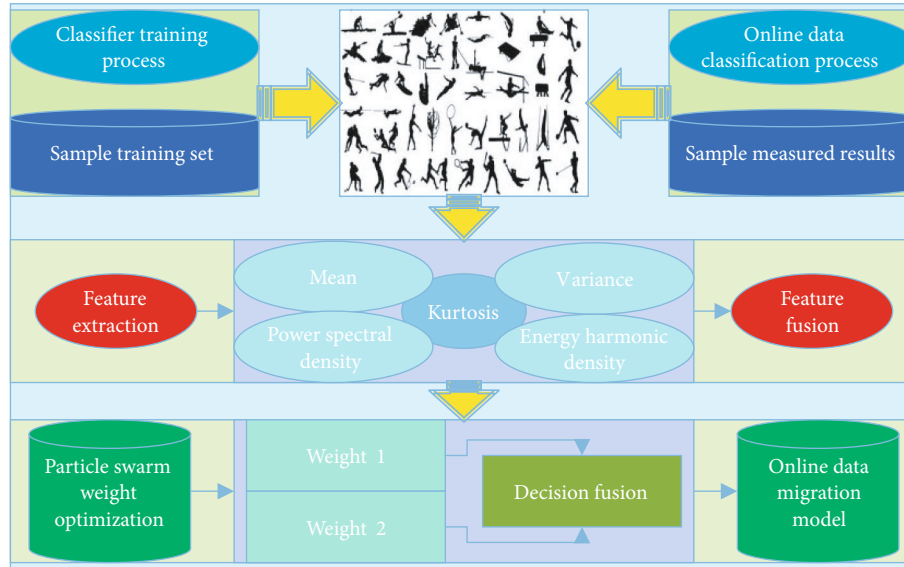
FIGURE 1: Online data migration model.

sample structural similarity generated by the structural similarity matrix is more biased towards the descriptive attributes that are structurally more similar to the categorical attributes, then the sample structural similarity as weights will play a positive role in the correction of information gain, thus reducing the interference of the multivalue bias problem on the selection of the joints.

3.3. Data Mining Classifier Design. In this paper, we propose an automatic extraction strategy for sports competition action data based on image overlap region feature migration machine learning, as shown in Figure 2. The strategy can be divided into three steps: (1) select the classifier with the highest generalization ability among ID3 algorithm classifiers to supervise the data classification and set it as the source classifier model; (2) propagate the source marker to the adjacent target image overlap region, randomly select a certain proportion of its marker samples to mix with the source training samples to form the pseudosample, and use the migration learning model to balance the difference of its temporal distribution; and (3) obtain the final classification model through the source classifier. The final classification model is obtained after continuous iterations, and the distribution difference between adjacent target image data is continuously reduced due to the increasing information of balanced temporal samples, to achieve the purpose of accurate extraction of sports competition action data (see Figure 2).

Using the automatic sports action data extraction strategy, only the source training samples of the source images and the classified images with overlapping areas with the images to be classified are involved in the classification, and then, the temporal difference balance based on the migration learning model and the source classifier model can complete the classification, and the information extraction of each image is completed one by one through the continuous iterative update to realize the large area sports competition data extraction. The action data extraction work is completed

one by one through continuous iterative updates. In this paper, the classifier algorithm with high adaptability and generalization ability is used as the classification model, which can effectively improve the accuracy of image classification results by learning more potential features. The migration learning model, as a tool for adaptive balancing between source and target images, can reduce the differences in data distribution between them due to temporal and spatial differences and has a weakening effect on improving the problem of obvious color differences in the extraction of thematic information covering large regions of sports competition action data. Selecting certain marker samples in the overlapping area together with the source training samples constitutes a pseudosample, which makes it possible to consider the spectral feature distributions of both the source and target images when performing temporal phase difference balancing and avoid the overbalancing situation. The close cooperation of the above methods to complete the large area grass information extraction will help to improve the extraction accuracy.

4. Analysis of Results

4.1. ID3 Algorithm Performance Analysis. Comparing the running time of the traditional ID3 algorithm and the improved ID3 algorithm, the execution time of the improved ID3 algorithm is shorter and the running time is reduced by the improvement, and the comparison results are shown in Figure 3(a). By comparing the classification error rate of the traditional ID3 algorithm and the improved low algorithm, the error rate of the improved ID3 algorithm is lower, and the comparison results are shown in Figure 3(b). By comparing the experiments of both the traditional ID3 algorithm and the improved ID3 algorithm in terms of algorithm execution time and classification error rate, it can be found that the dataset has reduced running time and lowered the
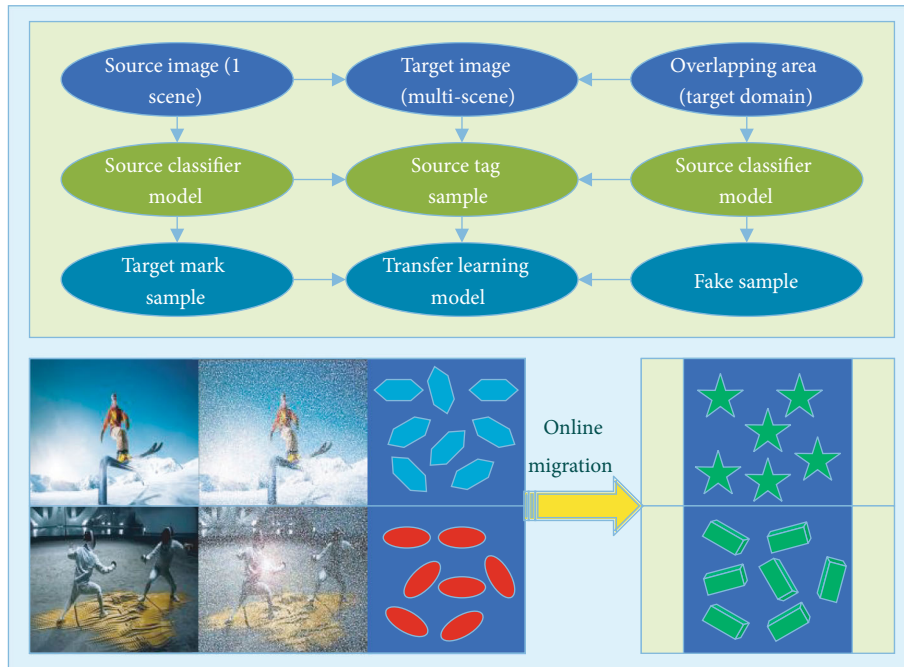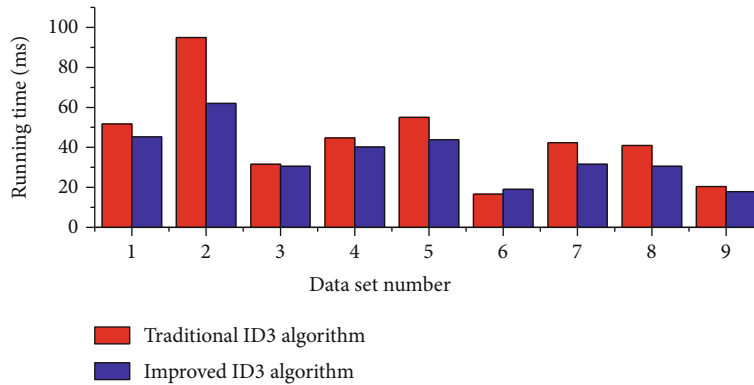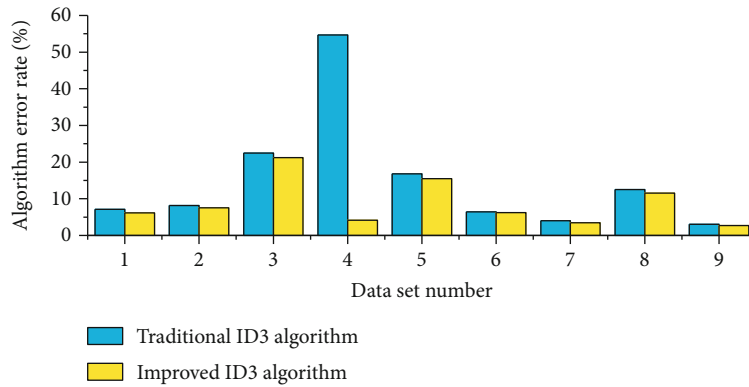
FIGURE 2: Flowchart of automatic extraction strategy for sports competition action data.

(a)

(b)

FIGURE 3: Comparison of the running time and error rate of the algorithm.
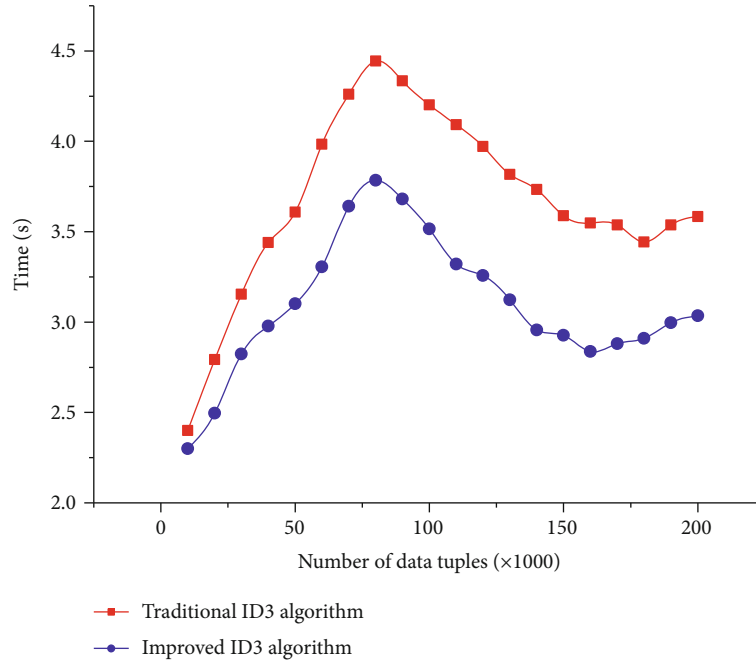
FIGURE 4: Modeling time comparison results.

average classification error rate after using the improved ID3 algorithm (see Figure 3).

Although the improved ID3 algorithm is proportional in time cost to the number of data tuples in the sample dataset, as the number of data tuples increases, this ratio is converging to the time to quantity ratio of the original algorithm. This means that the larger the sample dataset, the closer the decision tree algorithm is to the traditional ID3 algorithm in terms of practical performance. Then, we continue to process the time information to get the results of modeling time comparison between the traditional ID3 algorithm and the improved ID3 algorithm (Figure 4). Figure 4 shows that the time ratio between the improved ID3 algorithm and the traditional ID3 algorithm increases when the number of data tuples in the sample dataset grows from 10,000 to 20,000, while the time ratio between the improved ID3 algorithm and the traditional ID3 algorithm continues to decrease when the number of data tuples grows from 20,000 to 80,000, but at the value of 80,000, the decreasing trend of this ratio slows down at 160,000 items. This trend indicates that the time efficiency of the improved ID3 algorithm decreases when the number of data tuples is small, mainly due to the more complex algorithm, but the proportion of time spent by the algorithm decreases when the number of entries increases steadily. When the number of data tuples in the sample dataset is very large, the main time overhead has shifted to data I/O and processing and is no longer concentrated on the algorithm (see Figure 4).

The sample structural similarity model used by the improved ID3 algorithm does not have a large impact on the classification performance for datasets with an overall low structural similarity. When encountering datasets where the number of attribute values does not differ much and the structural similarity of the descriptive attributes is not more

obviously related to the structural similarity of the classified attributes in terms of structure, the structural similarity has less impact on the generation of the final decision tree structure, thus making the prediction rate of the improved ID3 algorithm more approximate to that of the traditional ID3 algorithm and making the classification accuracy between the two closer.

4.2. Online Data Migration Model Analysis. Figure 5 shows the experimental results of the VLSC dataset, and the following conclusions are obtained from the figure: (1) All migration learning methods perform better than online methods, the results indicate that the effect of training using only the target sample is limited, and using source samples with the same source or the same structure as the target sample is beneficial to the learning of the target sample. (2) The classification effect of kernel-based migration methods is significantly higher than that of online algorithms. Relative to the classical online algorithm, the classification accuracy of the kernel-based migration method is about 10% higher in the six tasks generated by the sets Caltech101 and SUN09 and exceeds 5% in the tasks of the datasets Label ME and VOC2009, which demonstrates the applicability of the kernel-based migration learning method. (3) Compared with the basic kernel-free online migration method, the overall accuracy of KOTL experiments is improved by about 2%, which indicates that the kernel-free online algorithm is suitable for the case of linearly separable source and target domains, while the kernel-based online migration method solves the problem and is also suitable for the case of linearly separable domains. (4) For the offline kernel migration method ARRLS, KOTL is also comparable to it and applicable to the online case. The offline kernel migration method requires a large storage space for caching all samples in both source and target
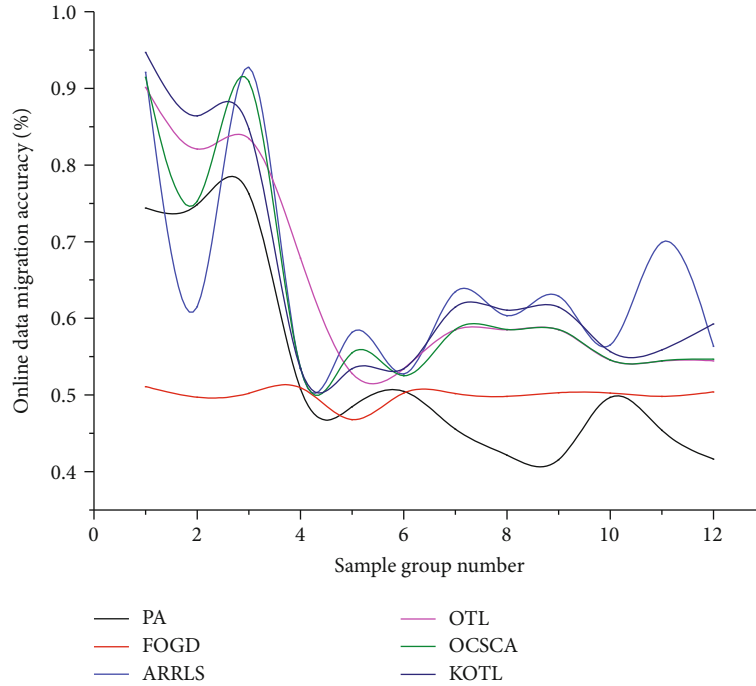
Figure 5: Online data migration accuracy comparison results.

domains, and the computational effort is also related to the number of samples (see Figure 5).

To verify the migration effect of the online data migration model proposed in this paper, as well as the enhancement effect on the target domain data under different sparsity, maintain the consistent time interval between samples, and maintain the temporal order of the data, this experiment selects the first 10%~90% samples from SCADA data as the target domain data and the remaining data as the test set for power prediction, using a long short-term memory network (LSTM) as the regression model for wind power prediction, and root mean square error (RMSE) is used as the measure. The root mean square error (RMSE) is the square root of the ratio of the square of the deviation of the predicted value from the true value to the number of observations n. In practical measurements, the number of observations n is always limited and the true value can only be replaced by the most trustworthy (best) value. The RMSE results of the prediction results on the test set are shown in Figure 6 for the comparison between the online data migration model and the PCTR model and the case where no source domain data are introduced. As the proportion of target domain data decreases and the data sparsity of the source domain training data increases, the gap between the prediction accuracy of the nonmigrated model and the migrated model becomes more obvious. The experimental results show that the online data migration model can effectively solve the problem of sparse target domain training samples and improve the accuracy of the prediction task by introducing source domain data and migrating the source domain knowledge (see Figure 6).

4.3. Analysis of the Effect of Sports Competition Action Data Mining. We have added the validation set test set. For each
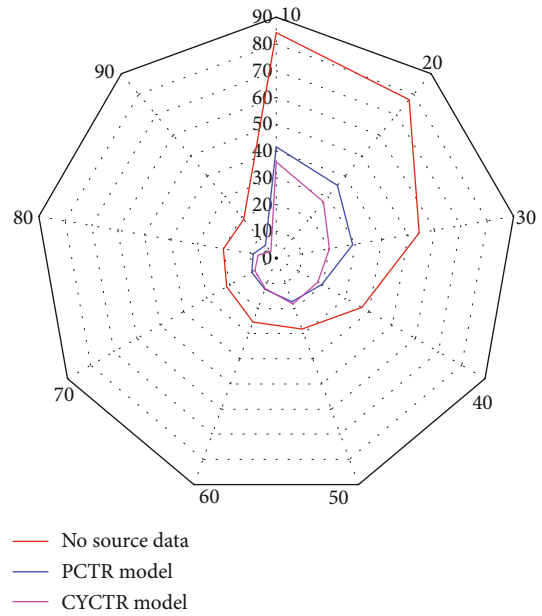

Figure 6: RMSE values of the model on the test set.

competitive sports development situation, input indicators, output indicators, and development indicators are used as indicator selection basis. Among them, the expenditure on competitive sports (including sports competition fees, sports training fees, and sports stadium construction fees) is taken as the input indicator, the points of the sports competition are taken as the output indicator, and the number of outstanding athletes is taken as the development indicator. Figure 7 shows the descriptive statistics about the three initial variables, including the mean, standard deviation, and the
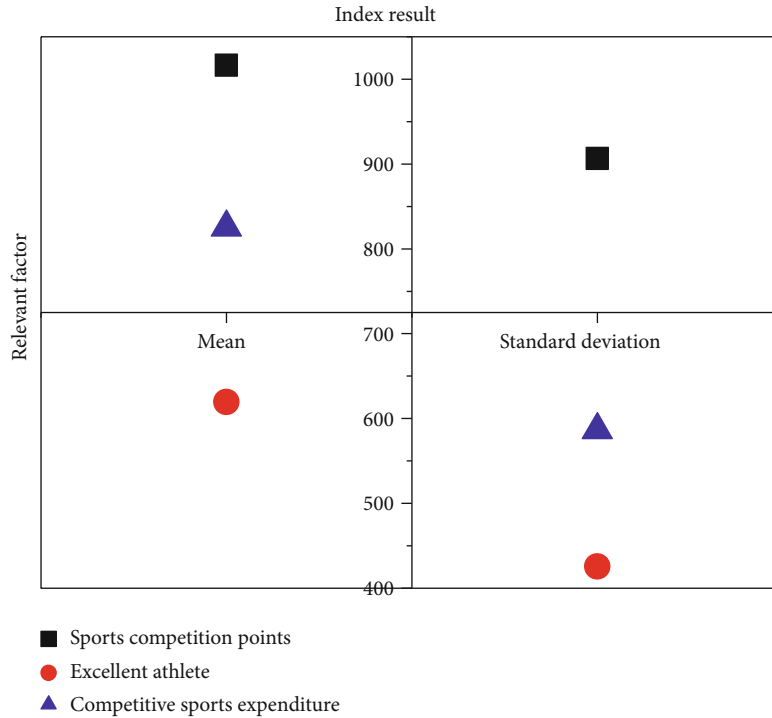
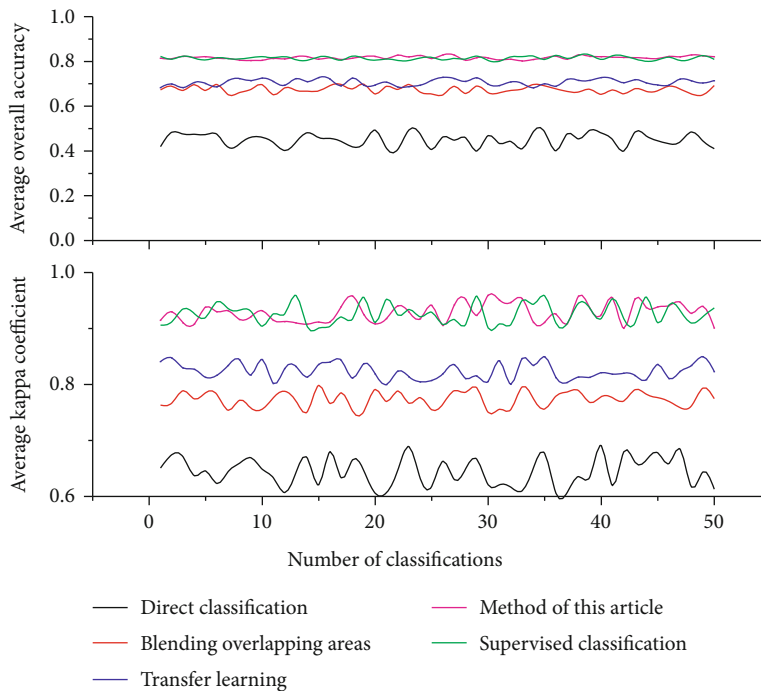FIGURE 7: Descriptive statistics of the initial variables.

FIGURE 8: Comparison of the overall classification accuracy of the five methods.

number of values used in the analysis N. Among them, the standard deviations of the points of sports competitions, excellent athletes, and competitive sports expenditures are 907.56, 426.47, and 586.61, respectively (see Figure 7).

Figure 8 shows the comparison of the average overall accuracy and average kappa coefficient of the target images under the five classification strategies. With the overall accuracy as the evaluation index, the relationship of (a) < (c) < (b) < (d) ≈ (e), and the kappa coefficient as the evaluation index, the relationship of (a) < (b) < (c) < (e) < (d) is satisfied. The overall classification situation still confirms the above conclusion that both fusion of overlapping

region labeled sample information and migration learning can improve the image classification accuracy, while the fusion of overlapping region labeled sample information and migration learning can add up the ability of both to improve the classification accuracy. The average overall accuracy obtained by this method is slightly equal to that of the supervised classification, the average kappa coefficient is greater than that of the supervised classification, and the two classification accuracy evaluation metrics are given the same weight, which indicates that this method is slightly better than the supervised classification method (see Figure 8).

When performing data mining algorithms, the sports competition dataset is used as the research object for the online migration success rate of sports competition action data, so the decision tree algorithm is used to analyze the main attributes in the sports competition data source that may affect the check-in rate and to identify the factors that are most likely to affect the sports competition movement so that these potentially relevant influencing factors can be scientifically applied to provide key future sports competition decision basis. The advantages of the decision tree algorithm are simple and easy to understand analysis, high classification accuracy, and high execution efficiency, so the decision tree algorithm is suitable to be applied to massive data mining. The ID3 algorithm selects the classification criterion of information gain, it selects the attribute with more attribute values as the split attribute, and the ID3 algorithm can only mine the data with nonlinearity.

Data classification is an important part of data mining, and the decision tree algorithm is one of them. It is an important method in data classification methods and has a mature theoretical foundation and a good development platform. The decision tree algorithm is relatively easy to understand, and the construction process is fast. What is more, decision trees can be easily converted into SQL statements for efficient access to the large number of various databases existing in financial systems. Data mining technology is an important technical tool to transform data into knowledge and value, but traditional data mining technology faces many challenges to extract the rich knowledge and value implied from big data. An important way to solve the problem of big data mining is to research and develop more efficient data mining algorithms based on the essential characteristics of big data.

5. Conclusion

In this paper, the clustering algorithm in data mining is applied to the sports competition action data migration model; using the theory related to the ID3 clustering algorithm and combining the characteristics of sports competition action data, the ID3 algorithm is combined and the ID3 algorithm is slightly improved to make it more suitable for sports competition action number nuggets. Then, the data of each dimension are clustered to derive the corresponding results, and the results are synthesized and analyzed to facilitate the subsequent recommendation algorithm for online data migration modeling. This paper presents a detailed analysis of sports competition action data and related theories and successfully applies the theoretical

knowledge of computer aspects to sports competition action data mining, paving the way for future cross-application of sports competition action analysis and computer disciplines. In this paper, only the ID3 algorithm in the data mining clustering algorithm has been fully studied and other clustering algorithms are not involved, and the algorithm based on division requires more data types and shapes of datasets and is also easy to fall into the local optimum; whether the sports competition pressure data in this paper is coincidentally applicable to other algorithms is not studied in-depth, and other types of clustering algorithms can be used for the data in this paper subsequently. It is possible to reach better conclusions. The clustering results have some errors and are only provided as a reference to the also-ran scholars for reference and cannot replace the actual also-ran judgments. Similarly, the recommendation results only assist the scholars to make suggestions and cannot be provided to the athletes as actual suggestions, but this, as a practical nature of the exploration, still needs continued in-depth research in the follow-up work. We provide an in-depth analysis of the online data migration model and the ID3 algorithm in sports game action data mining, and the model under study has a strong practicality.

Data Availability

All information is within the paper.

Conflicts of Interest

No competing interests exist concerning this study.

References

[1] S. Chen, "An effective going concern prediction model for the sustainability of enterprises and capital market development," *Applied Economics*, vol. 51, no. 31, pp. 3376–3388, 2019.

[2] B. V. Chowdary and Y. Radhika, "A survey on applications of data mining techniques," *International Journal of Applied Engineering Research*, vol. 13, no. 7, pp. 5384–5392, 2018.

[3] J. M. Torres, C. I. Comesaña, and P. J. Garcia-Nieto, "Review: machine learning techniques applied to cybersecurity," *International Journal of Machine Learning and Cybernetics*, vol. 10, no. 10, pp. 2823–2836, 2019.

[4] S. Shafqat, S. Kishwer, R. U. Rasool, J. Qadir, T. Amjad, and H. F. Ahmad, "Big data analytics enhanced healthcare systems: a review," *The Journal of Supercomputing*, vol. 76, no. 3, pp. 1754–1799, 2020.

[5] D. Draskovic, M. Cvetanovic, and B. Nikolic, "SAIL—software system for learning AI algorithms," *Computer Applications in Engineering Education*, vol. 26, no. 5, pp. 1195–1216, 2018.

[6] C. Schmidt and W. N. Sun, "Synthesizing agile and knowledge discovery: case study results," *Journal of Computer Information Systems*, vol. 58, no. 2, pp. 142–150, 2018.

[7] W. Liu, H. L. Ma, and A. Walsh, "Advance in photonic crystal solar cells," *Renewable and Sustainable Energy Reviews*, vol. 116, article 109436, 2019.

[8] X. Zhang, C. Zang, H. L. Ma, and Z. J. Wang, "Study on removing calcium carbonate plug from near wellbore by high-power

ultrasonic treatment," *Ultrasonics Sonochemistry*, vol. 62, article 104515, 2020.

[9] H. L. Ma, X. Zhang, F. F. Ju, and S. B. Tsai, "A study on curing kinetics of nano-phase modified epoxy resin," *Scientific Reports*, vol. 8, no. 1, article 3045, 2018.

[10] M. Ling, M. J. Esfahani, H. Akbari, and A. Foroughi, "Effects of residence time and heating rate on gasification of petroleum residue," *Petroleum Science and Technology*, vol. 34, no. 22, pp. 1837–1840, 2016.

[11] H. L. Ma and S. B. Tsai, "Design of research on performance of a new iridium coordination compound for the detection of Hg2+," *International Journal of Environmental Research and Public Health*, vol. 14, no. 10, article 1232, 2017.

[12] L. Y. Mo, W. H. Z. Sun, S. Jiang et al., "Removal of colloidal precipitation plugging with high-power ultrasound," *Ultrasonics Sonochemistry*, vol. 69, article 105259, 2020.

[13] D. Gao, Y. Liu, Z. Guo et al., "A study on optimization of CBM water drainage by well-test deconvolution in the early development stage," *Water*, vol. 10, no. 7, p. 929, 2018.

[14] S. B. Tsai and H. Ma, "A research on preparation and application of the monolithic catalyst with interconnecting pore structure," *Scientific Reports*, vol. 8, no. 1, article 16605, 2018.

[15] J. Xie and H. Ma, "Application of improved APO algorithm in vulnerability assessment and reconstruction of microgrid," *IOP Conference Series: Earth and Environmental Science*, vol. 108, no. 5, article 052109, 2018.

[16] A. Fujin, Y. Xiuzhao, and H. Ruochi, "Research into the super-absorbent polymers on agricultural water," *Water Management*, vol. 245, article 106513, 2021.

[17] Y. Shen, B. Biondi, and R. Clapp, "Q-model building using one-way wave-equation migration Q analysis — part 2: 3D field-data test," *Geophysics*, vol. 83, no. 2, pp. S111–S126, 2018.

[18] R. D. Hume, L. Berry, S. Reichelt et al., "An engineered human adipose/collagen model for in vitro breast cancer cell migration studies," *Tissue Engineering Part A*, vol. 24, no. 17-18, pp. 1309–1319, 2018.

[19] M. Patel, S. Chaudhary, and S. Garg, "Improved pre-copy algorithm using statistical prediction and compression model for efficient live memory migration," *International Journal of High Performance Computing and Networking*, vol. 11, no. 1, pp. 55–65, 2018.

[20] R. W. Allen, J. S. Collier, A. G. Stewart et al., "The role of arc migration in the development of the lesser Antilles: a new tectonic model for the Cenozoic evolution of the eastern Caribbean," *Geology*, vol. 47, no. 9, pp. 891–895, 2019.

[21] D. Xu and H. Ma, "Degradation of rhodamine B in water by ultrasound-assisted TiO_2 photocatalysis," *Journal of Cleaner Production*, vol. 313, article 127758, 2021.

[22] J. Hey, Y. Chung, A. Sethuraman et al., "Phylogeny estimation by integration over isolation with migration models," *Molecular Biology and Evolution*, vol. 35, no. 11, pp. 2805–2818, 2018.

[23] S. E. Vollset, E. Goren, C. W. Yuan et al., "Fertility, mortality, migration, and population scenarios for 195 countries and territories from 2017 to 2100: a forecasting analysis for the Global Burden of Disease Study," *The Lancet*, vol. 396, no. 10258, pp. 1285–1306, 2020.

[24] S. Zhang, A. Lorenzo, C. Zhou, Y. Cui, B. Gonçalves, and M. Angel Gómez, "Performance profiles and opposition interaction during game-play in elite basketball: evidences from National Basketball Association," *International Journal of Performance Analysis in Sport*, vol. 19, no. 1, pp. 28–48, 2019.