Hindawi
Mobile Information Systems
Volume 2021, Article ID 1875060, 11 pages
https://doi.org/10.1155/2021/1875060

Research Article

Prediction Model of Football World Cup Championship Based on Machine Learning and Mobile Algorithm

Yanyang Bai[1] and Xuesheng Zhang[2]

[1]*Sports Department, Guizhou University of Finance and Economics, Guiyang 550025, Guizhou, China*
[2]*Sports Department, North China Institute of Science and Technology, Sanhe 065201, Hebei, China*

Correspondence should be addressed to Xuesheng Zhang; zhangxuesheng@ncist.edu.cn

Received 21 May 2021; Revised 10 June 2021; Accepted 25 June 2021; Published 13 September 2021

Academic Editor: Sang-Bing Tsai

With the technological development and change of the times in the current era, with the rapid development of science and technology and information technology, there is a gradual replacement in the traditional way of cognition. Effective data analysis is of great help to all societies, thereby drive the development of better interests. How to expand the development of the overall information resources in the process of utilization, establish a mathematical analysis–oriented evidence theory system model, improve the effective utilization of the machine, and achieve the goal of comprehensively predicting the target behavior? The main goal of this article is to use machine learning technology; this article defines the main prediction model by python programming language, analyzes and forecasts the data of previous World Cup, and establishes the analysis and prediction model of football field by K-mean and DPC clustering algorithm. Python programming is used to implement the algorithm. The data of the previous World Cup football matches are selected, and the built model is used for the predictive analysis on the Python platform; the calculation method based on the DPC-K-means algorithm is used to determine the accuracy and probability of the variables through the calculation results, which develops results in specific competitions. Research shows how the machine wins and learns the efficiency of the production process, and the machine learning process, the reliability, and accuracy of the prediction results are improved by more than 55%, which proves that mobile algorithm technology has a high level of predictive analysis on the World Cup football stadium.

1. Introduction

The foreign preliminary matches of the World Cup champion team started early. Researchers also used some data analysis techniques and models and studied the data of previous football World Cup matches to get the trend of the football World Cup champions. For the shortcomings and deficiencies of the K-means algorithm, domestic and foreign scholars have given different improved algorithms according to different situations. They proposed an improved K-means algorithm for adaptive feature weights based on dynamic feature attribute weights. This method can reduce the dependence of the K-means algorithm on the initial center and increase the group effect.

Zhu hopes that the algorithm of the machine is a new algorithm that is applicable in almost any field. He came up with the application scheme of the algorithm in data acquisition. The output of a large number of mobile storage data and the effective positioning system of the storage system are the three stages of the cell tracking system, and then, use statistical methods to compare the results with a broad linear model. In combination, a quantitative analysis of the running characteristics of 381 nongoalkeeper football players in the 829 full-time matches during the 2018 World Cup in Russia was carried out to greatly improve the positioning speed and accuracy, but this method requires Consume a lot of resources, and it is not recommended to be widely used [1]. Lu et al. believes that traditional observation

is mainly for a single point, and the observation mechanism is used to predict the probability distribution within a small sample range and the idea of generating virtual data. This idea uses the relationship prediction model of the number of data three times to predict the sample through the observation mechanism probability distribution and generation of virtual data. He used virtual data. In order to support scientific data, he only created a new model of observation to solve problems, but these problems were only investigated in a small sample, and the written algorithm mechanism could not achieve better or higher the accuracy [2]. Li took some data, videos, statistics, and comparative analyses as examples to compare the scores and input characteristics of men on the eve of the World Championship finals in the 17th and 21st centuries. It has dropped to low level in the 19th World Championship. The clues we found were as follows: At the beginning of the World Cup, the number of teams exceeded the number of retired teams; the second zone is the best candidate for the game; 15 minutes before the end is the one with the highest scoring rate period. However, there are many ways to score goals, and the uncertainty and randomness on the court have led to prediction results [3].

In recent years, due to the needs of work and research in reality, more and more attention has been paid to the research and application of machine learning at home and abroad. Machine learning can continuously learn and optimize based on new data during the application process and improve the prediction model. The machine learning is applied to the prediction of the football championship, and the important information hidden in the data is excavated from the historical data of the World Cup football game. This can not only provide theoretical support for the fans' prediction research on the championship but also provide decision support for the company's leadership. Clustering algorithm is an important noncensored learning method for learning machines. The most extensive is the study of aggregation algorithms, which are based on mutually independent measurements. For example, methods such as K-means, K-center point, and the like can predict the desired results in a more detailed and accurate manner and are more efficient and faster than traditional methods.

2. Basic Method of Machine Adaptability

Machine learning is a combination of multiple technologies and comprehensive technologies, it mainly includes arithmetic statistics, mathematical analysis and other mathematical calculations, algorithm design, and other disciplines [4, 5]. Use computer-aided technology to realize artificial intelligence "learning."

This section mainly introduces commonly used machine learning algorithms, including regression, clustering, association, classification, and the like [6, 7]. These methods are widely used in the field of data mining.

(1) Regression model provides a predictable model, mainly focusing on the indefinite relationship between active variables and independent variables and establishing a model based on the relationship between the two variables. It is the most important restoration algorithm in robot learning: linear regression, logistic regression, ridge regression, etc. [8, 9].

(2) Classification is classified by data and by index, so existing data can be used for education and learning according to the existing classification, so as to identify and classify the difference between data and classify data. Commonly used classification algorithms include decision tree classification, naive Bayes classification algorithm [10, 11]. The most common method is to make a staged decision through a decision tree. Decision trees are classified and trained by categories, then the prediction is made according to the unknown criteria. ID3 is the basic method of decision tree, C4.5 (C5.0), CART, PUBLIC, are all improved algorithms on the basis of ID3 algorithm [12, 13]. The naive Bayes classification algorithm is an algorithm that uses the knowledge of probability and statistics for classification. It provides assumptions based on statistical probability and independence. In many cases, it is not realistic. Therefore, the segmentation rules of the algorithm lead to incorrect result. Therefore, the TAN (Tree Augmented Naive Bayes) algorithm was proposed. TAN can reduce the independent assumptions and improve the Bayesian network structure by increasing the relationship between attributes [14–16].

(3) Relevance system analysis is based on data collection to describe the relevant meaning and relevance of charts and patterns describing a certain performance [17]. The analyses of beer and diapers and adult and diapers is our typical example. Wal-Mart analytics recorded the surprising similarity between beer and adult diapers, so he put beer and diapers on the margins. Later investigations found that women in the United States usually take care of their children at home, so they often ask their husbands to remember to buy diapers after get off work, and their husbands will buy their favorite beer when buying diapers in the supermarket [18].

(4) Cluster (Cluster) Clustering is a process of bringing together things with the same nature according to a certain method, after combining similar objects in a certain way [19]. The input group is a sample, but an unidentified group. In grouping, the data are distinguished by characteristics [20]. The principle of grouping in grouping is that the more similar the group, the more different. Common clustering methods include fuzzy clustering, hierarchical clustering, and density clustering [21].

3. Application Research Experiment of Machine Learning

3.1. Introduction to Machine Vision. Machine vision is measured and evaluated by machinery [22]. The machine vision system means recording (camera to capture the target

and seeing) the target image of a specific group of targets and select to send such a message to get the shape of the object, in the signal allocation and its brightness, color, etc.; image transmission will generate a series of functions to track the target and adjust the visible objects according to the results just now. "The machine looks like a computer (maybe it is mobile) because it simulates human vision, and the simulation is the ultimate goal of the simulation. The actual processing of the image is mainly to process the image: zoom in, repeat, reducing the volume, edit, etc., as common photoshop is a powerful image processing software [23, 24]. Most machine vision includes the process of image processing [11]. Only after image process, can you find the required features in the image, so as to further execute other instructions. Some image algorithms studied in our actual engineering applications are actually machine vision, not pure image processing. In general, image processing technologies include image compression, improvement and recovery, matching, description, and identification; Parts are not independent in the field of practical technology, they are often interrelated and help each other to achieve practical results. Next, I will briefly introduce the image processing algorithms commonly used in machine vision.

3.2. Image Filtering Processing. Filtering is generally used in the image preprocessing stage to improve image information and facilitate subsequent processing [25]. Of course, this is not absolute. Filtering operations can be performed at any time if necessary in the image algorithm process. There are three commonly used filtering methods:

Mean filtering is also called linear filtering, and its main method is neighborhood averaging [26]. The basic principle of linear filtering is to replace each pixel value in the original image with the mean value, that is, the current pixel point (x, y) to be processed, select a template, which is composed of several pixels of its neighbors, and find the value of all pixels in the template. Average value and then assign the average value to the current pixel point (x, y) as the dark value $g(x, y)$ of the image at the processed point, namely,

$$g(x, y) = \frac{1}{m} \sum f(x, y). \tag{1}$$

M represents a complete pixel. Dragging the current pixel in the template can make the image smooth and fast and calculate the filter but cannot remove the sound noise.

(1) Median filtering is a nonlinear smoothing technique [27], which sets the gray value of each pixel to the median of the gray values of all pixels in a certain neighborhood window of that point. The realization process is

(1) In the window of the image, remove and sort the odd data.
(2) Use the sorted median value as the gray value of the current pixel.

The intermediate filter used to protect the edges is a classic method of image processing, which can capture sound, and the fringe phase analysis is of great benefit.

(2) Gaussian filtering is a linear smoothing filter [28], suitable for filtering Gaussian white noise, and has been widely used in the preprocessing stage of image processing. Gaussian filtering of the image is to calculate the pixel value of each point in the image. The calculation criterion is that the gray value of the point itself and the gray value of other pixels in the neighborhood are weighted and averaged to get the weighted average weight coefficient. It is sampled by a two-dimensional discrete Gaussian function and normalized. Discrete Gaussian convolution kernel dimension H: $(2K + 1) \times (2K + 1)$, its element calculation method is:

$$H_{i,j} = \frac{1}{2\pi\sigma^2} e^{(i-k-1)^2+(j-k-1)^2/2\sigma^2}. \tag{2}$$

Among them is the variance, k determines the dimension of the kernel matrix [29].

3.3. Image Segmentation. Digital sharing is a process to a process to segment images into specific areas and use them to remove targets. This is the most important step in copying photographs, and most of the existing technologies have passed. It is mainly divided into the following categories: threshold-based segmentation methods, region-based segmentation methods, edge-based segmentation methods, and specific theory–based segmentation methods. The target extracted after image segmentation can be used in image semantic recognition, image search, and other fields [30].

As shown in Figure 1, image segmentation is the technique and process of dividing an image into several specific areas with unique properties and proposing corresponding targets. The k-means algorithm is a common clustering algorithm [31]. In order to perform the calculation of the embedding space, the pixels in the image space will be separated by the symbols corresponding to their existence in a given space, and then, the image space will be inserted into the original image space to achieve the split of the image. The most commonly used threshold segmentation methods are the maximum between-class variance method (OTSU), the minimum error method, and the maximum entropy method. Among them, the OSTU algorithm is the most used.

The maximum between-class variance method OTSU algorithm, also known as the sum algorithm, is a binarization method that automatically selects a threshold based on the principle of decision analysis and least squares [32]. The basic idea is to use a certain gray for the image histogram. The degree value is divided into two, and the two most favorable two groups are extracted. Now, the threshold value is processed to two constants, which will be a more satisfactory result.

Grayscale $f(x, y)$ and y 0: L, the grayscale number is equal to the total number of pixels in the picture:

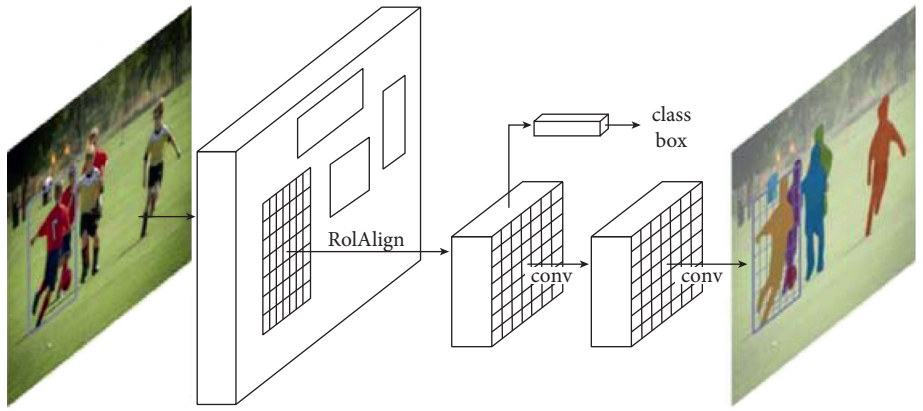

Figure 1: Schematic diagram of *K*-means algorithm after image segmentation (from https://blog.csdn.net//article/details/84641951).

$$N = \sum_{i=0}^{L} n_i. \tag{3}$$

When the threshold of k is established, the ashing stage will be divided into two groups C_0 and C_1, representing the background and the goal, respectively: $C_0 = 0$: K, then, $C_1 = K + 1$, then:

$$P_I = \frac{n_i}{N}. \tag{4}$$

Different thresholds for different types of exports are a small threshold that is automatically calculated, and the optimal threshold has been reached:

$$T_h = \max\left(\frac{\sigma^2}{2}\right)k. \tag{5}$$

The standard deviation represents the consistency of the numbers. When the difference between the two parts of the image is larger, if the difference of an object or another target occurs in the form of segments, the difference between the different parts is smaller. Therefore, the biggest difference means the smallest error [33].

3.4. Feature Extraction. Histogram of Oriented Gradient (HOG) feature is a feature descriptor used for object detection in computer vision and image processing. It composes features by calculating and counting the histogram of the gradient direction of the local area of the image. The HOG feature combined with the SVM classifier identification is widely used; especially, in the passport receiving aspect, the effect is quite good. The method of HOG + SVM was proposed by French researcher Dalal in 2005. Although there are many other technical methods, they are basically based on the idea of HOG + SVM [34]. The extraction process of HOG features is as follows:

In order to reduce the influence of lighting factors, the entire image needs to be normalized (normalized) first. In the texture intensity of the image, the local surface exposure contributes a larger proportion. This compression technique can effectively reduce the light and dark changes. Because the color is lowered, it has no effect on us and generally makes ourselves grayscale. The Gamma compression formula is given as:

$$I(x, y) = I(x, y)^{\text{gamma}}. \tag{6}$$

Gamma = 1/2 can be extracted, which can be achieved by calculating the right inclination angle of the image, the diagonal line of the subcoordinate, and the balance inclination. The gradient of the pixel (x, y) in the image is given as

$$G_x(x, y) = I(x + 1, y) - I(x - 1, y). \tag{7}$$

Among them, $I(x+1, y)$, $I(x-1, y)$, and $G_x(x, y)$, respectively, represent the horizontal gradient, vertical gradient, and pixel value at the pixel point (x, y) in the input image. The gradient magnitude and gradient direction at pixel (x, y) are

$$G(x, y) = \sqrt{G_x(x, y)^2 + G_y(x, y)^2}. \tag{8}$$

3.5. Traditional Modeling Analysis. Since the number of goals scored in football follows the Poisson model [35], we establish the following model:

$$\frac{X_{ij}}{Y_{ij}} = P\left(\theta_{ij}\right). \tag{9}$$

The algorithm calculates the four evolutionary stages of each particle: exploration development, convergence, and exit. It can calculate the average distance between the particle and all particles. In each generation, if we compare the distance measured by the average distance of the previous generation with the octave particle measurement to compare the average distance between the two values, a slight damage will cause the group's evolutionary status to be "restricted." If the former is relatively small, the group's evolutionary process stops: when the comparison result shows that the two values are between the two values, the evolutionary status will automatically comeback based on the result:

$$X_{ij} + Y_{ij} = \frac{P}{\theta_{ij}}. \tag{10}$$

The corresponding value of the model is 55%. Through machine learning algorithms and models, although machine

learning algorithms are obvious in many traditional models, it is a different situation for football games. The traditional model established based on the data is more than 10%:

$$(x + y)^n = \sum_{k=0}^{n} x^i y^j. \tag{11}$$

If these predictions are used in the 2013–2016 Premier League, the existing football data will be used in the algorithm (55% accuracy through machine learning and 45% accuracy from traditional models). Due to the shortage of seasonal data (produced by only 38 examples), the traditional model can adapt to the logical characteristics of the model:

$$(y + x)^n = y + \frac{n(n-1)x^2}{k}. \tag{12}$$

4. Mobile Algorithms and Predictive Models

4.1. K-Means Clustering Algorithm Analysis. The randomness of the initial clustering center point selection is the main disadvantage of the *K*-means clustering algorithm. The randomness of the first batch of odd numbers usually causes the same number of odd numbers to split more or less. Therefore, this function includes larger poles, but only one of many poles gives the best result on a global scale. In the nonconvex function shown in the figure below, due to the different initial clustering initial centers, the objective function decreases along the paths of the four different initial clustering centers, namely, Va, Vb, Vc, and Vd, and finds their respective minimums. Among them, the so-called minimums in Va, Vb, Vc, and Vd are all local minimums, but only the local minimum of Vc iteration is the global minimum. If the initial clustering center point is not properly selected, the *K*-means clustering algorithm often finds a local minimum and stops.

As shown in Figure 2, the main idea of the *K*-means *k*-value adaptive method is to obtain the championship prediction result of the World Cup football match through a *K*-means algorithm. Select k ($k = \text{int}\ n$, n is the amount of data) from the data set. The data center will be used as the starting point, and then, the data set will calculate the data base and enter the interval between the nearest two bases and concentrate them, and calculate and merge the previous evaluation value $E0$ and the combined evaluation value E. If log-log2E $2E0 < 1$, the combination condition is met, the combination is reasonable, and the next combination is performed. Combine the conditions, we output the value of k, and the algorithm ends. Each time *K*-means clustering can get a new cluster center and cluster number k and then use a new organization "k" to optimize the influence of the organization, by affecting the changes in the number of organizational cycles that occur. Give them more accurate predictions.

4.2. Various Improved Algorithms of K-Means. An improved *K*-means algorithm based on the new effectiveness evaluation index (IBWP) is proposed. The algorithm was refined combined with the density method, and an initial center point candidate set was generated. According to the maximum number of clusters in the cluster search, a new clustering effectiveness index is proposed, according to the index to search for the number of clusters, when the value of the index. When the optimal clustering is reached, the corresponding number of clusters is judged to be the optimal number of clusters, but the improved algorithm does not preprocess the abnormal data in the data set. If there is abnormal data, the clustering results will have a certain deviation.

As shown in Figure 3, in addition to *K*-means clustering, the number of clusters needs to be specified in advance, but when clustering a lot of data, the set of data objects is unknown. Therefore, it is very difficult to determine the number of classes. Traditionally, the minimum value of the clustering effectiveness evaluation function is used to determine the number of optimal clusters, but it is necessary to search for the minimum value of the clustering effectiveness evaluation function, which greatly increases the time complexity of the algorithm. In order to analyze the problems and shortcomings of these algorithms, another algorithm was derived, which accurately reduces the color according to the picture [36]. In addition, the pixel values in the image are calculated based on the statistical proportions of different colors in the image, so that the optimal formula of the image matrix solution is determined in the corresponding proportion. Improved algorithms will not mind accuracy, but better algorithms are project oriented and tags can also be installed in it and attach great importance to paint, but the network is neither explained nor accepted. It is recommended to use visualization to capture the characteristics of the network and to establish a new visualization technology—a new visualization method. This is a step scanner that uses strings to get the corresponding responses of neurons. It organizes the corresponding data graphically, reconstructs the matrix, and converts it into thermal imaging. The error that the algorithm bears is the error generated by the algorithm reflected on the image.

In order to prove the effectiveness of the k-value adaptive method in this article, experiments are carried out. The experiment selects the data in the UCI database to proceed. Because the number of clusters in the test set is known, the selection of the initial number of clusters has been fine tuned this time, so that the clustering results can be obtained faster. The median between the known number of clusters and the maximum number of clusters is used as the initial k-value. This experiment uses laboratory server equipment, 2TB hard disk, and 8.00G memory; the experiment is done under the Ubuntu platform, as shown in Table 1.

Because blind equalization requires iterations, the cyclomatic complexity and degree of repetition are quite high. If the method based on random forest classification is still used, it may become very complicated, which is not conducive to best practice planning and to reduce the complexity of the algorithm. This article uses the method of support vector machine recursive elimination to define the characteristics in multiple tail items as a specific multiorder characteristic, select a more important category in advance,

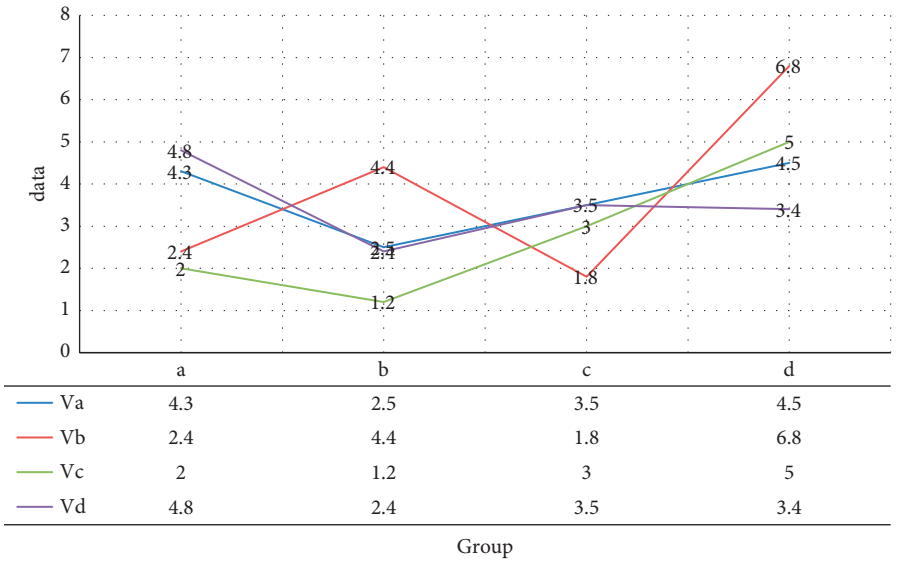

Figure 2: Initial cluster center point.

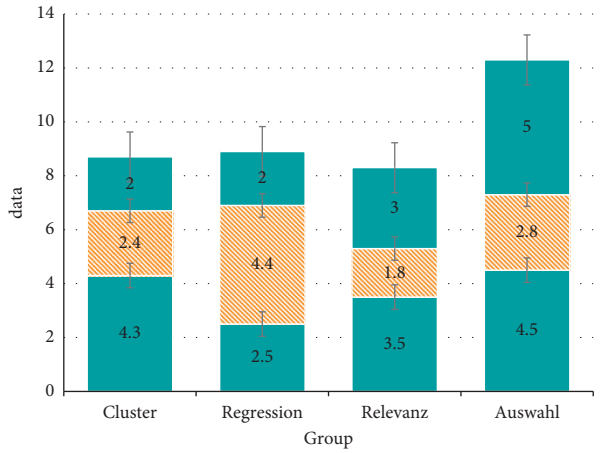

Figure 3: Simplicity naive bayes classifier.

Table1: The final cluster number is obtained.

	Data size	Data attribute	Data category
Iris	150	4	6
Wine	178	3	3
Sight	160	2	5

and then delimit the boundary to decide the result. The process of copying shows that the algorithm is presented in the closest way. Anyway, numbers belonging to a higher category alleviate the complexity as shown in Table 2.

Figure 4 shows the change of the evaluation value E of the two closest match information for each cluster merge. It can be seen from the figure that the evaluation value satisfies the merge condition between the number of clusters 10 and 1 and the number of clusters. When it is 4, the evaluation value E suddenly increases, which indicates that the dispersion degree and the aggregation degree within the group are too high, and the aggregation effect of the group is too low,

Table2: Interclass dispersion aggregation changes.

Uruguay	Mexico
Brazil	Belgium
Argentina	Romania
Spain	Austria
Bolivia	Yugoslavia
Chile	Argentina
Paraguay	Egypt
Peru	Italy

which does not meet the merger conditions. Therefore, the final number of clusters is 5, which is consistent with the actual number of clusters in the Iris data set.

For the adaptive algorithm in the data set wine, the interclass dispersion Disp, the intraclass clustering degree Aggr, and the evaluation value E vary, and the line graph of the evaluation value clearly show the transformation of the evaluation value E, as shown in Table 3:

It can be seen from Figure 5 that the k-value adaptive method optimizes the choice of k-value by merging the two closest classes in each clustering. The k-value of each cluster is updated once, and the final number of clusters can be quickly obtained. Therefore, the adaptive k-value algorithm is effective.

4.3. DPC and k-Means K-Value Adaptive Algorithm Fusion. DPC (Density Peak Based Clustering, abbreviated DPC) algorithm was developed by Alex Rodriguez and Alessandro; he published an elegant algorithm in the American "Science" magazine. The process of this algorithm is: find the maximum value of density as the cluster center, reference the data points according to the density, and then assign these data points to the classes with the closest distance and density greater than it, as shown in Table 4.

As shown in Figure 6, the most common view of the DPC algorithm is as follows: in the data of a World Cup football game, different population groups are more and more far

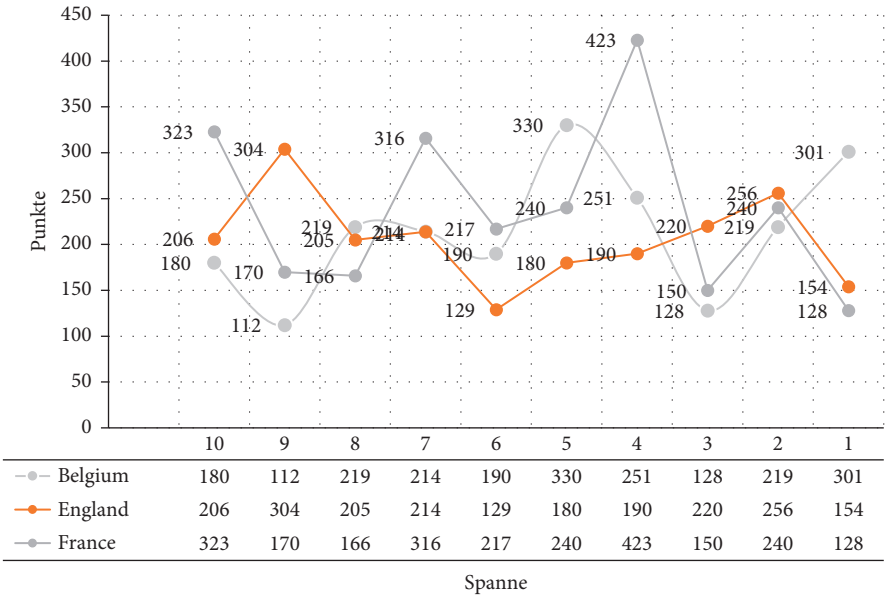

Figure 4: Evaluation value clustering number K-value.

Table 3: Interclass dispersion and intraclass aggregation vary.

k	10	9	8	7	6	5	4	3	2	1
Aggrk	443.89	473.78	520.75	678.9	816.88	1010.89	1321.9	1579.8	2132.3	443.89
Aggr		29.3	46.97	158.1	139.98	194.09	311.01	257.9	552.5	29.3
Dispk	68.76	78.9	94.71	126.46	141.53	155.78	194.54	221.41	240.43	68.76
Disp		10.14	15.81	31.75	15.07	14.25	38.76	26.874	19.02	10.14
E		2.89	2.97	4.97	9.28	13.62	8.02	9.6	29.04	2.89

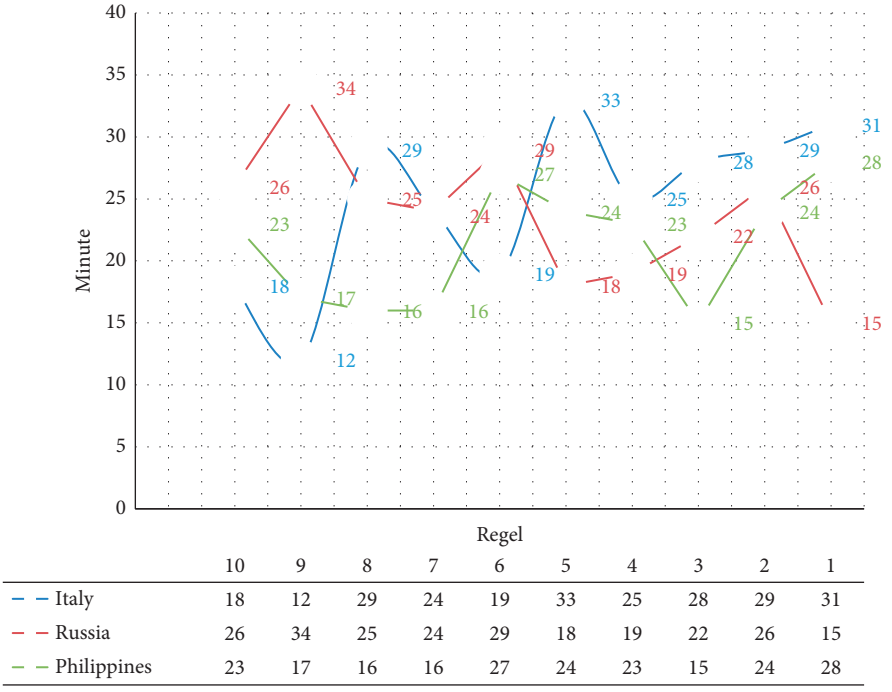

Figure 5: Adaptive process of E value.

away from the area with low population density. The extreme value of different population usually means that the area is very dense, whereas the closer population means the maximum distance between T and t, the extreme density points and abnormal points can be easily distinguished. Data points 1 and 10 are the extreme density values.

Table 4: DPC density clustering algorithm.

	Quality	Volume	Interspace
1	1.33	2.46	72
2	2.51	4.44	42
3	3.54	1.28	33
4	9.53	23.8	45
5	2.52	5.18	84.9
6	2.20	1.17	66
7	1.61	26.9	92.2
8	3.94	19.2	61.8
9	8.68	25	23.4

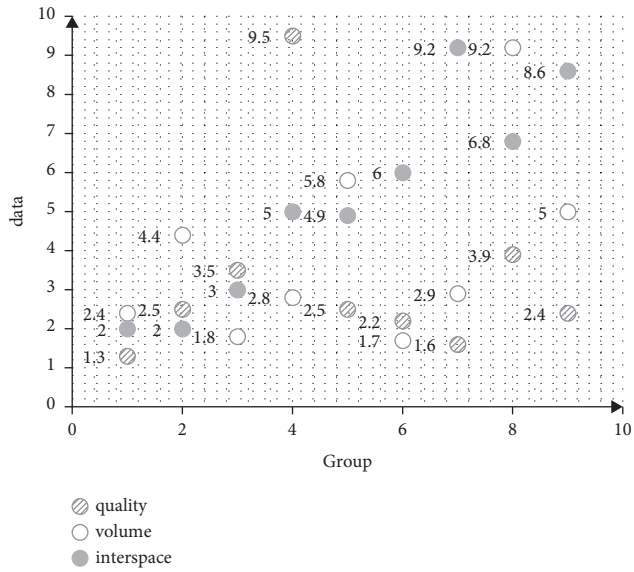

Figure 6: Data processing center density reference.

The advantage of hierarchical clustering is that the principle of the algorithm is easy to understand and simple to implement. The hierarchical relationship between clusters can be detected for any data set, and its scalability is relatively high. The disadvantage is that the time complexity is high, the algorithm cannot withdraw the previous champion prediction results, and the known errors cannot be modified.

This is also true when selecting terminal vectors, although this result precludes generalization of the same result. In other words, you can freely select one text after another, pull in two texts, and choose a vocabulary to use. We want to clean up these data and provide a basis for future analysis. Washing data is a necessary process to transform raw data into "usable" data. Washing data refers to cleaning up data to identify abnormal situations, missing or abnormal points. In order to prevent the model from being changed, the original data will be divided into training and experimental groups. The experimental scheme is designed according to the training of the model and the consistency of the texture. The algorithm was chosen to study the initial gregarious center results that might destroy his effect. Since the *K*-means algorithm randomly selects the initial clustering centers, this easily leads to unstable clustering results. The DPC algorithm can separate the local density extreme points in the data set from the ordinary data points, and select the local density extreme points as the cluster center. Therefore, taking advantage of the DPC algorithm and using the local density extreme points selected by the DPC algorithm as the clustering center of the *K*-means algorithm can effectively avoid the influence of the random selection of the initial center on the *K*-means algorithm. The results obtained are divided into two groups: 60% of the exercise results and 30% of the test results are calculated in order of 100%, and the average of the results is obtained. The data is scattered. Observe whether there are any abnormalities in the data, such as whether the shot is negative or more than 10 points; if the data is correct, you need to check whether it is true, and then further analysis. Because the main components of the DPC must be selected from the editing domain [37], the DPC algorithm clustering center needs to be manually selected and cannot be automatically completed, which will cause the algorithm's execution efficiency to become low. *K*-means requires the user to specify the value of k in advance, so the DPC algorithm needs to automatically obtain k cluster centers in some way.

As shown in Figure 7, the methods of DPCK-K-means and DPC-K-means for selecting initial clustering centers based on density is the same. They are only different in the use of the criterion function. Therefore, the DPCK-K-means algorithm is similar. Compared with the DPC-K-means

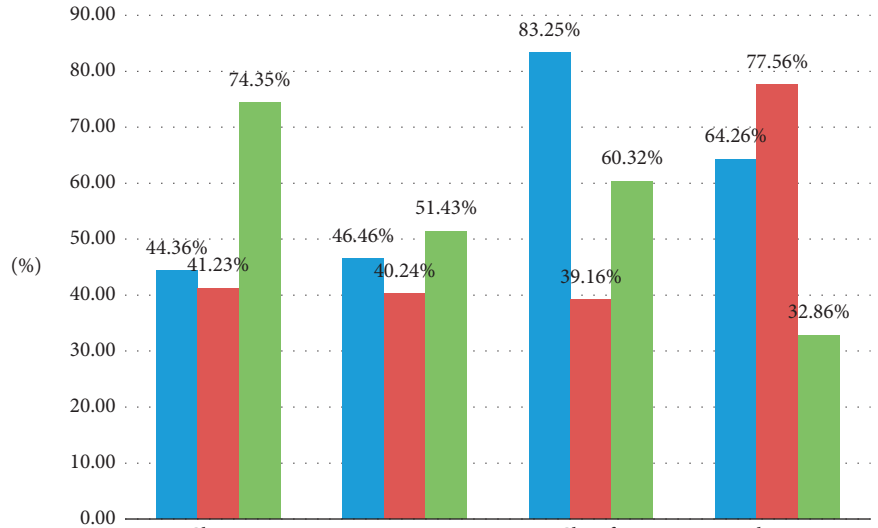

Figure 7: DPCK-means algorithm statistical percentage.

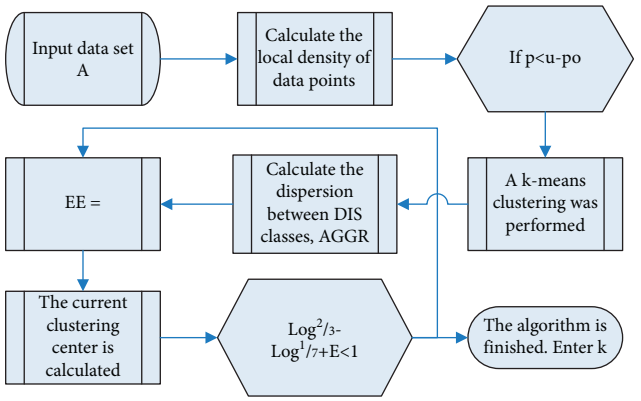

Figure 8: Flow chart of *K*-means K based on DPC.

algorithm, the classification accuracy is only slightly improved. However, the classification numbers of these two algorithms are basically consistent with the real classification results. There is a big difference in the classification of wine data, which is mainly because the characteristics of the data itself are not obvious, and the data within the class are not uniform.

As shown in Figure 8, according to the basic idea of algorithm improvement in the previous section, a specific improvement process is given. The process of the *K*-means *k*-value adaptive algorithm based on DPC uses the international general data set UCI for experiments, and the experiment passes traditional *K*-means clustering algorithm, DPC-K-means (the fusion of DPC and *K*-means algorithm) algorithm, and DPCK-K-means are compared and analyzed, and the improved prediction results are obtained.

5. Conclusions

In this article, the authors first introduce the disadvantages of the *K*-means algorithm, then analyze the research and improvement of the *K*-means algorithm by domestic and foreign scholars, and finally propose their own improved algorithm based on the DPC K-means *k* model (DPCK-K-means) prediction of the championship of the World Cup. On the one hand, through the experiment, the random football match is predicted and its effect is analyzed. On the other hand, the result of the actual football match is taken as the reference group. Also, the results of the match based on the evidence theory prediction method are used as the experimental group. The two sets of data are compared, and the efficiency and error of the method in this article are analyzed through simulation experiment analysis, which proves the

effectiveness of the algorithm, and greatly improves the effectiveness and accuracy of the prediction results.

Mobile algorithms are used in various fields of human production, not only in football matches but also in credit card standards, maps, smart cities, smart health care, smart transportation, system solutions, and document retrieval. When assessing the risk of credit card, default the use of algorithms can reduce the risk of default by matching with customers who lack credit and analyzing common characteristics. In the business field, mobile algorithms help to identify different user groups, increase their profits by analyzing the characteristics of these groups, and develop different marketing methods. There are still some shortcomings in this article. For example, in e-commerce, mobile algorithms can identify consumers who need similar scanners and help consumers understand the needs of different types of users, which will be improved in future work.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that there are no conflicts of interest with any financial organizations regarding the material reported in this article.

Acknowledgments

This work was supported by introducing talents to start scientific research projects of Guizhou University of Finance and Economics (2019YT048).

References

[1] T. Zhu, "The application of machine learning algorithms in data mining," *Digital Technology and Applications*, vol. 3, p. 166, 2017.

[2] Z. Lu, G. Chen, L. V. Zonglei, and G. Chen, "Study on probability distribution prediction models based on observational learning," *Computer and Digital Engineering*, vol. 44, no. 9, pp. 1635–1640, 2016.

[3] F. Li, "Analysis of goal characteristics in men's football matches-take the 17th to 21st Fifa world cup as an example," *Journal of Nanjing Institute of Physical Education*, vol. 2, no. 5, pp. 52–61, 2019.

[4] Y. Fu, W. Dong, L. Yu, and Y. Du, "Software defect prediction model based on combined machine learning algorithms," *Computer Research and Development*, vol. 54, no. 3, pp. 633–641, 2017.

[5] Z. Lv, L. Qiao, D. Chen, R. Lou, J. Li, and Y. Li, "Machine learning for proactive defense for critical infrastructure systems," *IEEE Communications Magazine*, 2020.

[6] W. Liu, W. Jiang, and Y. Ye, "Data mining practical machine learning technology," *Information and Computers*, vol. 405, no. 11, pp. 164-165, 2018.

[7] L. Li, Y. Lin, D. Cao, N. Zheng, and F. Wang, "Parallel learning-a new theoretical framework for machine learning," *Journal of Automation*, vol. 43, no. 1, pp. 1–8, 2017.

[8] L. Wei, "Research on software defect prediction techniques based on machine learning," *Journal of Changchun University*, vol. 27, no. 5, pp. 7–9, 2017.

[9] W. Zhou, W. Xiaoyan, Z. Chen, and Y. Zhou, "Research on electro-epileptic data analysis methods based on machine learning," *Medical Information*, vol. 39, no. 2, pp. 55–59, 2018.

[10] B. Liu, J. He, Y. Geng, and W. Zui, "Summary of advances in the frontiers of the basic system of parallel machine learning algorithms," *Computer Engineering and Applications*, vol. 53, no. 11, pp. 31–38, 2017.

[11] W. Lu, X. Wang, and T. Han, "Research on recommended methods in conjunction with link prediction and ET machine learning," *Modern Book Intelligence Technology*, vol. 4, pp. 38–45, 2017.

[12] J. Yang, "The application of machine learning algorithms in data mining," *Electronic Technology and Software Engineering*, vol. 4, p. 191, 2018.

[13] Y. S. Zhang and S. Liang, "Weighted slope one algorithm for converged machine learning," *Microcomputer Systems*, vol. 37, no. 8, pp. 1712–1716, 2016.

[14] G. Zhao and Z. Xu, "Research on the emotional analysis model of machine learning-based commodity reviews," *Information Security Research*, vol. 3, no. 2, pp. 166–170, 2017.

[15] Q. Xue, Y. Zhu, and J. Wang, "Joint distribution estimation and naïve bayes classification under local differential privacy," *IEEE Transactions on Emerging Topics in Computing*, 2019.

[16] C. Gong, "Youth football training concept and football reserve talent training," *Win the Future*, vol. 12, p. 359, 2017.

[17] M. Yu, F. Xiang, and Z. Zhu, "Educational application and innovation exploration of machine learning in artificial intelligence field of view," *The Journal of Distance Education*, vol. 35, no. 3, pp. 11–21, 2017.

[18] X. Li and Y. Guan, "Research on the application of computer vision and machine learning technology in 3D human animation," *Tomorrow Fashion*, vol. 2, p. 331, 2018.

[19] Y. Huang and L. X. Leihang, "Summary of quantum machine learning algorithms," *Journal of Computer Science*, vol. 41, no. 1, pp. 145–163, 2018.

[20] I. M. El-Hasnony, S. I. Barakat, M. Elhoseny, and R. R. Mostafa, "Improved feature selection model for big data analytics," *IEEE Access*, vol. 8, no. 1, pp. 66989–67004, 2020.

[21] B. Gong, R. Tao, and Z. Dong, "Discussion on some major issues of campus football in my country," *Journal of Shanghai University of Sport*, vol. 41, no. 1, pp. 61–67, 2017.

[22] S. Wan and S. Goudos, "Faster R-CNN for multi-class fruit detection using a robotic vision system," *Computer Networks*, vol. 168, Article ID 107036, 2019.

[23] S. Li, "Research on the bottleneck of youth campus football in my country," *Sports Culture Guide*, vol. 2, pp. 154–156, 2017.

[24] N. Krishnaraj, M. Elhoseny, M. Thenmozhi, M. M. Selim, and K. Shankar, "Deep learning model for real-time image compression in internet of underwater things (IoUT)," *Journal of Real-Time Image Processing*, vol. 17, no. 4, 2020.

[25] M. Elhoseny and K. Shankar, "Optimal bilateral filter and convolutional neural network based denoising method of medical image measurements," *Measurement*, vol. 143, pp. 125–135, 2019.

[26] Y. Yu, "Whether artificial intelligence will eventually surpass human intelligence-a discussion based on the fundamentals of machine learning and human brain cognition," *People's Forum Academic Frontier*, vol. 95, no. 7, pp. 14–23, 2016.

[27] X. Sun and B. Bai, "Machine learning-based NAO robot detection and tracking," *Journal of Changchun Polytechnic*

University (Natural Sciences Edition), vol. 39, no. 2, pp. 116–119, 2016.

[28] X. Huang, "Study on text recognition methods based on artificial intelligence machine learning," *Communication World*, vol. 13, p. 234, 2016.

[29] Z. Mu, "Personalized learning path cracking supported by learner data portraits-value of learning computing," *The Journal of Distance Education*, vol. 34, no. 6, pp. 11–19, 2016.

[30] W. A. Ma and Y. Shang, "Domestic and foreign power distribution cutting-edge technology dynamics and development," *China Journal of Electrical Engineering*, vol. 36, no. 6, pp. 1552–1567, 2016.

[31] A. K. Dutta, M. Elhoseny, V. Dahiya, and K. Shankar, "An efficient hierarchical clustering protocol for multihop internet of vehicles communication," *Emerging Telecommunications Technologies*, vol. 31, no. 5, p. e3690, 2019.

[32] M. Xiao, Z. Fandong, and F. Jufu, "Face recognition method based on sparsely represented characteristics of deep learning," *Journal of Intelligent Systems*, vol. 11, no. 3, pp. 279–286, 2016.

[33] Z. Sun, C. Lu, and Z. Shi, "In-depth study and progress," *Computer Science*, vol. 43, no. 2, pp. 7–14, 2016.

[34] X. Lu, "Key trends, challenges,and important technologies for the future of libraries-based on an analysis of the new media alliance horizon report: 2015 library edition," *Book Intelligence Knowledge*, vol. 2, pp. 26–32, 2017.

[35] Q. Wang, "On the qualitative nature of the content generated by artificial intelligence in copyright law," *Legal Science (Journal of Northwestern University of Political Science and Law)*, vol. 35, no. 5, pp. 148–155, 2017.

[36] S. Wan, Y. Xia, L. Qi, Y. H. Yang, and M. Atiquzzaman, "Automated colorization of a grayscale image with seed points propagation," *IEEE Transactions on Multimedia*, vol. 22, no. 7, pp. 1756–1768, 2020.

[37] Z. Chen and X. Zhu, "Online learner's academic performance prediction modeling study," *Based on Educational Data Mining China Electric Education*, vol. 12, pp. 75–81, 2017.