

## *Retraction*

# **Retracted: Machine Learning-Based Emotion Factor Analysis of Sport Fan Community**

### **Security and Communication Networks**

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

Copyright © 2023 Security and Communication Networks. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

In addition, our investigation has also shown that one or more of the following human-subject reporting requirements has not been met in this article: ethical approval by an Institutional Review Board (IRB) committee or equivalent, patient/participant consent to participate, and/or agreement to publish patient/participant details (where relevant).

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

### **References**

- [1] H. Guo, B. Liu, and Z. Yang, "Machine Learning-Based Emotion Factor Analysis of Sport Fan Community," *Security and Communication Networks*, vol. 2022, Article ID 2674987, 9 pages, 2022.

## Research Article

# Machine Learning-Based Emotion Factor Analysis of Sport Fan Community

Hengchuan Guo, Baoluo Liu , and Zhen Yang

*Luoyang Institute of Science and Technology, Luoyang 471023, Henan, China*

Correspondence should be addressed to Baoluo Liu; 200900500356@lit.edu.cn

Received 30 March 2022; Revised 25 April 2022; Accepted 29 April 2022; Published 14 May 2022

Academic Editor: Chin-Ling Chen

Copyright © 2022 Hengchuan Guo et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Football is one of the most popular sports in the world. As the popularity of football continues to grow worldwide, so does the number of incidents of violence on the pitch. Today, doping, match fixing, black whistles, and football hooliganism are ranked as the four most toxic aspects of sport. How to study the factors that cause aggressive behaviour of fans from a psychological perspective has become a key issue in the field of sports. Therefore, this study proposes a method for mining the psychological factors of sport fan community members based on machine learning clustering. Firstly, three different members of a large fan community, i.e., university students, office workers, and unemployed people, are used as research subjects to investigate the psychological factors influencing fans' aggressive behaviour using a questionnaire method. Secondly, the data obtained were mined and analysed using the K-means clustering algorithm in machine learning techniques. At the same time, a K-means initial clustering centre optimization algorithm based on principal component analysis (PCA) was proposed for the data characteristics of the interaction of psychological factors. The results show that the new algorithm significantly improves the quality of clustering compared with other optimization algorithms and accurately identifies the multiple factors that contribute to the occurrence of fan attacks.

## 1. Introduction

Football is the world's most popular, most widely played, and most influential sport and is known as the "world's number one sport." An exciting game of football attracts thousands of spectators and hundreds of millions of television viewers. The news about football takes up a lot of space in the world's newspapers and magazines. Today, football has become an integral part of people's lives. According to incomplete statistics, there are now about 800,000 teams playing regularly in the world and about 40 million registered players, including about 100,000 professional athletes [1–6].

The reason why football attracts such a large number of spectators and participants is ultimately down to the sport's appeal. People who play football regularly are conducive to developing good mental qualities, but from time to time we hear about an element of discord in football: aggressive

behaviour by fans. When a fan's favourite team loses a match at the time, it can happen that the fans cannot accept the reality of being eliminated. Led by a few fans, the majority of fans follow suit, abusing the players, then attacking the referee, attacking the police, and burning police cars.

Faced with the constant and gradually escalating fan unrest, governments have tried to do everything possible [7–9]. For example, Spain banned Manchester United fans who had been drinking during the Champions Cup match between Manchester United and Real Madrid. Britain punished football hooligans for ten years for not being allowed to watch football. Germany blacklisted some repeat offenders and banned them from attending matches abroad. Belgium kept fans from all countries tightly segregated while conducting massive anti-riot drills. However, these methods have had little effect. What was it about the match that got the crowd so excited? The crowd for a high-level match can range from a few thousand to 50,000 or 60,000. The

spectators vary in composition, from students to social workers, and some know a little bit about football and some know nothing about it, so they each go to the game with a different mentality [10–12]. What exactly is it that causes fans to repeatedly cause trouble during and after matches? Numerous scholars and experts at both home and abroad have made studies to address these aspects, but have never got to the root of such problems.

The main problem to be solved is to analyse the characteristics of fans' aggressive behaviour from the psychological point of view and try to use advanced machine learning technology to dig out what factors induce fans' aggressive behaviour from a large number of psychological questionnaire data, to carry out effective prevention. Clustering is an important method in machine learning [13–15] and an important element in data mining, which is widely used in many fields, including intelligent business, image pattern recognition, statistics and pattern analysis, and information retrieval. The K-means clustering algorithm [16] is widely used in various fields for its simplicity, ease of implementation, efficiency, and scalability. It is widely used in various fields.

Therefore, this study uses the K-means clustering algorithm to mine the required knowledge from a large amount of psychological questionnaire assessment data. Firstly, three different members of a large fan community, i.e., university students, office workers, and unemployed people, are used as research subjects, and a questionnaire is used to investigate the psychological factors affecting the aggressive behaviour of fans. Secondly, a modified K-means clustering algorithm from machine learning techniques was used to mine and analyse the obtained data. The experimental results validate the effectiveness of the proposed improved K-means clustering algorithm in mining the psychological factors of fan community members.

The rest of the study is organized as follows: in Section 2, a related research is studied in detail, while Section 3 provides the detailed improved K-means clustering algorithm based on PCA. Section 4 provides the detailed application of improved K-means clustering to psychological factors in fans. Section 5 provides the results and discussion. Finally, the study is concluded in Section 6.

## 2. Related Research

According to the mental venting perspective discussed by Freud and Lorenz, repressed aggressive forces may evolve into real aggression if they cannot be eliminated in some socially acceptable way. Bennett et al [17] analyse the problems created by football hooliganism from a sociological perspective. Romanet et al [18] point out the importance of the relationship between fans and the team they support. It is argued that stadium violence is actually created by those so-called fans. The so-called fans were not actually watching the game, but simply using the stadium as a place to vent their frustrations.

However, the conclusions drawn from the above studies are some personal empirical analyses and their reliability has yet to be verified. Furthermore, due to the limitations of

manual processing, the sample size of the data analysed by these methods is small and cannot be applied to a large number of fan members. As a result, the conclusions drawn on psychological factors are somewhat controversial. As the basic tools of scientific research move from the traditional “theory + experiment” to the current “theory + experiment + computation,” the importance of data mining and machine learning is becoming increasingly apparent. This is because the purpose of “computing” is often data analysis, and the core of data science is precisely the analysis of data to derive value. Clustering is an important element of data mining and is one of the fastest emerging areas of “new algorithms” in machine learning. Rochat et al. [19] used clustering algorithms to analyse the psychology of “swiping” on mobile apps. Kloos et al. [20] used a cluster randomization algorithm to deliver an online positive psychology intervention to nursing home staff. Aiyer et al. [21] proposed an analysis of psychological transition and neighbourhood relationships in adulthood using a multilevel clustering algorithm. Evidence was found for the direct effects of cluster membership and structural factors on neighbourhood relationships. It is evident from the above studies that clustering algorithms have greater potential for application in psychological factor analysis and mining.

The K-means algorithm, as a representative of division-based clustering algorithm, is again one of the top ten classical algorithms for data mining. In this study, the K-means clustering algorithm is used to mine the psychological factors of fan community members. Due to the complexity of fans' psychological factors, the initial clustering centre randomly selected in the typical K-means algorithm will lead to low clustering quality. The purpose of this study was to improve the K-means algorithm and use it to accurately dig out various factors that lead to fans' aggressive behaviour.

The main contributions of this study are as follows:

- (1) A K-means initial cluster centre optimization algorithm based on principal components analysis (PCA) is proposed for the data characteristics of psychological factor interactions. Compared with other existing clustering algorithms, the new algorithm is more accurate in mining psychological factors.
- (2) The improved K-means clustering algorithm was used to mine multiple causes of fan aggressive behaviour among members of the fan community and to obtain the proportion of each cause. Also, the occupational distribution of the sample of fan community members had to be mined. In addition, a summary analysis of the psychological factors that lead to the generation of fan violence is made.

## 3. Improved K-Means Clustering Algorithm Based on PCA

*3.1. Basic Principles of K-Means Clustering.* As a distance-based partition clustering algorithm, the K-means clustering algorithm has the advantages of simple algorithm structure,

high running efficiency, and wide application range [22–24]. The K-means clustering algorithm is generally optimized by the objective function shown as follows:

$$E = \sum_{j=1}^K \sum_{x \in C_j} \|x - m_j\|^2. \quad (1)$$

It can be seen that the objective function shown in equation (1) is a sum-of-square error calculation process, where  $E$  is the clustering criterion function,  $K$  is the total number of clusters,  $C_j$ ,  $j = 1, 2, \dots, K$  is the cluster  $j$  in the cluster,  $x$  is a clustering target in cluster  $C_j$ , and  $m_j$  is the average size of cluster  $C_j$ . The process of cluster analysis based on the K-means clustering algorithm is shown in Figure 1.

The input parameters of the K-means clustering algorithm are the value  $K$  and the number  $n$  of clustering targets in the dataset  $X$ . The output is the  $K$  clusters that minimize the clustering criterion function  $E$ . The basic flow of the K-means clustering algorithm is as follows [25]:

- Step 1: Input the parameters and initialize the  $K$  clustering centres
- Step 2: Calculate the value of  $E$
- Step 3: Update the centres of each cluster and calculate the new  $E$
- Step 4: Check whether the convergence condition is satisfied
- Step 5: Output the parameters and finish if yes, and skip to Step 2 if no

All division-based clustering algorithm methods have a significant problem [26–28]: they are sensitive to the initial clustering centre. Therefore, for K-means clustering algorithms, the initial clustering centre is a key factor in determining the quality of clustering. A good initial clustering centre not only can effectively avoid the algorithm from falling into local optimum but also can greatly reduce the time overhead of the algorithm, thus improving the algorithm clustering quality and time efficiency. Due to the complexity and interaction of fan psychological factors, randomly selected initial clustering centres can lead to poor clustering quality.

**3.2. Initial Clustering Centre Optimization Method.** To optimize the initial clustering centres, it is first necessary to know what kind of centres are good initial clustering centres. Consider two extreme cases: (1) if the classification of the clusters is known in advance and the centre of each cluster is used as the initial clustering centre, then the algorithm only needs one iteration to obtain a very good clustering result; (2) if  $k$  sample points that are close together and at the edge of the entire sample space are chosen as the initial clustering centres, not only will many iterations be required, but the final clustering result is likely to fall into a local optimum.

Therefore, a good initial clustering centre should satisfy two conditions: (i) they are separated from each other by a certain distance and (ii) they are as close as possible to the

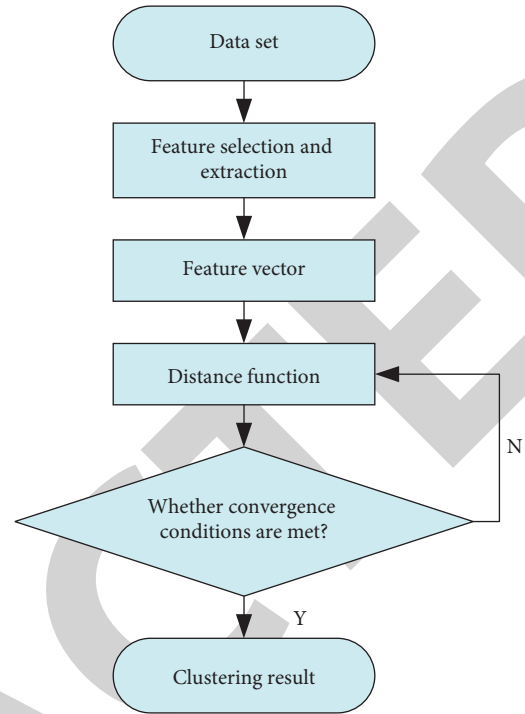


FIGURE 1: Cluster analysis schematic model.

centre of the intrinsically implied cluster. Existing distance-based and recursive-based refinements aim to make the initial clustering centres satisfy condition 1, while density-based refinements combine both condition 1 and condition 2. However, these methods introduce a number of additional parameters, which are often difficult to determine and are not borrowable when the datasets differ.

How do you find the initial cluster centre so that it satisfies both conditions? Consider a simple average height problem: there are 30 students (each of whom knows his or her own height). How do you divide them into “tall” and “short” groups by height (assuming 15 students in each group) and find the approximate average height of each group? A quick and easy way to do this is to have 30 students sorted by height from shortest to tallest, and then, the 7th and 22nd students will be the approximate average height of each group.

Borrowing from the idea of solving the average height problem, for one-dimensional data, the initial cluster centres can be obtained by first sorting and then averaging the middle points. For high-dimensional data, you need to find a way to sort them and then take the same mean points to get the initial cluster centres. As the most commonly used linear dimension reduction method, principal component analysis (PCA) based on multivariate statistics has an excellent performance in feature extraction of high-dimensional data. The research of Iannucci [29] shows that PCA can reduce the dimension of the original features by projection without losing the information as much as possible.

Therefore, this study proposes a PCA-based K-means initial clustering centre optimization algorithm. The main idea of this algorithm is as follows: firstly, the high-dimensional data are reduced to one-dimensional data by

principal component analysis; then the one-dimensional data are sorted in ascending order; then the one-dimensional data are clustered using the K-means algorithm; and finally, the initial clustering centre is obtained from the clustering results.

### 3.3. Steps in the Implementation of the Proposed Algorithm.

The new algorithm uses the K-means clustering algorithm itself to divide the sorted data into  $k$  subsets, which is a good way to reduce the initial clustering centre bias caused by asymmetric clusters in the data sample (i.e., some clusters have more sample points and some clusters have fewer sample points).

The steps of the PCA-based initial clustering centre optimisation algorithm can be briefly described as follows:

**Input:** dataset  $D$ . Suppose it contains  $n$  data sample points, each containing a  $c$ -dimensional attribute, and the number of clusters to be divided is  $k$ .

**Process:**

Step 1: reduce the original multidimensional data to one-dimensional data using the PCA algorithm [30], denoted as  $A_i$  ( $1 \leq i \leq n$ ).

Step 2: sort the one-dimensional data  $A_i$  in ascending order.

Step 3: sort the sorted  $A_i$  into  $k$  clusters using the K-means clustering algorithm, where the  $k$  initial clustering centres are shown as follows [31]:

$$\text{centre}_j = A_i \left( i = \frac{2 \times j - 1}{2 \times k} \times n, \quad 1 \leq j \leq k \right). \quad (2)$$

Step 4: divide the original data into  $k$  subsets based on the classified  $A_i$  and the one-to-one correspondence between  $A_i$  and the original multidimensional data.

Step 5: find the centroids of each of the  $k$  subsets.

Step 6: use the  $k$  sample points nearest to the centroids of the subsets in the original data as the initial clustering centres.

**Output:**  $k$  initial clustering centres.

**3.4. Complexity Analysis.** The main time spent in the principal component analysis is in finding the eigenvalues and eigenvectors. For an  $n \times d$  matrix, it takes  $O(nd)$  to calculate the covariance matrix and  $O(d^3)$  to perform an eigenvalue decomposition on a  $d \times d$  matrix. If the dataset is projected into the first  $m$  principal components, then only the first  $m$  eigenvalues and eigenvectors need to be found. This can be obtained using more efficient methods such as certain calculation, which has a time complexity of  $O(md^2)$ , where  $n$  is the number of data samples,  $d$  is the data dimension,  $k$  is the number of clusters, and  $t$  is the number of generations selected.

The time complexity of the K-means clustering algorithm is  $O(nkt)$ . Since  $m = 1$ , the total time complexity of the PCA-based K-means initial clustering centre optimisation algorithm is  $O(nd + d^2 + nkt)$ .

## 4. Application of Improved K-Means Clustering Algorithm to Psychological Factors in Fans

**4.1. Psychological Definition of Aggressive Behaviour.** According to sport psychologists, aggression is a purposeful act of hurting another person, which can harm their physical or mental health through words or physical actions. In other words, aggression is an intentional action with the aim of causing harm or suffering. Sport psychologists believe that three factors influence the development of aggressive behaviour: firstly, the individual's innate tendency to have a reaction; secondly, the excitability factor in the mind; and thirdly, the experience factor. In short, we see aggression as an individual's reaction to the different levels of anger he or she experiences. A mental model of how aggression arises is shown in Figure 2.

**4.2. Data Acquisition and Processing.** Currently, the main methods of data collection for psychological aspects include expert interviews, literature, questionnaires, and mathematical statistics. In this study, three different members of a large fan community, namely university students, office workers, and unemployed people, were used as the research subjects, and the questionnaire method was used to study the psychological factors affecting the aggressive behaviour of fans. The total number of people in a large fan community was 2927, of which 383 were university students, 2017 were office workers, and 228 were unemployed. Questionnaires were distributed regarding the influencing factors that trigger fan disturbances to understand the causes and control methods of fan disturbances.

After data collection by the above method, fans need to be classified. Based on the purpose of watching football at the venue, fans can be classified into the following five categories: (1) knowledge-seeking type: the main motivation for this type of viewer is to know the outcome of the match; (2) aesthetic type: these are the fans who appreciate the game as a work of art and are as culturally sophisticated as the inquisitive fans; (3) entertainment type: this type of fan comes to the stadium to be entertained, to amuse themselves, and to spend their leisure time; (4) common-seeking type: there is a social psychology of seeking social affiliation and recognition from others; and (5) venturing frenzy type: this segment of fans is particularly fascinated by the heated atmosphere of the stadium. There is no single factor that causes aggressive behaviour in fans. A 12-factor questionnaire was developed for these five categories of fans, as shown in Table 1.

## 5. Experiment and Result Analysis

**5.1. Manual Dataset Validation.** To test the effectiveness of the improved PCA-based K-means clustering algorithm, a comparison with the typical K-means [29] and multilevel clustering [21] was performed on an artificially simulated dataset. The CPU used for the experiments was an Intel(R) Core(TM) i7-3770 CPU @ 3.40 GHz, with 4 GB of RAM, a 64 bit WIN 10 Operating System, and a MATLAB 2016b software environment. A total of three artificially simulated

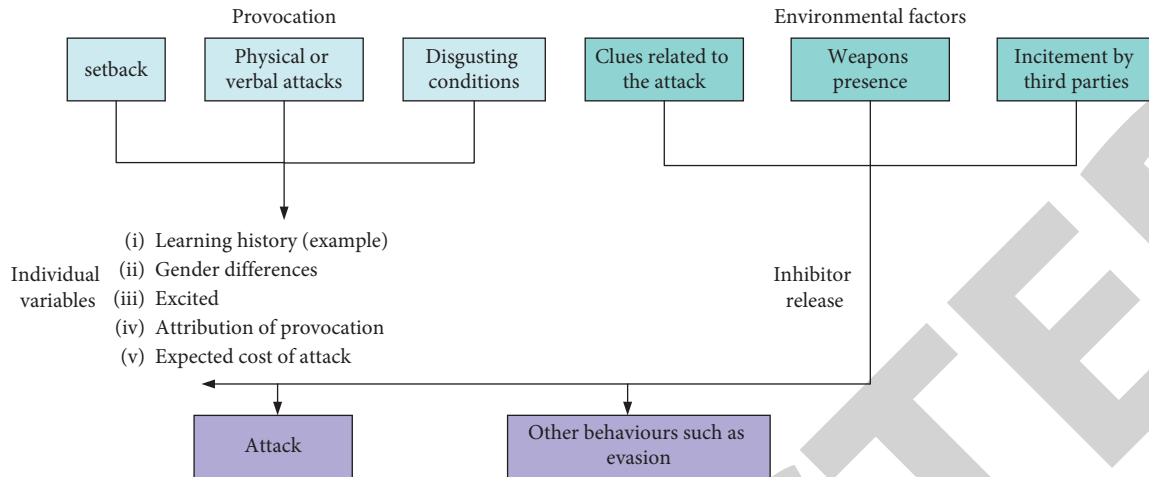


FIGURE 2: Psychological model for the production of aggressive behaviour.

TABLE 1: Questionnaire for the fan community (partial).

No.	Causes of aggressive behaviour by fans
1	The wrong cultural view of football
2	The sport of football itself and the characteristics of its playing process
3	Corruption in the racecourse
4	Moral attribution and player quality
5	Management factors
6	The negative role of the media
7	History of the relationship between the teams in the competition
8	Nature of the competition
9	Cultural bias and inadequate punishment
10	Modern social pressures
11	An increase in fan appreciation
12	Location of the competition
13	Others

datasets (randomly generated using a Gaussian normal distribution,  $\sigma=0.7$ ) were used, including simple and complex clustering structure features: Feature A, Feature B, and Feature C. The parameters of the artificially simulated datasets are shown in Table 2.

A two-dimensional scatter plot of the three datasets is shown in Figure 3. The experimental results of the improved K-means clustering algorithm on three artificial datasets are shown in Table 3.

As can be seen from Table 3, for both the manual dataset 1 and the manual dataset 2, the improved K-means algorithm obtained the smallest number of iterations and the sum of squared errors, indicating the highest quality of clustering. This indicates that the distance between the initial centre and the actual centre obtained by the improved K-means algorithm is the closest. For the artificial dataset 3, the multilevel clustering obtained the largest error sum of squares, the other 2 algorithms were the same, and thus the improved K-means algorithm had the highest quality of clustering. Although the improved K-means algorithm has more iterations than the multilevel clustering, the clustering quality of the multilevel clustering is poor.

Combining the three sets of manual simulation dataset experiments, it can be seen that the initial clustering centres

obtained by the improved K-means algorithm are close to the actual centres of the data samples, and the clustering quality obtained is optimal in all cases, and the number of iterations is also the lowest. Thus, the effectiveness of the improved K-means algorithm is verified.

**5.2. Clustering Mining Results.** A total of 2000 questionnaires were distributed to different fans and 1815 questionnaires were returned, with a return rate of 90.8%. The data from the questionnaires were analysed by applying a modified K-means algorithm for cluster analysis. The dimension of cluster analysis is 12, which corresponds to the 12 factors of the questionnaire. Taking the factors with serial number 1, serial number 2, and serial number 3 in Table 1 as examples, the initial cluster centroids obtained from the modified K-means clustering are shown in Table 4. The final analysis results of the psychological factors about fans' aggressive behaviour after cluster analysis are shown in Table 5.

The greater the proportion of factors, the greater their influence on the aggressive behaviour of fans. Table 5 shows that of the 12 factors that influence the development of aggressive behaviour among fans, "modern social pressure" has the greatest weight, followed by "increased appreciation

TABLE 2: Parameters of the artificially simulated dataset.

Datasets	Structural features	Number of attributes	Sample size for the first category	Sample size for the second category	Sample size for the third category
1	A	2	60	60	60
2	B	2	100	50	50
3	C	2	50	25	50

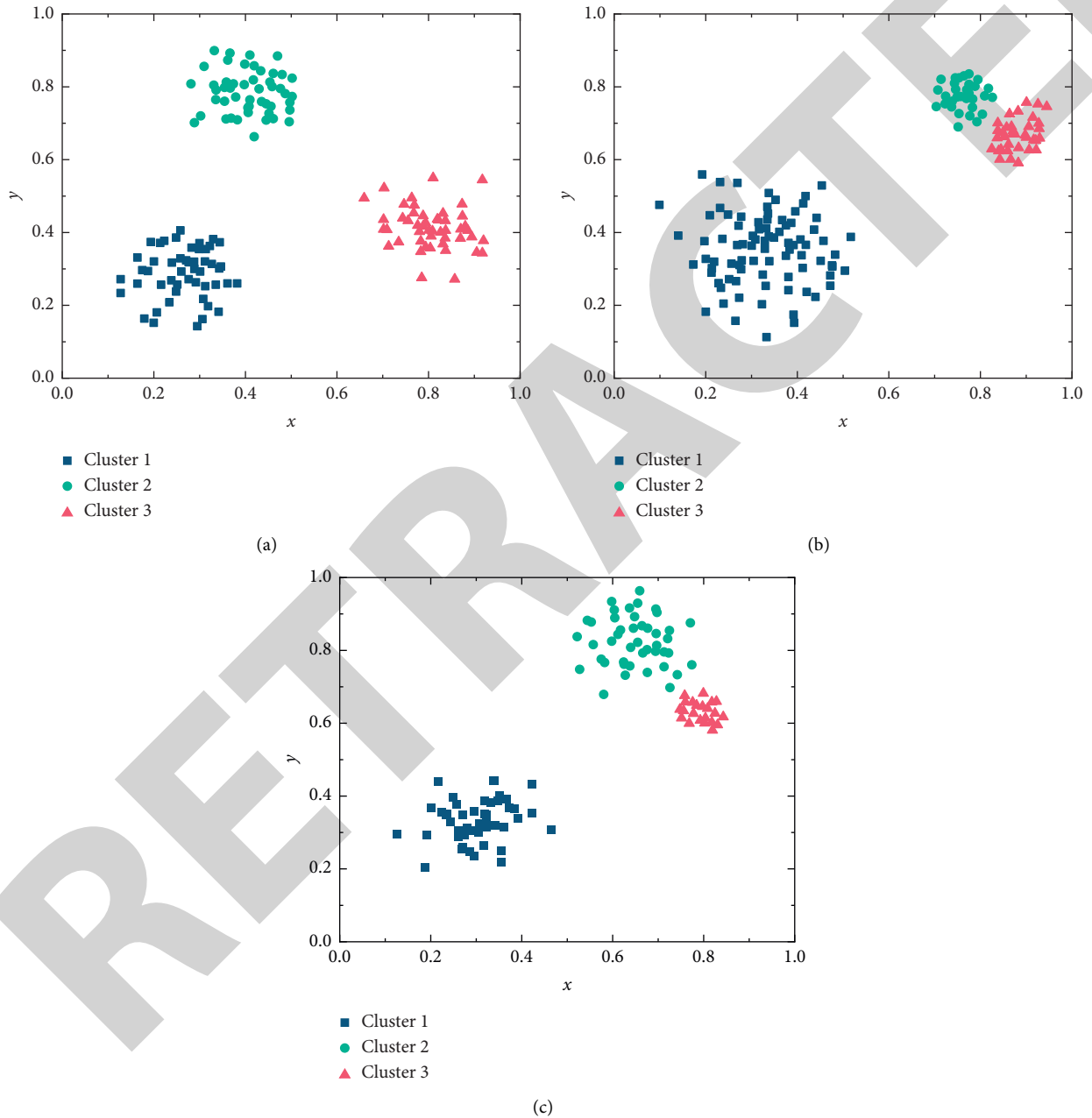


FIGURE 3: Two-dimensional plots of the three artificial datasets. (a) Dataset 1. (b) Dataset 2. (c) Dataset 3.



TABLE 3: Comparison of clustering results on the manual dataset.

Datasets	Clustering algorithms	Number of iterations	Sum of squared errors within a cluster
1	Improving K-means	2	83.530
	K-means	2	86.056
	Multilevel clustering	2	84.647
2	Improving K-means	3	150.782
	K-means	5	167.653
	Multilevel clustering	3	159.031
3	Improving K-means	3	74.052
	K-means	5	74.052
	Multilevel clustering	2	109.581

TABLE 4: Initial cluster centroids obtained by improved K-means clustering.

No.	Initial cluster centroids
1	(1.463326, 0.5131198, 0.6760331, 0.1187017, 0.2534607, 0.6067924, 0.6952479, 0.1956907, 0.8017820, 0.5259556, 0.1564559, 0.42580766, 0.37190083, 0.3680441, 0.3744835, 0.09604106, 0.4188017, 0.3913223, 0.3090909, 0.6013688, 0.5996901, 0.6358471, 0.2923554)
2	(1.666667, 0.2166667, 0.4333333, 0.8736559, 0.2500000, 0.7500000, 0.6666667, 0.7823129, 0.4270833, 0.7604167, 0.8019324, 0.70454545, 0.09090909, 0.1444444, 0.4375000, 0.88172043, 0.4166667, 0.4250000, 0.6166667, 0.6562500, 0.6562500, 0.5729167, 0.8229167)
3	(2.000000, 0.0000000, 0.0000000, 0.0483871, 0.0000000, 0.0000000, 0.1875000, 0.0000000, 1.0000000, 0.6875000, 0.2173913, 0.04545455, 0.21212121, 0.2000000, 0.2500000, 0.0000000, 1.0000000, 1.0000000, 0.9000000, 0.9375000, 1.0000000, 1.0000000, 0.2812500)

TABLE 5: Analysis results of psychological factors about fans' aggressive behaviour.

Psychological factors	Number	Proportion (%)
Modern social pressures	583	32.12
An increase in fan appreciation	220	12.12
Management factors	174	9.59
Moral attribution and player quality	152	8.37
Corruption in the racecourse	142	7.82
The negative role of the media	126	6.94
History of the relationship between the teams in the competition	112	6.17
Nature of the competition	95	5.23
Cultural bias and inadequate punishment	71	3.91
The wrong cultural view of football	63	3.47
The sport of football itself and the characteristics of its playing process	44	2.42
Location of the competition	21	1.16
Others	12	0.66

of fans.” To better analyse the causal factors of football spectator violence, a cluster analysis was conducted on the occupational distribution of this large fan community, as shown in Table 6.

In general, the fan groups in the large fan communities surveyed have the following distinctive features: (1) male-to-female ratio: men predominate, with a male-to-female ratio of 2.28:1; (2) age distribution: the highest proportion of people aged between 19 and 30, at 68.3%; (3) education level: mainly secondary education; (4) political outlook: mainly members of the Communist Youth League and the masses; (5) marital status: the proportion of unmarried people is higher than that of married people, with a lower divorce rate; (6) income level: the proportion of unmarried people is higher than that of married people (81.7%); (5) marital status: the proportion of unmarried people is higher than that of married people, and the divorce rate is lower; (6)

income level: mainly in the range of less than 3,000 yuan (81.7%); and (7) occupational distribution: mainly in the category of “enterprise employees” (36%).

5.3. *Analysis of Psychological Factors.* A summary of the triggers for violence by members of the fan community, based on the first 3 main psychological factors derived from the improved K-means clustering algorithm, is as follows:

- (1) Modern Social Pressure. The rapid development of all aspects of modern society, the accelerated pace of life, and increasingly fierce competition have put people under enormous psychological pressure. In the context of everyday life and work, this suppressed emotion cannot be released due to the constraints of self-imposed social identity. Depersonalization is the loss of one's identity in a crowd, and depersonalized



TABLE 6: Occupational distribution of the survey sample.

Occupational distribution	Proportion (%)
Business managers	9.1
Corporate staff	36
Science, education, culture, and health professionals	7.1
Professional and technical staff	5.3
General civil servants	2.4
Financial practitioners	4.3
Service industry practitioners	8.5
Individuals, private owners	4.7
Freelancers	13.7
Retired or unemployed	6.9
Others	2.01

individuals are prone to violent behaviour. Because emotions dominate thinking during the game, fans temporarily abandon the norms of everyday life and no longer care about their identity, giving rise to depersonalized and arbitrary expressions of their feelings. Fans will behave more boldly than usual and are prone to violent behaviour.

- (2) An Increase in Fan Appreciation. With the rapid development of the economy, fans are also able to see higher-level matches through media such as TV or newspapers, and at the same time, the level of fans' appreciation of football is increasing. Fans are also demanding more and more from the game. If the level of excitement that can be achieved on the field does not meet the growing level of appreciation, it can lead to discontent among spectators. Fans may provoke the referee and players.
- (3) Management Factors. Poor organizational management and improper maintenance of order on the field of play can also be a trigger for violence. Inappropriate enforcement tactics by match enforcers are directly related to fan violence. The failure of managers to disperse crowds assembled at the end of matches and the inability of live enforcers to stop some excesses in time can easily lead to spectator violence.

## 6. Conclusions

This study proposes a method for mining the psychological factors of fan community members based on improved K-means clustering. The proposed method can mine the data from a large number of psychological questionnaires to find out which factors induce fan aggression, so that effective prevention and control can be carried out. A PCA-based K-means initial clustering centre optimization algorithm is proposed for the data characteristics of the interaction of psychological factors, so that multiple causes of fan aggressive behaviour can be mined more efficiently, and the proportion of each cause can be obtained. Finally, a summary analysis of the psychological factors that lead to the generation of fan violence is made. Although the PCA-based K-means

initial cluster centre optimization algorithm can reduce the number of iterations, the time complexity in the initial cluster centre selection stage is high, and therefore, the overall time is still relatively high. Subsequent research will attempt linear discriminant analysis and kernel principal component analysis as an alternative to PCA.

## Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest to report regarding this study.

## References

- [1] D. C. Viano, C. Withnall, and D. Halstead, "Correction to: impact performance of modern football helmets," *Annals of Biomedical Engineering*, vol. 46, no. 1, pp. 208–210, 2018.
- [2] H. Sarmento, M. T. Anguera, A. Pereira, and D. Araújo, "Talent identification and development in male football: a systematic review," *Sports Medicine*, vol. 48, no. 4, pp. 907–931, 2018.
- [3] Constantinou and C. Anthony, "Dolores: a model that predicts football match outcomes from all over the world," *Machine Learning*, vol. 108, no. 1, pp. 49–75, 2018.
- [4] S. N. Jensen, A. Ivarsson, J. Fallby, S. Dankers, and A. M. Elbe, "Depression in Danish and Swedish elite football players and its relation to perfectionism and anxiety," *Psychology of Sport and Exercise*, vol. 36, pp. 147–155, 2018.
- [5] C. Eirale, "Hamstring injuries are increasing in men's professional football: every cloud has a silver lining?" *British Journal of Sports Medicine*, vol. 52, no. 23, p. 1489, 2018.
- [6] J. M. Buldú, J. Busquets, J. H. Martínez et al., "Using network science to analyse football passing networks: dynamics, space, time, and the multilayer nature of the game," *Frontiers in Psychology*, vol. 9, pp. 115–124, 2018.
- [7] S. Muniroglu and E. Subak, "A modified t-test for football referees to test agility, quickness and sprint performances," *Journal of Education and Training Studies*, vol. 6, no. 5, pp. 10–17, 2018.
- [8] J. P. Berkowitz and C. A. Depken, "A rational asymmetric reaction to news: evidence from English football clubs," *Review of Quantitative Finance and Accounting*, vol. 51, no. 2, pp. 347–374, 2018.
- [9] H. Folgado, B. Gonçalves, and J. Sampaio, "Positional synchronization affects physical and physiological responses to preseason in professional football (soccer)," *Research in Sports Medicine*, vol. 26, no. 1, pp. 51–63, 2018.
- [10] M. C. Leitner and F. Richlan, "Analysis system for emotional behavior in football (aseb-f): matches of fc red bull salzburg without supporters during the covid-19 pandemic," *Humanities and Social Sciences Communications*, vol. 8, no. 1, pp. 14–22, 2021.
- [11] Y. Nagashima and A. Horikawa, "Association between 24 h urinary sodium and potassium excretion and dietary intake in Japanese male adolescent football players," *Adolescents*, vol. 1, no. 4, pp. 461–472, 2021.

- [12] B. Nordberg, D. W. Lopez-Hernandez, A. Bueno et al., "A-122 verbal fluency switching and clustering performance in retired professional football players," *Archives of Clinical Neuropsychology*, vol. 36, no. 6, p. 1172, 2021.
- [13] A. Challa, S. Danda, B. S. D. Sagar, and L. Najman, "Power spectral clustering," *Journal of Mathematical Imaging and Vision*, vol. 62, no. 9, pp. 1195–1213, 2020.
- [14] Y. Xu and M. Wei, "Multi-view clustering toward aerial images by combining spectral analysis and local refinement," *Future Generation Computer Systems*, vol. 117, pp. 138–144, 2021.
- [15] H. Wu and Z. Qu, "Gray clustering model based on the degree of dynamic weighted incidence for panel data and its application," *Grey Systems: Theory and Application*, vol. 10, no. 4, pp. 413–423, 2020.
- [16] A. E. Ezugwu, A. M. Ikotun, O. O. Oyelade et al., "A comprehensive survey of clustering algorithms: state-of-the-art machine learning applications, taxonomy, challenges, and future research prospects," *Engineering Applications of Artificial Intelligence*, vol. 110, no. 4, pp. 104743–105197, 2022.
- [17] P. G. Bennett, M. R. Dando, and R. G. Sharp, "Using hypergames to model difficult social issues: an approach to the case of soccer hooliganism," *Journal of the Operational Research Society*, vol. 31, no. 7, pp. 621–635, 1980.
- [18] I. Romanet, N. Graillon, M. K. Le Roux et al., "Hooliganism and maxillofacial trauma: the surgeon should be warned," *Journal of Stomatology, Oral and Maxillofacial Surgery*, vol. 120, no. 2, pp. 106–109, 2019.
- [19] L. Rochat, F. Bianchi-Demicheli, E. Aboujaoude, and Y. Khazaal, "The psychology of "swiping": a cluster analysis of the mobile dating app Tinder," *Journal of Behavioral Addictions*, vol. 8, no. 4, pp. 804–813, 2019.
- [20] N. Kloos, C. H. Drossaert, E. T. Bohlmeijer, and G. J. Westerhof, "Online positive psychology intervention for nursing home staff: a cluster-randomized controlled feasibility trial of effectiveness and acceptability," *International Journal of Nursing Studies*, vol. 98, no. 10, pp. 48–56, 2019.
- [21] S. M. Aiyer, R. Zahnw, and L. A. Mazerolle, "Developmental transitions during adulthood and neighborliness: a multilevel cluster analysis," *Journal of Community Psychology*, vol. 48, no. 2, pp. 237–257, 2020.
- [22] Z. S. Chen, X. Zhang, W. Pedrycz, X. J. Wang, and K. S. Chin, "K-means clustering for the aggregation of hflts possibility distributions: n-two-stage algorithmic paradigm," *Knowledge-Based Systems*, vol. 227, pp. 107230–108119, 2021.
- [23] R. D. Dana, D. Soilihudin, R. H. Silalahi, D. A. Kurnia, and U. Hayati, "Competency test clustering through the application of principal component analysis (pca) and the k-means algorithm," *IOP Conference Series: Materials Science and Engineering*, vol. 1088, no. 1, pp. 012038–012126, 2021.
- [24] A. Rehman, J. Jaafar, M. Omar, S. Basri, I. Abdul, and Q. Uddin, "Suitable personality traits for learning programming subjects: a rough-fuzzy model," *International Journal of Advanced Computer Science and Applications*, vol. 8, no. 8, pp. 153–162, 2017.
- [25] W. Tao and T. Liu, "Building ontology for different emotional contexts and multilingual environment in opinion mining," *Intelligent Automation and Soft Computing*, vol. 24, no. 1, pp. 65–72, 2018.
- [26] N. Gokilavani and B. Bharathi, "Test case prioritization to examine software for fault detection using pca extraction and k-means clustering with ranking," *Soft Computing*, vol. 25, no. 7, pp. 5163–5172, 2021.
- [27] S. Zhang, C. Bi, and M. Zhang, "Logistics service supply chain order allocation mixed K-Means and Qos matching," *Procedia Computer Science*, vol. 188, pp. 121–129, 2021.
- [28] I. D. Borlea, R. E. Precup, A. B. Borlea, and D. Iercan, "A unified form of fuzzy c-means and k-means algorithms and its partitional implementation," *Knowledge-Based Systems*, vol. 214, no. 23, pp. 106731–107121, 2021.
- [29] L. Iannucci, "Chemometrics for data interpretation: application of principal components analysis (PCA) to multivariate spectroscopic measurements," *IEEE Instrumentation and Measurement Magazine*, vol. 24, no. 4, pp. 42–48, 2021.
- [30] D. Hong, L. Balzano, and J. A. Fessler, "Asymptotic performance of PCA for high-dimensional heteroscedastic data," *Journal of Multivariate Analysis*, vol. 167, pp. 435–452, 2018.
- [31] Z. Sheikhsosseini and H. Monkaresi, "Delineation of potential seismic sources using weighted K-means cluster analysis and particle swarm optimization (PSO)," *Acta Geophysica*, vol. 69, no. 6, pp. 2161–2172, 2021.