A Method for Integrating Sports Information Resources Based on Fuzzy Clustering Algorithm

Xiaoxian Xu[1](✉) and Qiao Wu[2]

[1] Physical Education Department, Xi'an Shiyou University, Xi'an 710065, China
13289273723@163.com
[2] Changchun University of Architecture and Engineering, Changchun 130119, China

Abstract. To improve the accuracy of sports information resource integration, a fuzzy clustering algorithm based method for sports information resource integration is studied. First, classify the sports information resources. According to the classification results of resources, build the sports information resource model. Use different sports concepts as nodes and their relationships as edges to build the concept network model. Based on the concept network model, denoise the sports information data. Based on the denoised sports information data, use the fuzzy clustering algorithm to cluster the sports information cluster analysis, Obtain relevant clusters of data, and then adjust the clustering algorithm parameters accordingly through statistical analysis of the clustering results to obtain accurate and effective integration results of sports information resources. The experimental results show that the accuracy of sports information integration in this method is the highest at 98%, the recall rate is the highest at 96%, the F1 is the highest at 0.97, and the longest time is 3.77 s, indicating the practicality of this method.

Keywords: Sports information resources · Fuzzy clustering algorithm · Conceptual network model

1 Introduction

In contemporary society, sports competitions have gradually developed from a simple sports event to providing entertainment and commercial value for people [1, 2]. Sports competitions [3, 4] are no longer just pure competitions between individuals and teams, but there are many factors and issues that need to be considered. Sports information resources are generated to address these issues. It is also closely related to sports, training, management, and business [5]. Firstly, in terms of sports, sports information resources can help players and coaches better understand their opponents' tactics and lineup, in order to respond faster and more accurately. In terms of training, sports information resources can provide more data and statistical results, which can be compared with past experience and judgments, helping coaches better improve and adjust training plans to improve the level of athletes. In terms of management, sports information resources can provide more useful information for event organizers and management institutions. For

L. Yun et al. (Eds.): ADHIP 2023, LNICST 548, pp. 3–19, 2024.
https://doi.org/10.1007/978-3-031-50546-1_1

example, TV broadcast data, event ticket booking, etc. [6]. By analyzing this data, managers can develop better marketing strategies, increase advertising revenue, and better operate competitions.In terms of business, with the increasing popularity of sports competitions, more and more enterprises and brands have noticed the commercial value of sports events and have invested in the precise marketing strategy of sports events. At this time, sports information resources have become particularly important [7]. Through the use of sports information resources, major brands can obtain more user data and traffic, improve brand value and brand influence, and enhance the stickiness with consumers. In short, sports information resources play a very important role in promoting the development of sports industry. The accuracy and timeliness of sports information resources are directly related to the smooth development of sports events, the healthy development of sports industry and the audience's feeling of watching sports events.

Therefore, the integration of sports information resources is a very important part of the current development of the sports industry. With the continuous development and popularization of computer technology and information network technology, sports information resources now widely cover athlete personal information, technical and tactical data, competition videos, ticket sales, marketing, news reports, and other aspects of information [8]. However, these different types of information are usually stored in different databases, using different systems and platforms, and cannot be shared or interacted with each other. Therefore, integrating various sports information resources is very important for promoting the development of the sports industry and improving the level and quality of sports. Effective integration of sports information resources can improve data validity: Data from different sources often have problems such as duplication, missing, and errors. After integration, it can effectively eliminate duplication, fill in gaps, verify, and improve the reliability and effectiveness of the data. Ability to better analyze and predict: The integrated data is richer and more comprehensive, enabling better data mining and analysis, and making predictions and analyses on the performance of sports athletes, event results, etc. Can provide support for management and decision-making: The integrated data can provide basic data for personnel engaged in sports management and decision-making, analyzing trends and needs, formulating better plans and decisions, and better meeting the needs of athletes and spectators. Can improve efficiency and reduce costs: The integration of sports information resources can realize the automatic collection, processing and sharing of data, reduce the duplication of work, improve work efficiency, and reduce the cost of data management and processing. Can promote marketing: Integrated data can provide strong support for the marketing of sports industry, promote sports events and brands, and provide better commercialization opportunities for sports enterprises. In short, the integration of sports information resources is very important for comprehensively improving the management and competitive level of sports, promoting the healthy development of sports industry and realizing the sustainable development goals. Some relevant researchers have studied the integration methods of sports information resources, for example, literature [9] proposes the integration study of college sports teaching and ethnic traditional sports resources, literature [10] proposes the optimization and integration of Chinese ice and snow sports resources based on the perspective of H-O-S theory, but the accuracy of the above methods is low, which affects the effect of sports information integration.

The fuzzy clustering algorithm can process fuzzy data and handle fuzzy and uncertain data without the need to divide the data into clear categories. At the same time, it has a certain degree of flexibility, and the clustering results obtained by fuzzy clustering algorithms can be fuzzy, rather than unique hard classification results. This makes the clustering results more flexible and can better reflect the similarity between data. And the anti noise ability is strong: the fuzzy clustering algorithm has a strong ability to resist outlier and noise. Suitable for data analysis and processing in various fields. This article studies the integration method of sports information resources based on fuzzy clustering algorithm.

2 Research on Sports Information Resource Integration Based on Fuzzy Clustering Algorithm

2.1 Classification of Sports Information Resources

Sports information resources refer to various data and information related to the field of sports, which can be used to support various decisions and activities, including but not limited to athlete personal data, competition results, event arrangements, schedule, player data, hot topics, historical records and trend analysis. We can analyze the characteristics of excellent athletes, factors that win in competitions, historical development trends of sports events, et al. [11]. Sports information resources can be classified according to different dimensions, and the following are some common classification methods:

1. Competitive events: including football, basketball, volleyball, track and field, swimming and other kinds of sports.
2. Event type: including professional league, national championship, World Championship, Olympic Games, World Cup and other types of sports events.
3. Data types: including game results, player data, team data, tactical analysis, technical statistics and other different types of data.
4. Data sources: including official data, third-party data, social media data and other types of data sources.
5. Time dimension: including historical data, real-time data, forecast data and other data of different time dimensions.
6. Business applications: including sports analysis, data analysis, media reports, on-site operation of sports events, event promotion and other business applications.

Different classification methods can provide demand oriented data extraction and integration for different business needs, meeting the data needs in different scenarios. The above classification process has completed the classification of sports information resources, laying a solid foundation for the subsequent construction of sports information resource models.

2.2 Construction of Sports Information Resource Model

Based on the above classification results of sports information resources, the model of sports information resources is constructed. The conceptual network model is a computer model that simulates the structure of human knowledge. The purpose of constructing sports information resource model is to prepare for the introduction of fuzzy clustering algorithm and the construction of sports information resource integration method. The better the sports information resource model, the better the integration effect. The core idea of the model is to express knowledge as concepts, and to construct knowledge network by expressing the relationship between concepts. Conceptual network model, also known as conceptual map model, is a graphical model that represents concepts and their relationships. It can be mainly used for knowledge management and knowledge representation. It can help people organize complex knowledge and information, and facilitate people to understand and use these knowledge and information [12]. It can also be used in the fields of natural language processing, information retrieval, text classification and machine translation. In this paper, the concept network model is adopted to construct the sports information resource model.

The conceptual network consists of nodes and arcs. Construct a conceptual network model by using different sports concepts as nodes and their relationships as edges. The network includes two types of nodes: concept nodes and document nodes. The arc connecting nodes expresses the correlation between nodes and quantifies the strength of the relationship using weights [13]. Set the concept node set as:

$$C = (c_1, c_2, \cdots, c_n) \tag{1}$$

The document node set is:

$$D = (d_1, d_2, \cdots, d_n) \tag{2}$$

$c_i \xrightarrow{\mu} c_j$ represents the correlation weight between concept nodes c_i and c_j as μ, which can also be expressed as $f\left(c_i, c_j\right) = \mu$. $d_i \xrightarrow{\eta} c_j$ represents the correlation weight between d_i and concept c_j as η, which can also be expressed as $f\left(d_i, c_j\right) = \eta$.

Rule 1: If there are nodes c_i, c_j and c_k, the weight of the correlation relation between $f(c_i, c_k) = \alpha, f\left(c_k, c_j\right) = \beta$, c_i and c_j is $F(c_i, c_k) = \min(\alpha, \beta)$.

Rule 2: If multiple paths are connected between node c_i and c_j, the correlation between c_i and c_j is the maximum path weight. The conceptual network model is shown in Fig. 1 below.

As shown in Fig. 1, the basic elements of the conceptual network model are mainly divided into two categories: nodes and edges.

A node represents a concept or entity, usually represented by text or icons. Nodes may include the following content: name, description, attributes, and labels. The specific element content is shown in Table 1.

Edges represent relationships or connections between nodes. Edges can also have attributes. There are several common types of edges: associative edges, inherited edges, combined edges, and aggregated edges. The specific element content is shown in Table 2.

Building a sports information resource model based on the above element content, the specific process is as follows: Document d is preprocessed (word segmentation, removal of stop words) and represented as keyword set $T = \{t_1, t_2, \cdots, t_n\}$. The frequency of each keyword appearing in the main text, title, keyword, hyperlink, and hyperlink

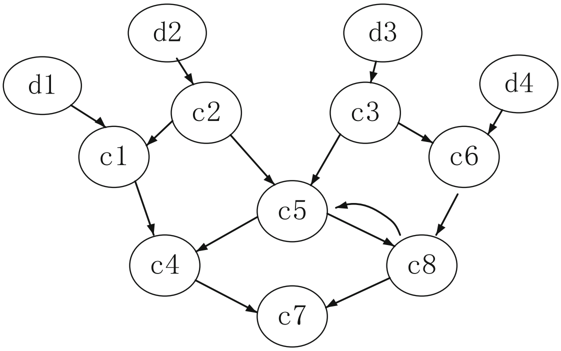


Fig. 1. Schematic diagram of conceptual network model

Table 1. Main elements of conceptual network model nodes

Element	Element content
Name	The name of the concept or entity that the node represents
Description	Description of a concept or entity corresponding to a node, including text, pictures, or videos
Attribute	Attributes of the concept or entity corresponding to the node, such as brand, color, size, et al.
Label	Nodes can use multiple labels to identify and classify the concepts or categories to which they belong

Table 2. Main elements of conceptual network model

Element	Element content
Associated edge	Indicates a relationship between two nodes, such as “basketball player” and “basketball game”
Inherited edge	Indicates that one node is a specialization of another node. For example, “China National Football team” is a subclass of “Asian Cup Football match”
Combined edge	Indicates that one node is a part of another node, for example, “Player v is a component of the” Chinese Women’s Volleyball team”
Convergent edge	Indicates that one node contains other nodes, but other nodes can belong to different instances of other nodes. For example, “football team” includes “football player”

description in the statistical word set is represented as $tf_{other}(t_i)$, $tf_{title}(t_i)$, $tf_{key}(t_i)$, $tf_{url}(t_i)$, and $tf_{Anchore}(t_i)$. The frequency calculation formula for keyword t_i is:

$$tf(t_i) = tf_{other}(t_i) + A_1 \cdot f_{title}(t_i) + A_2 \cdot f_{key}(t_i) + A_3 \cdot f_{url}(t_i) + A_4 \cdot f_{anchor}(t_i) \quad (3)$$

where, A_1, A_2, A_3 and A_4 are adjustment coefficients. Then the formula for calculating the weight of feature words in the document is:

$$w_t \rightarrow d(t_i, d) = tf(t_i) \times \log\left(\frac{N}{df(t_i)} + 0.5\right) \tag{4}$$

Among them, $df(t_i)$ represents the number of documents containing keyword t_i, and N represents the total number of documents. Words are the manifestation of concepts, and a concept node may contain multiple corresponding words. Let the words corresponding to concept node c_i form a set $T_i = (t_1, t_2, \cdots, t_m)$, represented in vector form $c_i = \langle w_{i1}, w_{i2}, \cdots, w_{im}\rangle$, where w_{ij} represents the weight of keyword t_j in concept node c_i. Calculate the correlation between document d and concept c_i as follows:

$$rel_{d\rightarrow c}(d, c_i) = \frac{\sum_{j=1}^{m} w_{t\rightarrow d}\left(t_j, d\right) + w_{t\rightarrow c}\left(t_j, c_i\right)}{Max(Tn(d), Tn(c_i)) * 2} \tag{5}$$

In the formula, $w_{t\rightarrow c}\left(t_j, c_j\right)$ represents the weight of keyword t_j in concept node c_i, $Tn(d)$ represents the total weight of all keywords in document d, and $Tn(c_i)$ represents the total weight of keywords contained in concept node c_i.

There are semantic associations between words contained in the same document. This correlation is formally manifested as the co-occurrence of words [14]. We use these phenomena to find correlations between concepts. Some samples are selected to constitute sample set S, $S = (S_1, S_2, \cdots, S_M)$ and M as the number of documents. Set the concept node set $C = (c_1, c_2, \cdots, c_n)$, and calculate the correlation between the documents in the sample set and the concept node. For concept node c_i, its correlation with the document can be expressed in vector form: $c_i = \langle e_{1i}, e_{2i}, \cdots, e_{Mi}\rangle$, e_{ji} represents the correlation between document d_j and concept node c_i. The formula for calculating the correlation between conceptual nodes c_i and c_j is:

$$rel_{c\rightarrow c}(c_i, c_i) = \frac{\sum_{k=1}^{M} e_{ki} e_{kj}}{\sqrt{\sum_{k=1}^{M} e_{ki}^2 \sum_{k=1}^{M} e_{kj}^2}} \tag{6}$$

The generation of concept nodes in conceptual networks can be achieved through two methods: clustering and gradual addition. When using clustering methods, each keyword in the initial stage corresponds to an independent concept node. Calculate the correlation of concept nodes, and according to the set threshold, concept nodes with correlation exceeding the threshold are merged into new nodes.

Gradually add methods to implement concept mapping using HowNet. CNKI is a common sense knowledge base based on concepts and their characteristics, revealing the relationships between concepts and the characteristics they possess. CNKI describes various relationships between concepts and their attributes. By querying CNKI, we can obtain the concepts corresponding to feature words. We map the keywords in the document to the corresponding concepts. Due to the existence of multiple semantics in some feature words, there may be a one-to-many mapping situation, so auxiliary processes need to be taken to confirm the correct conceptual mapping relationship.

Define co-occurrence: If feature words x and y appear in the same sentence in document d_i, they are considered co-occurrence, and the expression of interword co-occurrence rate CO is as follows:

$$CO(x, y) = \frac{freq(x, y)}{freq(x) + freq(y)} \tag{7}$$

Among them, $freq(x, y)$ is the number of sentences where the feature word x and y co appear, and $freq(x)$ is the number of sentences where the feature word x appears. Document feature word set $T = \{t_1, t_2, \cdots\}$, co-occurrence feature word set for feature word $t_i \in T$ with multiple mapping relationships:

$$T_{CO}(t_i) = \left\{t_j \middle| CO\left(t_i, t_j\right) \geq \psi \right\} \tag{8}$$

ψ is the default threshold. The concept of t_i is c_1, c_2,... c_k. The feature word t_i belongs to the concept c_i. The possibility is calculated as:

$$p_{ti}(c_i) = \sum_{t_k \in T_{CO}(t_i)} CO(t_i, t_k)\delta_{rel(c_i, C(t_k))} \tag{9}$$

Among them,

$$d_{rel(c_i, c_j)} = \begin{cases} 0.7 & c_i \text{and } c_j \text{ are co-occurrence concepts} \\ 0.5 & c_i \text{and } c_j \text{ are agents or beneficiaries} \\ 0.2 & c_i \text{and } c_j \text{ are synonymous} \\ 0.1 & \text{other} \end{cases} \tag{10}$$

Among them, $C(x)$ represents the concept of feature word x. Choose the concept with the highest degree of membership as its feature word.

For concept node c, the page set is:

$$D(c) = \{d | d \in D, rel_{d \to c}(d, c) > \lambda\} \tag{11}$$

λ indicates the threshold. According to the above page set, calculate the weight of the key words corresponding to the concept node:

$$w_{t \to c}(t_i, c) = \frac{\sum_{d_j \in D(c)} w_{t \to d}\left(t_i, d_j\right)}{m} \tag{12}$$

In the formula, m is the number of documents in document set $D(c)$. Normalize the $w_{t \to c}(t_i, c)$ value.

Weights can reflect the importance or relevance of a keyword in a specific text or field. Finally, we can calculate the weight of each concept node as the sum of the keyword weights contained in its label. This can obtain an indicator that reflects the importance or correlation of nodes in the conceptual network. By labeling and weight calculation of nodes mentioned above, we can obtain a more comprehensive and accurate conceptual network model, and can more effectively apply sports information resource models when processing large amounts of data and conducting complex reasoning. Based on the above process, the construction of a sports information resource model was completed, and support was provided for the subsequent preprocessing of sports information resources. In the process of building the sports information resource model, this paper labels the information nodes, combines the weights, and optimizes the correlation between concept nodes to build the feature word set of sports information resources. In order to further optimize the resource model, the weights of the keywords corresponding to the concept nodes are further determined. A more perfect sports information resource model is constructed.

2.3 Preprocessing of Sports Information Resources

There is often some noise in sports information resources, which affects the integration of sports information resources, so it is necessary to preprocess and reduce the noise of resources. Based on the model of sports information resources, the representation form of sports information resources is determined. According to the representation form of sports information resources, the denoising process of sports information data is carried out using the principle of wavelet transform [15, 16]. In general, the data after wavelet transform consists of two parts. Are respectively the low frequency part corresponding to the effective signal and the high frequency part corresponding to the noise. Therefore, when using wavelet transform to denoise data, an appropriate threshold should be selected to control the denoising accuracy. The expression of threshold function is as follows:

$$\hat{w}_{i,k} = \begin{cases} w_{j,k}, & \left|w_{j,k}\right| \geq \lambda \\ 0, & \left|w_{j,k}\right| < \lambda' \end{cases} \tag{13}$$

where, $w_{j,k}$ is the wavelet coefficient after threshold processing, $\lambda = \delta_n\sqrt{2\ln N}$ and δ_n are the standard deviation of noise, and N is the number of discrete points.

Let the observation data be $f(t)$, which can be expressed as the combination of the original data $s(t)$ and noise $n(t)$, namely $f(t) = s(t) + n(t)$. The objective of wavelet denoising is to estimate the original signal $s(t)$ and remove the noise $n(t)$, which can be achieved by the following formula:

$$\hat{s}(t) = \sum_{j=1}^{J}\sum_{k=0}^{2^j-1} wthresh\left(\left|W_{j,k}(f(t))\right| - \lambda_j W_{j,k}(f(t))\right) \tag{14}$$

Based on the above content, complete the preprocessing of sports information data, laying the foundation for the subsequent integration of sports information resources.

2.4 Sports Information Resource Integration Based on Fuzzy Clustering Algorithm

In order to improve the integration effect of sports information resources, fuzzy clustering algorithm [17–19] is introduced. Based on the above preprocessed sports information data, the fuzzy clustering algorithm is used to cluster the sports information cluster analysis to obtain the relevant clusters of the data. Then, through the statistical analysis of the clustering results, the clustering algorithm parameters are adjusted accordingly to obtain more accurate and effective clustering results.

Due to the numerous factors that affect the integration of sports information resources, and the significant correlation between each factor and the integration of sports information resources, different aspects of the impact are comprehensively considered, and the representativeness and comparability in the resource classification process are followed to determine the clustering objects and related indicators. Based on the classification results of sports information resources mentioned above, select a clustering center, which is the basis for resource classification.Assume that the given data set is $X = (x_1, x_2, \ldots, x_N)$, that is, the data set contains N samples, where $x_j(j = 1, 2, \ldots, N)$ is a D-dimensional data point. If the data set X is considered to be classified into class C, $V = (v_1, v_2, \ldots, v_c)$ represents the clustering center. According to membership matrix u_{ij}, each data point is assigned to C clustering centers. u_{ij} represents the probability that the j data point belongs to the i clustering center, where $u_{ij} \in [0, 1]$ can reflect the probability of the data point belonging to the corresponding category according to the probability. The membership function also needs to satisfy $\sum_{i=1}^{c} u_{ij} = 1$, that is, the sum of probabilities from a pixel point to each cluster center is always equal to 1. After the membership matrix is determined, each pixel can be divided into the category with the largest corresponding membership value according to the principle of maximum similarity, as shown in Fig. 2.

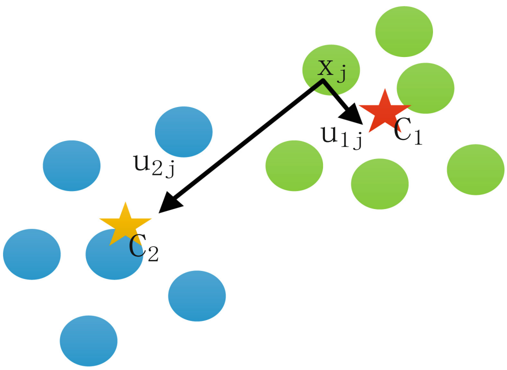


Fig. 2. Schematic diagram of pixel membership division

Among them, the membership value is inversely proportional to the distance, and the distance between data points x_j and c_2 is larger, resulting in smaller membership values. Therefore, x_j is classified into the category where c_1 belongs. The similarity of similar data points in the final clustering results is higher, while the similarity of data points in different classes is lower. Based on the fuzzy C-means clustering algorithm, an alternating iteration scheme is used to minimize the objective function J to obtain the optimal segmentation result. Among them, the objective function represents the weighted sum of Euclidean distances from the data point to the corresponding clustering center.

The definition formula of objective function J is as follows:

$$J(U,V)=\sum_{i=1}^{C}\sum_{j=1}^{N}u_{ij}^{m}\left\|x_j-v_i\right\|^2 \tag{15}$$

where, m is a membership weighted value, which can be set artificially. Generally speaking, m is 2 according to experimental experience. Membership degree must meet: $\sum_{i=1}^{C}u_{ij}=1,\forall j\in\{1,2,\ldots,N\}$. v_i represents the i cluster center, and $\left\|x_j-v_i\right\|^2$ represents the Euclidean distance between data points and the cluster center. A better clustering result is represented by a smaller distance between data points in one category and a larger distance between data points in different categories. So the iterative process is constantly adjusting the membership to minimize the weight and objective function J. Firstly, the membership degree is calculated and then the clustering center is calculated, and then the membership degree is recalculated. Then the membership degree is iterated until the clustering center and the membership matrix do not change too much. The optimal conditions for the implementation of each iteration are as follows:

$$\begin{cases} U^{(t+1)}=\arg\min\limits_{U} J_m^{(FCM)}\{U,V^{(t)}\} \\ \quad U \\ U^{(t+1)}=\arg\min\limits_{V} J_m^{(FCM)}\{U^{(t+1)},V\} \end{cases} \tag{16}$$

Among them, t represents the iteration step. Usually, random initialization is used to obtain $U^{(0)}$ or $V^{(0)}$, and then alternating updates are made.

U and V until convergence. Due to the constraint conditions of the membership function being: $\sum_{i=1}^{C}u_{ij}=1,\forall j\in\{1,2,\ldots,N\}$. Therefore, the Lagrangian factor can be introduced to solve the objective function J:

$$L(u_{ij},v_i)=\sum_{i=1}^{C}\sum_{j=1}^{N}u_{ij}^{m}\left\|x_j-v_i\right\|^2-\lambda\left(\sum_{i=1}^{C}u_{ij}-1\right) \tag{17}$$

where the undetermined coefficient λ is the Lagrange factor. In this way, the minimum problem of objective function J with respect to membership function u_{ij} can be transformed into the minimum problem of function $L(u_{ij},v_i)$. According to the constraints of the membership function [20], it can be obtained:

$$\lambda=m\left[\left\|x_j-v_i\right\|^{\frac{2}{m-1}}\right]^{m-1} \tag{18}$$

By incorporating formula (14) into formula (13), the iterative formula for the membership function u_{ij} can be obtained:

$$u_{ij}=\sum_{l=1}^{C}\left(\frac{\left\|x_j-v_i\right\|}{\left\|x_j-v_l\right\|}\right)^{\frac{-2}{m-1}} \tag{19}$$

The iterative formula of cluster center v_i can be obtained by differentiating v_i:

$$v_i = \frac{\sum_{j=1}^{N} u_{ij}^m x_j}{\sum_{j=1}^{N} u_{ij}^m} \tag{20}$$

The fuzzy C-means clustering algorithm usually initializes the membership matrix U randomly in the process of algorithm iteration, obtains the cluster center matrix V through formula (20), and then updates U, so as to iterate repeatedly to obtain the solution of the objective function J. In summary, the algorithm flowchart of the fuzzy C-means algorithm is shown in Fig. 3.

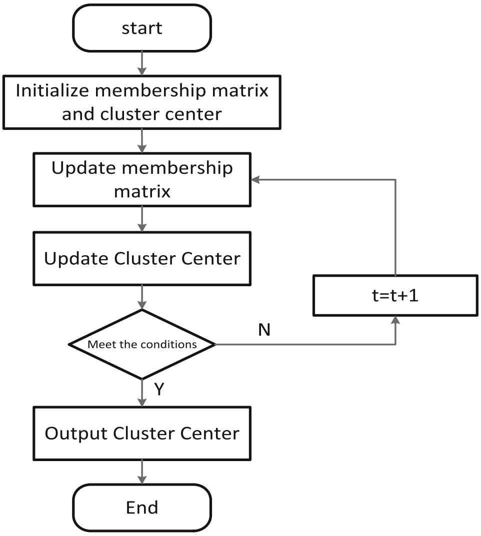


Fig. 3. Algorithm flowchart of fuzzy C-means algorithm

Before iteration, the clustering algorithm initializes two clustering centers, and then updates the membership matrix continuously according to iteration operation, and then updates the clustering centers according to the updated membership function until the iteration stop condition set by the initial parameter is satisfied, the algorithm stops iteration and outputs the clustering result, which is the integration result of sports information resources. It can help to improve the efficiency of sports information resources. At this point, the design of sports information resource integration method based on fuzzy clustering algorithm has been completed, which classifies sports information resources and considers the influence of resource model on resource integration effect. Therefore, the sports information resource model is constructed by introducing weight algorithm, labeling processing and correlation calculation, and considering the existence of certain noise in sports information resources. It seriously affects the integration effect. Therefore, before introducing the fuzzy clustering algorithm, the wavelet transform algorithm is introduced to preprocess the sports information resources and introduce the fuzzy clustering algorithm. Through the combination of these algorithms, the integration method of sports information resources is constructed and the integration effect is improved.

3 Experimental Analysis

3.1 Preparation for Experiment

This article is based on the integration method of sports information resources using fuzzy clustering algorithm for sports information resource integration. To verify the performance of this method, experimental tests were conducted using the integration research of university sports teaching and ethnic traditional sports resources (Method 1) and the optimization and integration of China's ice and snow sports resources based on the H-O-S theory perspective (Method 2) as comparative methods. Retrieve sports information resources through a database for experimental testing. This experiment adopts the sports special database, which has collected all kinds of sports resources since the establishment of People's Sports Publishing House, divided into seven categories: classic textbooks, academic monographs, competition rules and judging laws, sports videos, sports books, sports dictionaries, Olympic channels, and more than 100 series. The library contains more than 600 kinds of teaching materials, 2,269 kinds of academic books, more than 700 kinds of public books, more than 1,000 kinds of videos, more than 90,000 professional entries, and more than 4,000 professional example sentences and phrases. The database content is systematic and thematic. And the database combined with the requirements of the new national curriculum standards, through video, books, monographs, sports dictionaries, competition rules, referee law, and Olympic topics and other forms of all-round display in sports teaching, research and daily application of resources. In order to verify the data security performance and sharing time efficiency of the method proposed in this article, the entire experiment needs to be completed on a powerful server. The server software and hardware parameters are shown in the table below (Table 3).

Table 3. Software and hardware parameters of the server

Parameter	Model number
Server	Inter(R)Core(TM)i7-7700HQ CPU@3.8 GHz
Database	MySQL 5.5
Database management tool	Navicat
Environment	D2RO, D2R Server
Code writing software platform	Python 4.8
Memory	512G
Operating system	Windows10

3.2 Experimental Indicators

The accuracy of sports information resource integration is generally quantitatively evaluated using evaluation indicators such as Precision, Recall, and F1 in the field of information retrieval.

Among them, Precision represents the ratio of the number of relevant resources retrieved to the number of all resources retrieved, expressing the proportion of relevant information content in the search results. The closer it is to 1, the more truly relevant resources in the search results, that is, the more accurate the search results are. The calculation formula is:

$$P = \frac{TP}{(TP + FP)} \tag{21}$$

TP (True Positive) indicates the number of resources retrieved correctly, and FP (False Positive) indicates the number of resources retrieved incorrectly.

Recall refers to the ratio between the number of retrieved resources and the actual number of all relevant resources, that is, the number of retrieved resources. The closer it is to 1, the more truly relevant resources can be found, that is, the more comprehensive the search results. The formula is as follows:

$$\mathrm{Recall} = \mathrm{TP}/(\mathrm{TP} + \mathrm{FN})$$

$$R = \frac{TP}{(TP + FN)} \tag{22}$$

Among them, FN (False Negative) represents the actual number of related resources that exist but cannot be retrieved.

F1 value is the weighted harmonic mean of Precision and Recall, which can quantify the accuracy and integrity of the integration results. The closer to 1, the higher the overall accuracy and recall, and the best comprehensive evaluation effect. Its calculation formula is:

$$F1 = 2 \times P \times R \frac{R}{(P + R)} \tag{23}$$

The F1 value can be regarded as the comprehensive performance of Precision and Recall. The larger it is, the higher the accuracy and completeness of the integrated results will be.

3.3 System Function Test Analysis

Before verifying the performance of the sports information resource integration method based on fuzzy clustering algorithm designed in this paper, it is necessary to ensure that all functions of the method in this paper can operate normally, otherwise the accuracy of the experimental analysis may be affected. Therefore, systematic functional testing and analysis of the method in this paper are carried out, and the analysis results are shown in Table 4.

Table 4. Results of systematic functional test of sports information resource integration method

Serial number	Test item	Operation condition
1	Classification of sports information resources	Normal
2	Information resource label	Normal
3	Multiple semantic feature word mapping	Normal
4	Sports information resources denoising	Normal
5	Clustering of different types of resources	Normal
6	Clustering of resources of the same type	Normal

As can be seen from Table 4, all functions of the method in this paper can operate normally and have the basis for experimental analysis. During the experiment, the operation of each module function of the method will not affect the integration effect of sports information resources.

3.4 Accuracy of Sports Information Resource Integration

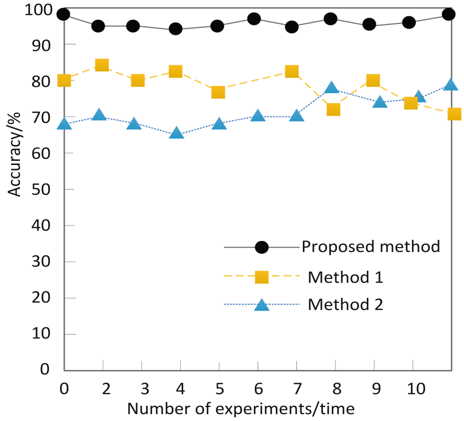
To verify the effectiveness of the method proposed in this article, experimental tests were conducted using the accuracy of information resource integration as the experimental indicator. The test results are shown below.

As shown in Fig. 4, after applying the method proposed in this article, the accuracy, recall, and F1 of sports information resource integration are higher than those of Method 1 and Method 2. Among them, the maximum accuracy of the sports information integration method in this article is 98%, the maximum recall rate is 96%, and the maximum F1 is 0.97. It can be seen that the sports information integration effect of this method is the best.

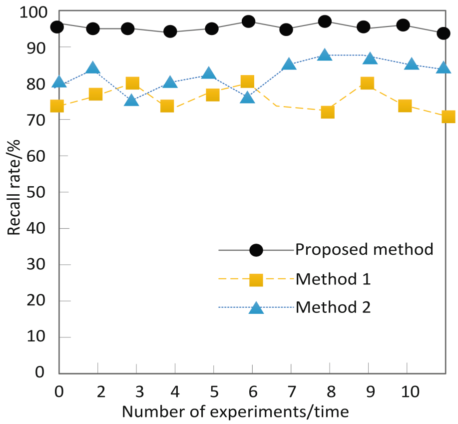
3.5 Efficiency of Sports Information Resource Integration

To further validate the practicality of the method proposed in this article, experimental testing was conducted using the efficiency of information resource integration as the experimental indicator. The test results are shown in Fig. 5.

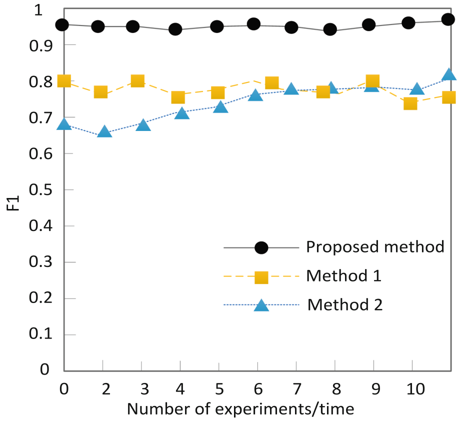
As shown in Fig. 5, the integration efficiency of sports information resources after the application of the proposed method is higher than that of methods 1 and 2. Among them, the longest time of sports information integration of the method in this paper is 3.77 s, the longest time of the method in method 1 is 6.34 s, and the longest time of the method in method 2 is 8.21 s. It can be seen that the method in this paper has the shortest time of sports information integration, the highest efficiency and practicability.



(a)Accuracy comparison result



(b) Comparison results of recall rates



(c) F1 comparison results

Fig. 4. Integration accuracy of sports information resources

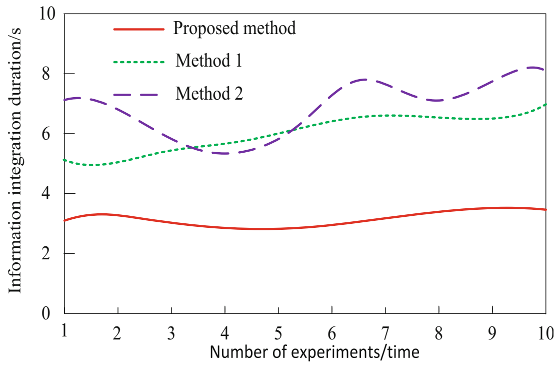


Fig. 5. Integration efficiency of sports information resources

4 Conclusion

With the continuous development and popularization of computer technology and information network technology, sports information resources have now widely covered athletes' personal information, game video, marketing and other aspects of information. However, these different types of information are usually stored in different databases, using different systems and platforms, and cannot be shared or interacted with. Therefore, the integration of various sports information resources is very important to promote the development of sports industry and improve the level and quality of sports. In this regard, this paper studies the integration method of sports information resources based on fuzzy clustering algorithm. The experimental results show that the accuracy rate of sports information integration of the proposed method is 98%, the recall rate is 96%, the F1 maximum is 0.97, and the longest time is 3.77 s, indicating that the proposed method is practical.

References

1. Escamilla-Fajardo, P., Alguacil, M., García-Pascual, F.: Business mode1 adaptation in spanish sports c1ubs abetaording to the perceived context impact on the socia1 cause perfonnance. Sustainability **13**(6), 1–11 (2021)
2. Wei, L., Xing, L.: Sociological analysis of Chinese sports viewers with differences in social capital. Complexity **2021**(6), 1–7 (2021)
3. Subathra, P., Elango, M., Arumugam, S.: Influence of mental training on aggression and sports competition anxiety among volleyball players. Gorteria **34**(1), 377–382 (2021)
4. Ju, L., Huang, L., Tsai, S.B.: Online data migration model and ID3 algorithm in sports competition action data mining application. Wirel. Commun. Mobile Comput. **2021**(7), 1–11 (2021)
5. Wang, J.: Research on the integration path of school and city sports resources based on network platform. Contemp. Sports Technol. **11**(18), 233–235 (2021)
6. Yi, S.: The value evolution and resource integration path of school physical education under the "Double Reduction Policy". Sci. Technol. Stationery Sport. Goods (23), 130–132 (2022)

7. Tian, H.: Resource integration and inheritance innovation of traditional sports culture in local universities-a case study of south lion talent training in Jiaying University. J. Jiaying Univ. **40**(6), 82–87 (2022)
8. Chen, H., Nan, M.: Research on the archive resources integration mode of large-scale sports events from the perspective of digital humanities. Shanxi Arch. (01), 27–39 (2022)
9. Deng, X.: Research on the integration of college physical education teaching and ethnic traditional sports resources. China Adult Educ. (04), 51–53 (2021)
10. Wang, Z., Li, K.: The optimization and integration of snow and ice sports resources in China based on H-O-S theory. China Winter Sports **44**(3), 49–53 (2022)
11. Zheng, C.: The exploration and research of school physical education informationization teaching in DT Era. Contemp. Sports Technol. (6), 130–133 (2021)
12. Jiang, H., Fu, Z.: Research on the construction and application of a school school joint sports resource sharing platform based on regional digitalization. Chin. Sci. Technol. J. Database (Full Text Edition) Educ. Sci. (11), 199–202 (2022)
13. Wang, Y.: Algorithm design of air logistics resource optimal allocation based on fuzzy clustering. Electron. Des. Eng. **30**(15), 20–24 (2022)
14. Wang, Y., Li, X.: Hybrid recommendation algorithm combining wolf colony algorithm and fuzzy clustering. Comput. Eng. Appl. **58**(5), 104–111 (2022)
15. Meng, L., Liu, L., Peng, Y., et al.: A data hiding scheme based on U-Net and wavelet transform. Knowl.-Based Syst. **223**(8), 22–36 (2021)
16. Izat, S., Ruslan, M., Peter, C., et al.: Prediction of gamma ray data from pre-stack seismic reflection partial angle stacks using continuous wavelet transform and convolutional neural network approach. J. Appl. Geophys. **197**(6), 523–533 (2022)
17. Li, K.X., Zhao, J., Liu, T., et al.: A fuzzy spectral clustering algorithm for hyperspectral image classification. IET Image Process. **15**(12), 2810–2817 (2021)
18. Zhu, Y., Han, Y.: Marine environment monitoring based on virtual reality and fuzzy C-means clustering algorithm. Mob. Inf. Syst. **2021**(7), 1–11 (2021)
19. Chen, Z.: Using big data fuzzy K-means clustering and information fusion algorithm in english teaching ability evaluation. Complexity **2021**(5), 1–9 (2021)
20. Wang, X., Hao, L.: Event-triggered fault detection for Takagi-Sugeno fuzzy systems via an improved matching membership function approach. Inf. Sci. **593**(6), 35–48 (2022)