Journal of Information Security Research

DLINE JOURNALS

Print ISSN: 0976-4143
Online ISSN: 0976-4151

JISR 2025: 16 (2)
https://doi.org/10.6025/jisr/2025/16/2/45-52

Improved K-means Algorithm and its Application in University Public Physical Education

Li Huang
School of Physical Education
Hunan University of Arts and Science
415000, Changdu, Hunan. China
152524094@qq.com

ABSTRACT

In the current context, where national sports have established new goals and requirements for university sports, the limitations and one-sidedness of traditional teaching methods have become evident. For the current university public physical education, there is a need for more scientific and effective strategies, which would result in improved teaching efficiency and more engaging teaching content. By introducing the particle swarm algorithm, we can better utilize K-means technology to extract data features and relationships in physical education courses and verify the effectiveness of clustering through information theory. We can also apply this algorithm to university public physical education to improve teaching quality. The application of the improved K-means algorithm can significantly enhance the quality of physical education, providing valuable data support and effectively improving the efficiency of course management, thus better meeting the needs of students.

Keywords: University Public Physical Education, K-means Algorithm, Particle Swarm Algorithm

Received: 10 October 2024, Revised 12 December 2024, Accepted 27 January 2024

1. Introduction

By offering public physical education courses in universities, we can gain a deeper understanding of the fundamental principles of sports and better train our athletes. This helps us know sports better and lays a solid foundation for our future. In response to a series of major university education reforms in China, university public physical education courses are undergoing significant changes, including improvements in theory, curriculum, activities, and skills. The K-means algorithm is an advanced information mining method that effectively integrates various technologies, such as computer science, data analysis, intelligent database systems, and neural networks, to explore potential influencing factors and transform them into modern teaching assessment

data, thereby supporting educational managers' decisions [1]. To promote physical education effectively, we need to establish a comprehensive decision tree to measure all possible influences. In this way, we can formulate specific measures tailored to different situations to enhance teaching quality [2]. For example, in business English teaching, we can use the K-means algorithm to control and analyze candidates' performance, thereby optimizing course content and promoting the continuous development of teaching quality [3]. In this article, we discuss how to utilize the K-means algorithm to optimize university public physical education teaching and better identify the advantages and disadvantages of courses, thereby enhancing the overall quality of the courses. The K-means algorithm is a popular data analysis technique that can extract more information from existing data resources and explore the correlations between them, helping us better understand and master this technology and apply it to university physical education classrooms, effectively improving students' physical fitness, and also allowing for timely adjustments to teaching plans, achieving optimal teaching results [4].

2. Related Work

With the development of technology in the early 21st century, an increasing number of sub-Is have begun to explore more precise evaluation criteria to assess and evaluate educational resources better. Among them, the fuzzy comprehensive evaluation method has been particularly emphasized, as it can effectively assist us in assessing educational resources in the classroom. In recent years, the rapid development of UML and KDD has brought new opportunities for many professional studies. However, despite these advancements, many scholars can only attempt these theories in theory without putting them into practice. Therefore, data mining technology is emerging as a discipline that combines fuzzy mathematics, statistics, machine learning, logical reasoning, and other fields to better meet diverse needs and gain increasing attention in various fields [5].

Currently, in university sports events, numerous meaningful studies focus on the practical operations and training involved. For example, Yue conducted an in-depth analysis of sports training and theoretical training in one paper, proposing a new model for sports training and preparation [6]. Under the guidance of CaoD's "cycle theory," we should explore and construct the implementation of public physical education courses from different perspectives, including organizing course content and formats meticulously based on students' personalized needs and building a comprehensive and well-developed physical education curriculum system [7]. Through interdisciplinary collaboration led by MaJ. M, in-depth exploration of college life and health courses based on psychology has been significantly promoted [8]. By comprehensively utilizing various methods such as clustering algorithms, genetic algorithms, and immune methods, we can more accurately predict future development trends and obtain more precise results. Notably, clustering algorithms can more accurately predict future development trends and outcomes. By employing immune algorithms, we can mimic the mechanisms of the human immune system to monitor and control intracellular environments effectively [9].

With technological advancements, an increasing number of algorithms have been developed, including fuzzy clustering and neural network clustering. Fuzzy clustering algorithms employ a hierarchical classification approach, starting from a lower level, while neural networks utilize a multilevel, progressive clustering algorithm. These algorithms offer several advantages, including higher convergence rates, more efficient computation, and easier access to accurate results. However, fuzzy clustering algorithms still have some evident problems, such as their inability to fully reflect the similarity of data, leading to limitations in clustering accuracy to some extent. In contrast, fuzzy clustering algorithms focus more on the similarity of data and set the sum of membership degrees to be greater than 1, which can better reflect the accuracy of clustering, resist noise interference, and

thus more accurately predict clustering results. This method is particularly suitable for handling samples with noise [10].

In this study, we explored various factors that influence university public physical internships. These factors include students' interests, health status, thinking abilities, and experiences in participating in competitions. Additionally, they include teachers' teaching methods, professional knowledge, and course content. In recent years, with the advancement of technology, an increasing number of researchers have begun using the K-means algorithm to analyze and interpret large datasets more accurately, thereby revealing abundant information and providing a more comprehensive depiction of numerous variables and trends, ultimately offering a clearer understanding of various variables and trends.

3. Improving the K-Means Clustering Algorithm

The success of the k-means clustering algorithm depends on the selection of initial cluster centers, which is crucial. If the initial cluster centers are chosen improperly, the clustering results will be significantly affected, and the expected outcomes may not be achieved. To address this issue, this paper introduces the particle swarm optimization (PSO) algorithm to substantially improve the k-means clustering algorithm [11, 12]. By employing the PSO, a more effective value of '*k*' can be obtained, and the clustering results are validated using the Elbow method, making the experimental results more reasonable and effective. Compared to traditional gradient-based methods, the PSO algorithm can more accurately find the local optimal solution, thereby enhancing the overall performance of the clustering.

3.1 Particle Swarm Optimization (PSO) Algorithm

The development of the Particle Swarm Optimization (PSO) algorithm can be traced back to the foraging behavior of bird flocks. PSO enables the exchange of various relevant data and information among individuals in the group, empowering them to make independent decisions and transform the data and results of other groups into more reasonable patterns to achieve better outcomes [13,14]. By utilizing individual optimal solutions and integrating various solutions, PSO can effectively control the subsequent evolution of individuals, resulting in a more accurate spatial distribution. Based on the fundamental PSO algorithm, a D-dimensional space can be partitioned into several particle swarms, with each particle representing a specific solution. These particles automatically adjust their starting points, endpoints, and movement directions based on given conditions to determine the best results. The following are detailed calculation methods: [Description of the actual PSO calculation methods can be inserted here.

$$v_{id}^{k+1} = w_{id}^{k} + c_1 rand_1() \times \left(p_{id} - x_{id}^{k} \right) \qquad (1)$$

Here, we use several random functions to generate data, including "rand," "(),," and "()." The data generated by these functions include "*oa*," which describes the motion of particle *i* before the k^{th} iteration; "*o*," which represents the inertia weight function; "*c*2" and "*c*2," which are weight factors; "*pa*," which describes the personal best position of particle *i*; and "*pga*," which represents the position of particle *i* before the k^{th} iteration, all within the range [0,1].

3.2 Information Criterion

The concept of criterion was first proposed in thermodynamics in 1948 and has evolved into a more concrete term over time. It not only involves the expected value of a random variable but also a more complex mathematical

model, namely, information. It helps us better understand and grasp information, enabling us to make more effective use of this data. Let X be a discrete random variable with n different parameters, represented as {.*x*, .*xg*, ..., .}, and the probability distribution function of the occurrence of these parameters is expressed by *pi*, *P:*, ..., *p*, and the corresponding information criterion definition expression, where *p* = (*pi, P:*, ..., *p*,).

$$H(X) = H\left(p_1, p_2, \cdots, p_n\right) = -\sum_{i=1}^{n} p_i \log_2 p_i \tag{2}$$

The information criterion has two significant characteristics: first, all its probability values are positive, meaning they are greater than zero; second, it reaches its maximum value only in the case of P2.

3.3 Improved K-means Algorithm

By extracting all data objects into a space, we can construct a system composed of several particles. By applying formula (3), we can extract k particles from these particles and use this information to determine their positions, velocities, and other features.

$$m_j = \frac{1}{n} \sum_{j=1}^{n_j} x_j \tag{3}$$

After the previous processing, we obtain the effective value of *k*. Next, we can utilize the traditional *k*-means clustering algorithm to cluster all sample points, generating *k* cluster points and a corresponding set. The k-means algorithm has a significant flaw: the value of *k* is often influenced by external factors, making it difficult to determine precisely which data attributes can be used, and consequently, the value of *k* is uncertain. The correct value of *k* is a critical factor that affects the experimental results, and an incorrect value of *k* can lead to experimental failure. Therefore, when conducting a clustering analysis, the correct method must be employed to obtain more accurate conclusions. By continuously iterating analysis, *k*-means can effectively extract the local optimal solution, thereby maintaining the stability of the system. See Figure 1 for illustration

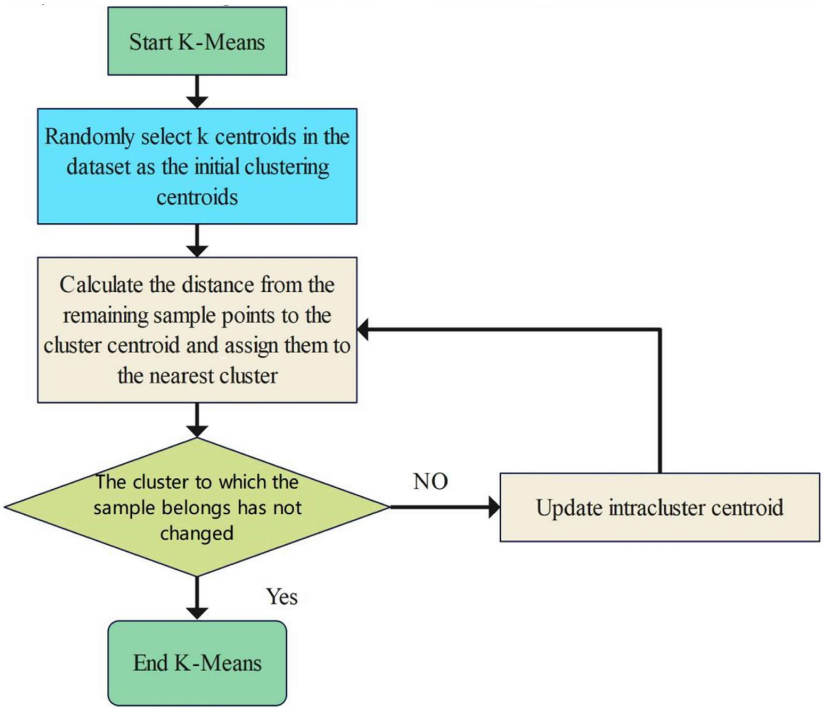

Figure 1. K-means Clustering Algorithm Flowchart

The basic steps of the algorithm are as follows: first, initialize the target points, then classify them based on their distances, and update their mean values at each iteration. Finally, the algorithm stops running when no objects (or a minimum number of objects) are reassigned to different target points. The error sum of squares reaches a local minimum.

4. Algorithm Application and Data Analysis

K-means and Particle Swarm Optimization (PSO) are classification algorithms based on attribute values, but their approaches differ. K-means precisely defines the nearest distance between two points through multiple iterations, whereas PSO requires frequent observation and adjustments to achieve more stable classification results [15, 16]. Therefore, both K-means and PSO exhibit high flexibility; however, the latter is more stable, accurate, and reliable. K-means has a strong element of randomness, which means its iterative results may exhibit certain biases. However, it excels in handling multiple attributes (high dimensionality), making it more flexible than PSO. Hence, in data analysis, its multidimensional advantages can be fully demonstrated.

4.1 Application of Improved K-means Algorithm

By randomly sampling 400 students from 11 universities, we established a database for physical education students, including their personal information such as age, gender, intelligence level, and physical fitness, obtained from various sources [17, 18]. Through this survey, we collected 28 pieces of information regarding students' interest in sports, participation frequency, teacher's skill level, teaching content, teaching methods, and teaching effectiveness. These data will provide essential references for better understanding the students' physical education levels.

4.2 Data Preprocessing

After careful data processing, we transformed the content of 28 unfilled questionnaires into 25 retrievable items, thus constructing a complete data model with multiple features. Through logistic regression analysis of

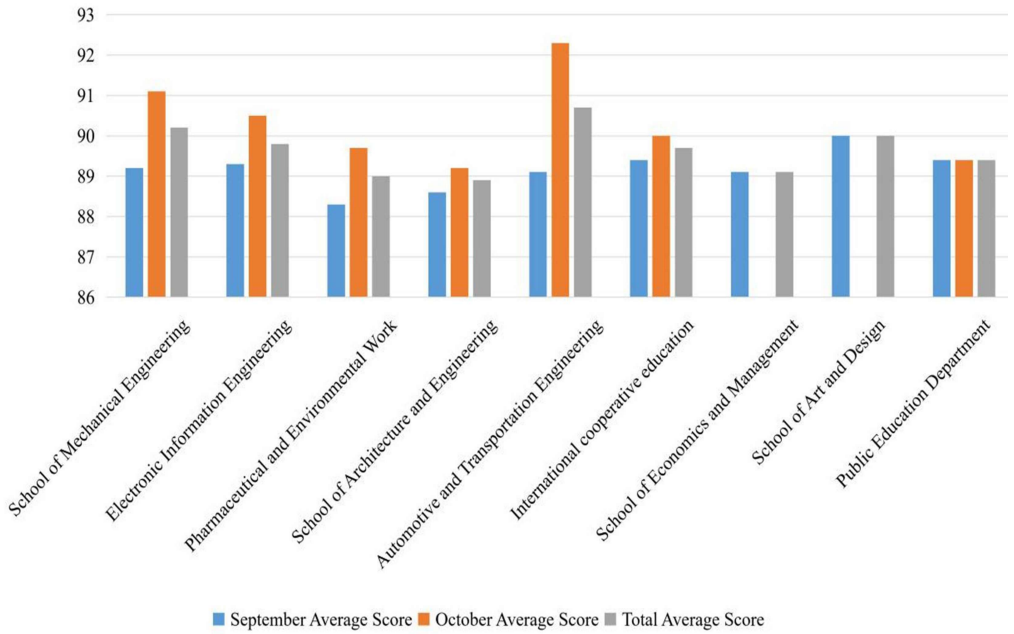

Figure 2. Statistics of Physical Education Teaching Quality Evaluation Data in Various Colleges

32 analytical variables, we found that, considering R = 0.512 and a = 0.05, we could select 8 variables with statistically significant p-values to build a decision tree model for evaluating the effectiveness of university physical education courses. After a meticulous division of the eight analytical variables, we selected the optimal model, as shown in Figure 2.

Based on Figure 2, we can see how the physical education teaching quality evaluation data vary among different colleges. Out of 347 training set tuples, 85% of valid data were randomly selected from each set and used as training samples (295) and testing samples (52) to construct a decision tree model reflecting the teaching quality of university physical education dance courses. The C5 decision tree in SPSS Clementine 12.0 can be used for model fitting, where each branch of the decision tree should have a specific feature to ensure that the feature provides a higher information gain, surpassing the average level of other features.

4.3 Analysis of Improved K-means Algorithm Results

The basic structure of the K-means algorithm reveals that "student participation frequency" and "student competition frequency" have a significant influence on students' performance in public physical education courses, with "student participation frequency" having a more substantial impact, contributing to improved students' competitiveness and overall performance. The K-means algorithm can extract valuable information from a large amount of sample data and transform it into results that can be better predicted and evaluated, thereby assisting teachers and parents in better understanding children's learning and enhancing the quality and effectiveness of teaching. We can classify the closest value by generating reference values and then sequentially checking all unmarked values. After one round of iteration, we can randomly select reference points from each class. As the number of iterations increases, the variation in classification will gradually decrease, eventually reaching a stable state where the classification itself does not change. As shown in Figure 3, this program demonstrates relatively satisfactory efficiency for K-means, with a fast-processing speed for the information. Although there were initial issues with the space stack overflow, they were quickly resolved by limiting the number of classifications processed at once. Overall, K-means exhibits a fast speed but requires higher space requirements, which have been addressed. Particle Swarm Optimization has lower efficiency due to its algorithmic characteristics, resulting in repetitive processing. However, optimization efforts have significantly improved its speed, reducing the processing time to around 2 minutes when handling 10,000 significant data points, which is much faster than the unoptimized version, which would take at least 30 minutes. Although this algorithm has lower efficiency and higher space requirements, optimizations have mitigated these challenges.

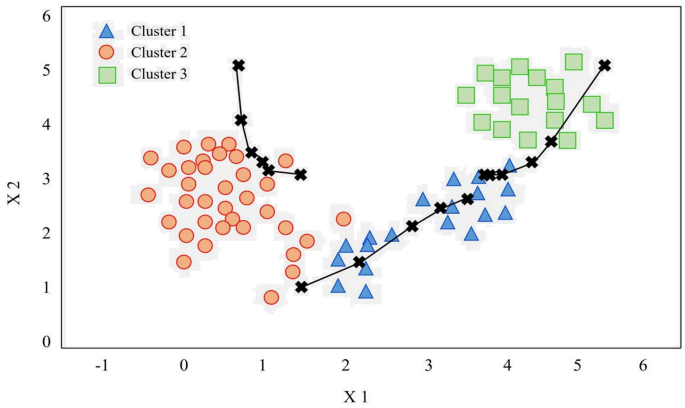

Figure 3. K-means Classification Statistics (canteen and money attributes)

5. Conclusion

The importance of physical education courses has been widely recognized, as they lay the foundation for our future development. We need to explore more effective ways to manage and control our courses in a targeted manner. To achieve this, various measures, including the use of Particle Swarm Optimization and other techniques, should be taken to optimize our *K*-means algorithm and refine it for greater effectiveness. By doing so, we can better manage our courses and establish a solid foundation for our future development. Through an in-depth analysis of the K-means algorithm, we find that it exhibits excellent performance, not only effectively helping us explore and discover unknown information but also bringing great convenience and timeliness to our classroom design and practice activities, making our classroom activities more engaging and creating favorable conditions for our learning and growth.

References

[1] He, G., Lu, X. L. (2021). An improved QPSO algorithm and its application in fuzzy portfolio model with constraints. *Soft Computing, 25*(12), 7695–7706.

[2] Zhao, Y., Chen, J., Li, W. (2022). Improved whale algorithm and its application in cobot excitation trajectory optimization. *International Journal of Intelligent Robotics and Applications, 6*(4), 615–624.

[3] Xu, K., Ran, S. (2023). An improved order-encoded quantum image representation model and its application. *International Journal of Quantum Information, 21*(4).

[4] Yiping, W., Buqing, S., Jianjun, W., et al. (2021). An improved multi-view collaborative fuzzy C-means clustering algorithm and its application in overseas oil and gas exploration. *Journal of Petroleum Science and Engineering, 197*, 108093.

[5] Chen, Z. Y. (2022). A computational intelligence hybrid algorithm based on population evolutionary and neural network learning for the crude oil spot price prediction. *International Journal of Computational Intelligence Systems, 15*(1), 68.

[6] Yue, Y., You, H., Wang, S., et al. (2021). Improved whale optimization algorithm and its application in heterogeneous wireless sensor networks. *International Journal of Distributed Sensor Networks, 17*(5), 277–323.

[7] Cao, D., Chen, Y., Chen, J., et al. (2021). An improved algorithm for the maximal information coefficient and its application. *Royal Society Open Science, 8*(2), 201424.

[8] Ma, J. M., Shan, S. J., Su, R. J., et al. (2021). An improved island algorithm and its application in model optimization of micro soft robot. *Journal of Physics: Conference Series, 1944*(1), 012036.

[9] Wu, H., Zhang, X., Song, L., et al. (2022). A hybrid improved BRO algorithm and its application in inverse kinematics of 7R 6DOF robot. *Advances in Mechanical Engineering, 14*(3), 259–267.

[10] Yiping, W., Buqing, S., Jianjun, W., et al. (2021). An improved multi-view collaborative fuzzy C-means clustering algorithm and its application in overseas oil and gas exploration. *Journal of Petroleum Science and Engineering, 197*, 108093.

[11] Abo Elyousr, F. K., Sharaf, A. M., Darwish, M. M. F., et al. (2022). Optimal scheduling of DG and EV parking lots simultaneously with demand response based on self adjusted PSO and K means clustering. *Energy Science & Engineering, 10*(10), 4025–4043.

[12] Zhang, H., Peng, Q. (2022). PSO and K-means-based semantic segmentation toward agricultural products. *Future Generation Computer Systems, 126*, 82–87.

[13] Feng, Z., Niu, W., Zhang, R., et al. (2019). Operation rule derivation of hydropower reservoir by K-means clustering method and extreme learning machine based on particle swarm optimization. *Journal of Hydrology, 576*, 229–238.

[14] Sheikhhosseini, Z., Mirzaei, N., Heidari, R., et al. (2021). Delineation of potential seismic sources using weighted K-means cluster analysis and particle swarm optimization (PSO). *Acta Geophysica, 69*, 2161–2172.

[15] Yang, Y., Liao, Q., Wang, J., et al. (2022). Application of multi-objective particle swarm optimization based on short-term memory and K-means clustering in multi-modal multi-objective optimization. *Engineering Applications of Artificial Intelligence, 112*, 104866.

[16] Jamali-Dinan, S. S., Soltanian-Zadeh, H., Bowyer, S. M., et al. (2020). A combination of particle swarm optimization and Minkowski weighted K-means clustering: Application in lateralization of temporal lobe epilepsy. *Brain Topography, 33*, 519–532.

[17] Bai, L., Song, Z., Bao, H., et al. (2021). K-means clustering based on improved quantum particle swarm optimization algorithm. In *2021 13th International Conference on Advanced Computational Intelligence (ICACI)* (pp. 140–145). IEEE.

[18] Paul, S., De, S., Dey, S. (2020). A novel approach of data clustering using an improved particle swarm optimization-based K–means clustering algorithm. In *2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)* (pp. 1–6). IEEE.