



Optimizing the Students' Exercise Behavior Analysis: A Clustering Model

Yinhui Hao, Chunyan Zhao

School Of Accounting, Tongling University, Tongling

Anhui, 244000. China

haoyinhui125@126.com

ABSTRACT

This paper proposes an optimized Ant Colony Algorithm (ACA) model to cluster and mine the physical exercise behavior characteristics of college students. Recognizing that students' unstructured exercise habits often lead to inefficacy or injury, the study aims to provide more accurate behavioral analysis. The model first constructs a behavior representation by tracking the centroid movement of students' bodies in video sequences to generate feature vectors. It then employs an ACA, inspired by ants' foraging behavior, to cluster these vectors efficiently. The algorithm is optimized to address common ACA issues like slow convergence and stagnation by improving pheromone updating and path selection. Experimental results on standard and custom datasets show the model achieves higher F-measure values, greater stability, and faster convergence compared to other algorithms, effectively avoiding local optima for superior clustering accuracy.

Keywords: Ant Colony Optimization (ACO), Exercise Behavior Clustering, College Students, Physical Activity Analysis, Behavior Feature Mining, Computer Vision, Centroid Motion Tracking, Algorithm Convergence, F-measure Evaluation, Injury Prevention in Exercise

Received: 19 May 2025, Revised 3 August 2025, Accepted 17 August 2025

Copyright: with Authors

1. Introduction

Physical fitness is essential for the ongoing development of both individuals and society, and fostering physical fitness has always been a significant aspect of modern education. In contrast to high school students, college students enjoy greater freedom in managing their time, with teachers primarily serving in a guiding capacity. As a result, without obligatory exercise requirements and strong oversight, numerous college students exhibit considerable disparities in how they allocate time between academics and physical activity [1]. The duration and intensity of their exercise often fail to meet the physical fitness development needs of college students. Surveys indicate that while most college students acknowledge that physical activity can enhance their fitness, they struggle to scientifically regulate their exercise levels due to a lack of structured understanding and training during their

work outs. Concurrently, many college students engage in incorrect exercise practices, which can lead to potential injuries and prevent them from reaching the objective of scientific exercise.

To enhance the accuracy of college students' physical activity habits and provide them with scientific feedback throughout their exercise sessions, information technology, big data, and artificial intelligence have been increasingly implemented in the study of physical exercise behaviors [2]. Systems for monitoring and providing feedback on college students' physical activities have been established, enabling students to adjust their exercise habits based on the input received promptly and ensuring the correctness of their athletic actions. Given that the exercise behaviours of college students are dynamic, recognising these behaviours presents significant challenges, and various recognition techniques exhibit substantial differences in accuracy. Furthermore, adapting to the identification of different characteristics of exercise behavior is difficult, resulting in a limited range of applications. Additionally, some algorithms used in the clustering analysis of behaviors based on their characteristics face issues such as long processing times, significant inaccuracies, and a high reliance on manually set parameters, leading to insufficient precision in the results obtained [3]. Therefore, this paper incorporates the ant colony algorithm into the model for extracting characteristics of college students' physical exercise behaviors, aiming to optimize the clustering path through the ant colony algorithm to enhance the effectiveness and efficiency of behavior feature clustering. The goal is to better achieve the aims of clustering analysis and exploration of college students' physical exercise behavior characteristics.

2. Identification and Analysis of Physical Exercise Behavior

As the focus on physical activity grows, many researchers have identified problems such as improper techniques, excessive training, and injuries occurring during exercise. These issues not only fail to fulfill the aim of enhancing physical fitness but also pose risks to both physical and mental well being [4]. Furthermore, college students demonstrate two contrasting trends in physical activity. Some students, influenced by age, personality, and other factors, are inclined towards adventurous sports, while others prefer to engage in minimal or no physical activity. Relevant surveys reveal that many college students have reported sustaining injuries during physical activity, primarily due to a lack of professional guidance and scientific knowledge [5]. In addressing this issue, some researchers have employed digital and mobile technologies to capture the actions of individuals exercising and subsequently offer guidance and corrections from professionals based on the recorded footage [6]. Nevertheless, this technique is hindered by a time lag, as exercisers are unable to adjust their movements promptly. To tackle this challenge, certain scholars have developed a physical exercise behavior error detection system leveraging neural networks, specifically using convolutional neural networks to identify and compare exercise movements and provide immediate feedback to participants [7]. However, this model exhibits relatively high computational complexity and lacks application stability, indicating the need for further enhancements. Other academics have applied computer vision technology to recognize the actions of physical exercisers and deliver timely feedback through intelligent algorithms [8]. Furthermore, some researchers have created aggregation behavior recognition models for team sports, employing the ant colony algorithm to identify and assess group sports behavior and provide relevant feedback based on established criteria [9]. At present, there remains a considerable amount of error in detecting physical exercise behavior, especially concerning dynamic movements, leading to relatively low accuracy. Continued optimization and investigation in this area are essential.

3. Ant Colony Algorithm-Based Clustering Mining Model for College Students' Physical Exercise Behavior

3.1 Construction of the College Students' Physical Exercise Behavior Model

Given that the human body demonstrates significant flexibility and undergoes variations in position and velocity across different directions during movement [10], to more effectively analyze the changes in college students' body movements during physical activity, their bodies can be viewed as a cohesive unit. By examining the variations in the centroid position and the velocities produced in various directions, it is possible to significantly reflect the characteristics of college students' exercise behavior. Assuming the college student's body acts as the centroid, it will experience movements in both horizontal and vertical planes during exercise. Concurrently, the movements in various directions can be broken down into horizontal and vertical motion components according to their respective speeds. Utilizing Formula (1), the corresponding binary image description matrix can be generated:

$$h_{m,n} = \sum_{x=1}^i \sum_{y=1}^j x^m x^n f(x, y) \quad (1)$$

Wherein, the description of the order moment is h , and the corresponding numbers are m and n , respectively. The corresponding coordinates of the different order moment numbers on the abscissa of m and ordinate are x , and the description of the Binary image is $f(x, y)$.

According to the above formula, we can analyze the change of the centroid position of college students in their physical exercise behavior. If the centroid coordinate in the Binary image with the sequence number G is expressed as (x, y) , we can describe the position change of the centroid in the horizontal and vertical directions in the next image through the corresponding calculation formula, as shown in Formula (2):

$$\begin{cases} K_s = |x_{g+1} - x_g| \\ K_c = |y_{g+1} - y_g| \end{cases} \quad (2)$$

When collecting images, there is a time interval between each frame of video and every other frame of image. If time interval is considered, the velocity of the centroid motion in the horizontal and vertical directions is described as formula (3):

$$\begin{cases} v_s = \frac{25K_s}{2} \\ v_c = \frac{25K_c}{2} \end{cases} \quad (3)$$

Following the aforementioned expressions, images of college students engaged in physical exercise must be captured from the initial video sequence and then undergo real time normalization processing to create the corresponding image sequence dataset. Subsequently, the individual images can be processed in binary mode to extract relevant local histogram features. Additionally, based on the decomposition of the centroid motion of college students in both horizontal and vertical directions, corresponding sets of feature vectors will be produced.

3.2 Ant Colony Algorithm-Based Behavior Feature Clustering and Mining Model

The ant colony algorithm is inspired by the natural behavior of ants foraging for food. As they search for the same food source, ants deposit pheromones along the routes they traverse, and these pheromone levels gradually diminish over time [11, 12]. This implies that routes with greater distances tend to have lower pheromone concentrations, which decreases the likelihood of other ants opting for those routes. When a higher number of ants select the same route, the pheromone concentration on that path increases continuously, usually indicating

the shortest route to the food source. Leveraging this concept, if the physical exercise behavior feature vectors of college students are regarded as targets for the ant colony to locate, and the search area is segmented into a two-dimensional grid with each distinct grid housing a single feature vector, the ant colony will seek the target feature vectors according to specific objectives. As ants randomly navigate the two dimensional grid and come across a feature vector representing physical exercise behavior, they will assess the likelihood of pursuing that target based on the similarity between the target and the neighboring feature vectors in the grid [13, 14]. If the probability falls short of the set criteria, the ant will randomly move to another grid. To prevent certain feature vectors from being overlooked, this study establishes a probability for both the ant colony and the movement of feature vectors within the grid. When ants transporting the feature vectors of college students' physical exercise behaviors encounter vacant grids or grids where the feature vectors they carry show high similarity to neighboring feature vectors, they can place the feature vectors based on a probability that correlates with the density of the feature vectors. The likelihood of identifying the feature vector is positively associated with its density. If the determined density does not satisfy the criteria, the ant will continue to carry the feature vector. The similarity between two feature vectors can be quantified using Euclidean distance, where a result of 0 indicates a high degree of similarity, while a result of 1 signifies a low degree of similarity. Utilizing the calculated Euclidean distances, a two-dimensional consistency matrix can be constructed, from which local density can be derived, as illustrated in Formula (4):

$$f(w_m) = \frac{\sum w_m (l + \frac{d(w_m, w_n)}{2})}{2} \quad (4)$$

Among them, the neighborhood edge length of lattice r is 1 and the area is 1,1.

Assuming that the probability of the feature vector of college students' physical exercise behavior during ant movement is F , and the corresponding probability of dropping it is $F(W)$, the calculation formula is shown in (5):

$$\begin{cases} F_p(w_m) = \left(\frac{A_1}{A_1 + f(w_m)} \right)^2 \\ F_d(w_n) = \begin{cases} 2f(w_m) & f(w_m) < A_2 \\ 1 & f(w_m) \geq A_2 \end{cases} \end{cases} \quad (5)$$

The threshold constant is represented as A .

Given that the practical application of ant colony optimization algorithms often encounters challenges, including the high complexity of data processing during the algorithm's initial phase, suboptimal convergence efficiency, and a significant chance of stagnation during the solution process, which can adversely affect the clustering outcomes of college students' physical exercise behavior feature vectors, this paper addresses the optimization of the issues associated with ant colony optimization algorithms [15, 16] The optimization process is primarily focused on two key areas: the method for updating pheromones and the mode of selecting paths by ants. To mitigate the issue of potentially overlooking the optimal path during the initial search phase of the ant colony, a strategy involving random perturbation is implemented in the pheromone update process. Specifically, when an ant classifies any feature vector into a category, the corresponding pheromone concentration on that path will increase, leaving an identifiable mark. Consequently, other ants will be

influenced by the pheromone laid down by the first ant, leading to an increase in pheromone levels on that specific path [17, 18]. As various ants in the colony continuously seek the best path, the pheromone levels on different routes remain in a state of flux. Once all ants have completed their searches, the increment in pheromone levels across different paths is assessed, allowing for the determination of the optimal path selection range based on these findings.

4. Ant Colony Algorithm Based College Students' Physical Exercise Behavior Feature Clustering and Mining Model

The efficiency of clustering and mining algorithm models is essential for their real world applicability. Hence, this study compared two alternative clustering algorithms with the optimized Ant Colony Algorithm driven behavior feature clustering and mining model. To effectively assess the adaptability of these clustering algorithm models across various datasets, two datasets namely Iris and Robotnavigation were utilized for the experiments. The results from the experiments are illustrated in Figure 1. The two datasets employed in the experiments include Iris and Robot navigation, with the latter containing a higher number of samples than the former. As seen in Figure 1 (a), the F -measure values of the two alternative clustering algorithm models are relatively comparable, but there is a significant difference when compared to the F -measure value of the algorithm model presented in this study. Moreover, the overall fluctuation range appears to be greater for the other two models. Conversely, the F -measure value of this study's algorithm model exhibits a comparatively smaller fluctuation range, signifying greater stability in the model. In Figure 1 (b), the data indicates a certain decline in the F -measure values of the other two clustering algorithm models, while the F -measure value of this study's algorithm model remains largely unchanged compared to Figure 1 (a). This consistency is primarily due to the increase in sample size within this dataset, which has led to a rise in the error rates of the other two clustering

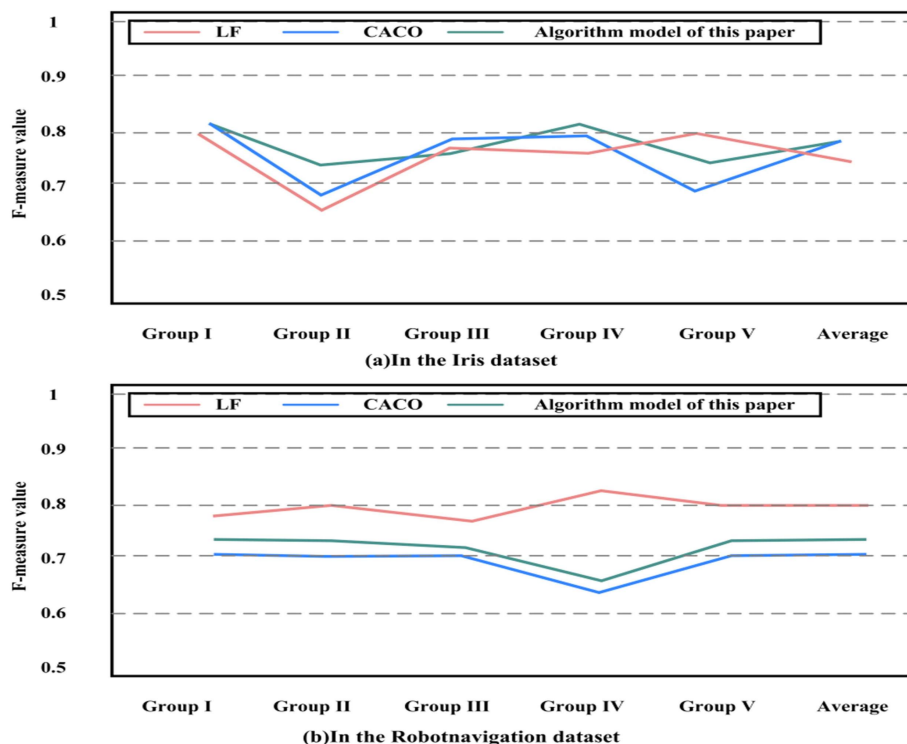


Figure 1. Comparison of F -measure values of three algorithms in two different datasets

algorithm models, causing a noticeable drop in their F -measure values. In contrast, the clustering algorithm model presented in this study has a well defined purpose, maintaining its error rate fluctuation even as sample data increases. Therefore, the F -measure value of this study's model stays stable across both datasets. This suggests that, relative to the other two algorithm models, this study's algorithm model demonstrates superior reliability and stability. Additionally, when considering the results from both datasets, it is evident that while this study's algorithm model does not exhibit a significant advantage in F -measure values with smaller datasets, it converges more rapidly than the other two algorithms.

To enhance the practical validation of the effectiveness of the Ant Colony Algorithm based model for clustering and mining physical exercise behavior features among college students, this research performed separate clustering analyses for common activities and physical movements in college students' exercise routines. The findings are illustrated in Figure 2 and Figure 3. In Figure 2, the clustering F -measure values for all common

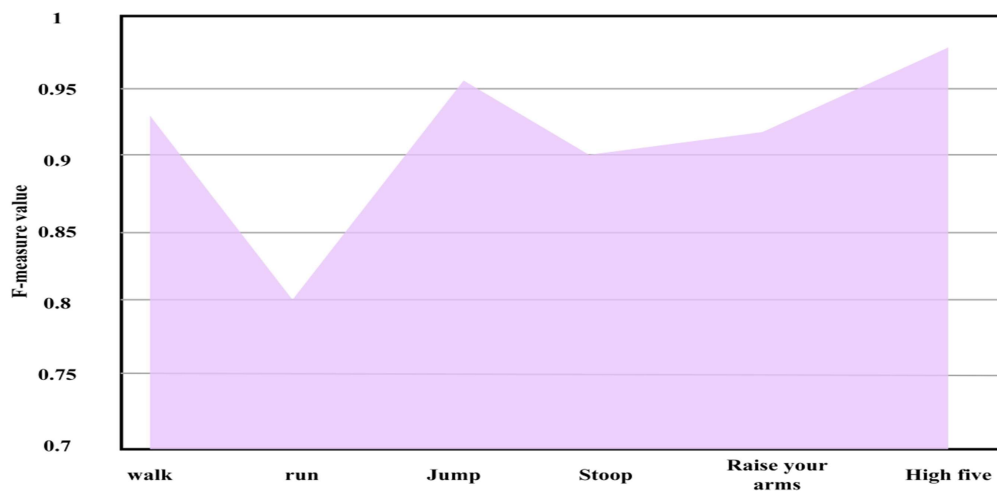


Figure 2. Clustering effect of common physical exercise behaviors in college students

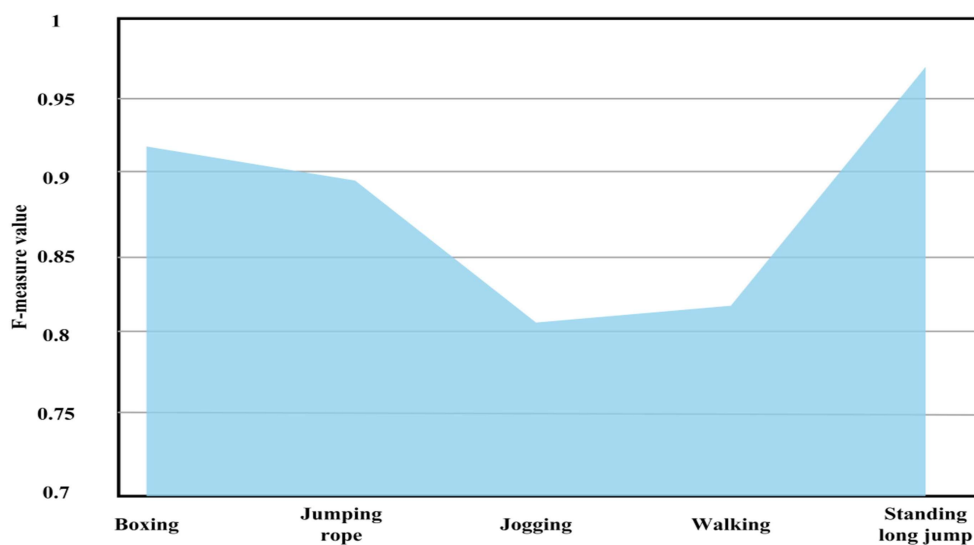


Figure 3. Clustering effect of physical exercise movements in college students

physical exercise behaviors are slightly above 0.8, while the F -measure values for other activities do not drop below 0.9. In Figure 3, the F -measure values for slow running and brisk walking, among other physical exercise behaviors, are just above 0.8, whereas the F -measure values for other activities exceed 0.9. This is primarily due to the high similarity between running and walking behaviors in physical exercise, which leads to a relatively lower recognition accuracy. Nonetheless, the overall clustering performance of this study's algorithm model is commendable, successfully avoiding stagnation by not getting trapped in local optimal solutions. It also enhances the accuracy of recognizing and clustering college students' physical exercise behaviors, yielding improved clustering outcomes.

In summary, the Ant Colony Algorithm based model for clustering and mining physical exercise behavior features in college students can adapt to various scenarios and datasets. It demonstrates a rapid convergence rate, with the error rate remaining consistent regardless of the sample dataset size, showcasing strong stability and reliability. Additionally, it effectively circumvents the issue of becoming trapped in local optima, thereby enhancing the accuracy of behavior recognition and clustering, which leads to superior clustering results in practical implementations.

5. Conclusion

Physical exercise among college students is a crucial aspect of their personal development and a vital foundation for achieving self worth and self improvement as future professionals. However, many college students struggle to make informed decisions regarding their exercise behaviors and manage exercise intensity due to a lack of systematic and professional knowledge and skills in physical fitness. To address this, this study developed a feature extraction model for college students' physical exercise behavior and integrated the Ant Colony Algorithm for behavior clustering. To achieve optimal clustering results, enhancements were implemented on the ant colony algorithm concerning pheromone concentration updates and ant path selection. The experimental outcomes revealed that the Ant Colony Algorithm based model for college students' physical exercise behavior feature clustering and mining demonstrated effective and consistent clustering results across various datasets. Its accuracy remains unaffected by an increase in the number of dataset samples, indicating remarkable stability and reliability. Overall, this algorithm model is capable of achieving commendable clustering results for diverse physical exercise behaviors, although additional optimization is necessary for the clustering of running and walking activities.

References

- [1] Zhang, H., Ye, Y., Cai, X., et al. (2020). Efficient action recognition algorithm based on human joint points. *Computer Engineering and Design*, 41(11), 176-182.
- [2] Chen, S., Ma, Q., Zhao, J., et al. (2019). Research on the method of action detection and action recognition based on upper extremity motor scores. *Chinese Journal of Rehabilitation Medicine*, 34(6), 707-710.
- [3] Salazar, M. O. L., Deneubourg, J. L., Sempo, G. (2013). Information cascade ruling the fleeing behaviour of a gregarious insect. *Animal Behaviour*, 85(6), 1271-1285.

- [4] Zhang, X. Y., Duan, H. B., Luo, Q. N. (2014). Levenberg-Marquardt based artificial physics method for mobile robot oscillation alleviation. *Science China Physics, Mechanics & Astronomy*, 57, 1771-1777.
- [5] Gao, S., Wang, Y., Cheng, J., et al. (2016). Ant colony optimization with clustering for solving the dynamic location routing problem. *Applied Mathematics & Computation*, 285, 149-173.
- [6] Ji, J., Song, X., Liu, C., et al. (2013). Ant colony clustering with fitness perception and pheromone diffusion for community detection in complex networks. *Physica A: Statistical Mechanics & Its Applications*, 392(15), 3260-3272.
- [7] Ýnkaya, T., Kayalýgil, S., Özdemirel, N. E. (2015). Ant colony optimization based clustering methodology. *Applied Soft Computing*, 28, 301-311.
- [8] Moradi, P., Rostami, M. (2015). Integration of graph clustering with ant colony optimization for feature selection. *Knowledge-Based Systems*, 84, 144-161.
- [9] Hong, Y., Chen, L., Mo, L. (2019). Optimization of cluster resource indexing of the Internet of Things based on improved ant colony algorithm. *Cluster Computing*, 22, 7379-7387.
- [10] Kan, Z., Yucelen, T., Doucette, E., et al. (2017). A finite-time consensus framework over time-varying graph topologies with temporal constraints. *Journal of Dynamic Systems, Measurement, and Control*, 139(7), 071012.
- [11] Paniri, M., Dowlatshahi, M. B., Nezamabadi-Pour, H. (2020). MLACO: A multi-label feature selection algorithm based on ant colony optimization. *Knowledge-Based Systems*, 192, 105285.
- [12] Azaryuon, K., Fakhar, B. (2013). A novel document clustering algorithm based on ant colony optimization algorithm. *Journal of Mathematics and Computer Science*, 7, 171-180.
- [13] AlFarraj, O., AlZubi, A., Tolba, A. (2019). Optimized feature selection algorithm based on fireflies with gravitational ant colony algorithm for big data predictive analytics. *Neural Computing and Applications*, 31, 1391-1403.
- [14] Yang, Y., Ni, X., Wang, H., et al. (2012). Parallel implementation of ant-based clustering algorithm based on Hadoop. In *Proceedings of the International Conference in Swarm Intelligence* (pp. 190-197). Springer Berlin Heidelberg.
- [15] Han, X. H., Quan, L., Xiong, X. Y., et al. (2017). A novel data clustering algorithm based on modified gravitational search algorithm. *Engineering Applications of Artificial Intelligence*, 61, 1-7.
- [16] Hameurlaine, M., Moussaoui, A., Cherroun, H. (2012). AntMeans: A new hybrid algorithm based on ant colonies for complex data mining. *International Journal of Computer Applications*, 975, 8887.
- [17] Shi, Q., Liu, B., Guan, Q., et al. (2020). A genetic ant colony algorithm-based driving cycle generation

approach for testing driving range of battery electric vehicle. *Advances in Mechanical Engineering*, 12(1), 1687814019901054.

[18] Agbehadji, I. E., Millham, R. C., Abayomi, A., et al. (2021). Clustering algorithm based on nature-inspired approach for energy optimization in heterogeneous wireless sensor network. *Applied Soft Computing*, 104, 107171.