Abstract

The integration of Fuzzy C-Means (FCM) clustering with the Entropy-Based Smarter Adaptive Population Reduction (E-SAPR) Particle Swarm Optimization (PSO) algorithm introduces a powerful and innovative hybrid approach to address the complexities of clustering and optimization in modern data-driven applications. Clustering, a fundamental technique in unsupervised machine learning, involves grouping data into meaningful categories based on similarity. However, traditional clustering methods such as k-means and even standalone FCM often fall short when dealing with overlapping clusters, high-dimensional datasets, and noisy data. These limitations can lead to poor cluster assignments, reduced accuracy, and high computational costs.

Similarly, classical PSO algorithms, while widely used for global optimization due to their simplicity and efficiency, suffer from key challenges including premature convergence, stagnation, and high computational demand, especially when a large swarm size is used. These issues make them suboptimal for real-world clustering tasks where both exploration (global search) and exploitation (local search) need to be balanced efficiently.

The E-SAPR PSO algorithm addresses these inherent challenges of PSO by introducing a dynamic mechanism for population control that relies on entropy—a measure of randomness or diversity within the swarm. Unlike conventional PSO that maintains a fixed swarm size throughout the optimization process, E-SAPR adaptively reduces the swarm population based on entropy values, thereby preserving diversity early in the search process and focusing computational resources as convergence progresses. This entropy-driven approach ensures that the swarm does not lose diversity prematurely and avoids getting trapped in local optima, a common problem in traditional PSO implementations.

In the proposed hybrid model, FCM is used for soft clustering, where data points can belong to multiple clusters with varying degrees of membership. This soft assignment is particularly beneficial in handling overlapping data and ambiguous cluster boundaries, which are often present in real-world datasets. The E-SAPR PSO is then employed to optimize the cluster centroids produced by FCM, using the entropy-guided mechanism to adaptively refine solutions and improve convergence speed and accuracy.

The core strength of this hybrid algorithm lies in its ability to combine the interpretability and flexibility of FCM with the adaptive, exploration-exploitation balancing capabilities of E-SAPR PSO. This combination allows for effective handling

of complex, noisy, and high-dimensional datasets, significantly improving clustering performance.

To evaluate the effectiveness of the hybrid FCM + E-SAPR PSO approach, extensive experiments were conducted on synthetic and real-world datasets of varying sizes and complexities. Performance metrics such as the Xie-Beni (XB) index and the Davies-Bouldin (DB) index were used to measure clustering quality. The XB index assesses the compactness and separation of clusters, while the DB index evaluates intra-cluster similarity and inter-cluster distance. In both metrics, lower values indicate better clustering results. The experimental results revealed that the proposed hybrid algorithm consistently outperformed traditional FCM and standalone PSO in terms of both clustering accuracy and computational efficiency.

In addition to improved clustering quality, the hybrid algorithm also demonstrated superior scalability and adaptability. As the size and dimensionality of the datasets increased, the entropy-based adaptive reduction of the swarm allowed the algorithm to maintain performance without a proportional increase in computation time or memory usage. This makes the approach suitable for large-scale data processing tasks and real-time applications.

The broader implications of this research are significant. By introducing entropy-based population adaptation into clustering optimization, the proposed algorithm opens new avenues for efficient, scalable, and intelligent clustering techniques. The methodology is well-suited for applications in domains such as bioinformatics (e.g., gene expression data clustering), image processing (e.g., segmentation and object recognition), anomaly detection (e.g., fraud detection and cybersecurity), and medical diagnostics. In all these areas, accurate and efficient clustering is essential for extracting actionable insights from complex datasets.

In conclusion, the hybrid FCM + E-SAPR PSO algorithm presents a robust and versatile solution for modern clustering challenges. It not only advances the state-of-the-art in clustering optimization but also demonstrates the practical benefits of integrating entropy-driven adaptive strategies with soft clustering techniques. This work sets the stage for future research into hybrid intelligent systems that combine the strengths of multiple algorithms to achieve higher efficiency, adaptability, and performance in data-intensive environments.