

Peer Review

Review of: "Fixed-Sized Clusters k-Means"

Yugo Nakayama¹

1. Nissan (Japan), Yokohama, Japan

This paper presents an interesting approach to k-means clustering with fixed-size constraints using the Hungarian algorithm. While the theoretical foundation is sound, several areas require additional work before the manuscript can be considered for publication.

Insufficient Empirical Validation of Computational Complexity

The authors state that their algorithm has a time complexity of $O(n^3)$ due to the Hungarian algorithm implementation in the assignment phase. However, the paper lacks concrete empirical measurements of execution time and memory usage for datasets at the claimed upper limit of approximately 5000 data points. To substantiate the practical applicability of the algorithm, detailed runtime analyses should be provided, especially for high-dimensional data (10 dimensions or greater). The authors should conduct and report benchmarks that examine how computation time scales with both the number of clusters (k) and the dimensionality (d) of the input data. This would provide readers with a clearer understanding of the algorithm's performance boundaries in real-world scenarios.

Methodological Limitations in Comparative Experiments

The experimental comparison with existing methods, particularly Bradley's Constrained k-means mentioned in the literature review, is inadequate. The authors do not present MSE comparisons under identical initialization conditions, which is essential for a fair assessment of clustering quality. Furthermore, the robustness of the algorithm should be evaluated under challenging constraints, such as when cluster size ratios are highly imbalanced (e.g., when the maximum to minimum cluster size ratio exceeds 10:1). Such analyses would demonstrate whether the proposed method maintains its effectiveness under more demanding conditions that might arise in practical applications.

The current experimental section focuses primarily on a single case study (seating arrangement), which, while interesting, does not provide sufficient evidence for the general effectiveness of the algorithm across different data distributions and constraints. I recommend including comparative analyses using standard benchmark datasets from the UCI repository or similar sources to strengthen the empirical validation of the proposed approach.

Conclusion

While the paper presents a novel approach to fixed-size cluster k-means, addressing the highlighted limitations would significantly strengthen the contribution and enhance the paper's impact in the clustering literature.

Declarations

Potential competing interests: No potential competing interests to declare.