Journal of Computational Intelligence and Secure Systems of Artificial Intelligence
Volume 01, Issue 03, July 2025
www.jcissai.com

Real Coded Genetic Algorithm for Graph Clustering Based on SVD

M Sudarshan
Department of Computer Science Engineering, Guntur Engineering College, Guntur
Corresponding Author: sudarshancse2233@gmail.com

To Cite this Article
Sudarshan, "Real Coded Genetic Algorithm for Graph Clustering based on SVD", Journal of Computational Intelligence and Secure Systems of Artificial Intelligence, Vol. 01, Issue 03, July 2025, pp:16-18.

Abstract: *This work uses a random point bipartite graph to present a new genetic algorithm-based graph clustering model. The model makes use of uniformly distributed random points in the data space, and gap between these points and the test points is measured and regarded as closeness. An adjacency matrix is produced using test points and random points. Correlation coefficients are calculated using the provided bipartite graph to produce a similarity matrix. To find the cluster centers, the eigenvectors of the weighted similarity matrix's singular value decomposition are taken into account and fed into an exclusive GA model. The model's performance has been compared to current standard algorithms, and it has been tested using standard datasets.*

Keywords: *Singular Value Decomposition, Bipartite Graph, Cluster Validity, , Clustering, Cluster Validity, Algorithm Graph, Index Genetic*

I. Introduction

Grouping items together based solely on their fundamental proximity to one another is known as clustering. An unsupervised NP-hard grouping issue is clustering. The literature offers a variety of clustering techniques. While Everitt et al. [2] discuss clustering and its underlying algebra, Xu has presented a performance analysis of conventional clustering models. A representation that models the objects and their relationships is called a graph. Typically, the graph's vertices or nodes represent the objects, and edges show the connections between two things. The nature of the relationship is not constrained by graph theory. Relationships can be included from any real-world relationship.

These days, clustering on graphs has become very popular. In graph clustering, the proximities are represented by edges, while the data objects are represented by vertices. That is, if there is an edge connecting two nodes, they are close to one another. The weight matrix is also taken into account for proximity measures in the case of a weighted graph. This indicates that having an edge is insufficient; the closer two things are, the smaller the edge weight. Data points are clustered using typical graph theoretic techniques after the problem has been translated to a graph model by encoding the multiple points into objects and the gap between objects using the different corners.

A web application is created [5]. Schaeffer conducts a survey of graph-based clustering [6]. The study employs a bipartite graph-based data clustering technique in which additional random points are taken from the data points to construct a bipartite graph. The sample points' proximity is determined by how similar they are to the bipartite graph. Due to Bollobas, there is a good body of literature on graphs and bipartite [3].

II. The Mathematical Context

A search-based optimisation method for NP-hard tasks is the genetic algorithm. Biological genetics, which states that a parent's chromosome determines the characteristics of their child, served as the inspiration for the genetic algorithm. It is also thought that the baby may outperform its parents and inherit the positive traits of both parents. The chromosome is made up of several genes. Every gene is in charge of a particular characteristic of the progeny. Genes determine characteristics like nose shape and hair color.

The similar idea has been passed down in genetic algorithms. Each component of the solution is encoded in a gene, and the gene sets are used to construct chromosomes. Every chromosome represents a potential remedy for the issue at hand. Many of these solutions are produced as the starting population in a genetic algorithm. Crossbreeding, also known as crossover, is permitted in the population. Chromosome pairs are selected for mating in this situation. New children are produced as a result. After that, the progeny are sent to a fitness function to assess their fitness or proximity to the best solution.

This is accomplished by determining an objective function and measuring the traits of the progeny. The GA's elitism model chooses the best n chromosomes from the existing pool and the other ones from the progeny using the survival of the fittest rule. The solution set develops in this manner. Random gene mutations may occur during this evolution to produce a completely different solution. We refer to this as mutation. It is anticipated that an acknowledged solution would emerge after a finite number of generations [6].

III. The Methodology of Research

An experimental setting has been selected to evaluate the suggested model's performance. For this arrangement, three factors are taken into account. These are the issues to take into account, the dataset that includes these issues, and the measurement scales that will be used to gauge the results. One can detect the clustering issues primarily in two ways. One is based on separability, while the other is based on the clusters' shapes. The issue of separability refers to how well-separated the clusters are. The issue is a touching problem if the clusters are not sufficiently separated. There are two sorts of cluster shapes.

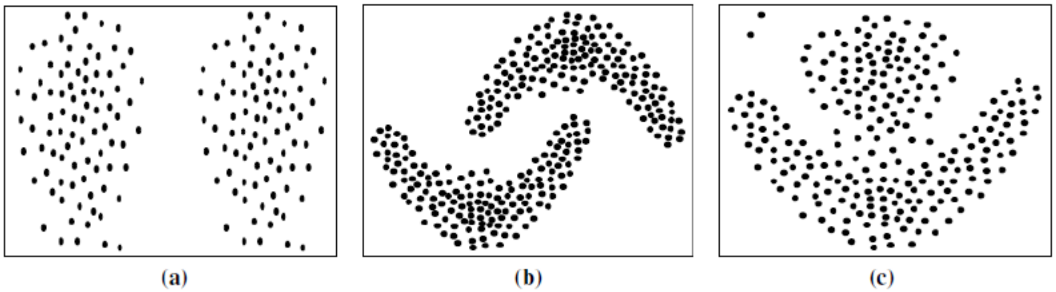

Fig 1: Back Propagation Algorithm

IV. Results and Discussions

As may be observed, the suggested model's index values are higher than those of the full linkage hierarchical model. This suggests that the suggested model outperforms the full linkage hierarchical model across all external indexes in the seed’s dataset. Additionally, the suggested model outperforms the Single Linkage Hierarchical on the seed dataset. However, the suggested model's performance on the seed’s dataset is comparable to that of the regular k-means algorithm and the average linkage hierarchical model. Across every externally validity assessment done, the proposed model surpasses not only the complete linkage hierarchical model and the single linkage hierarchical model, but also the average linkage hierarchical model, and the standard k-means model on the iris dataset.

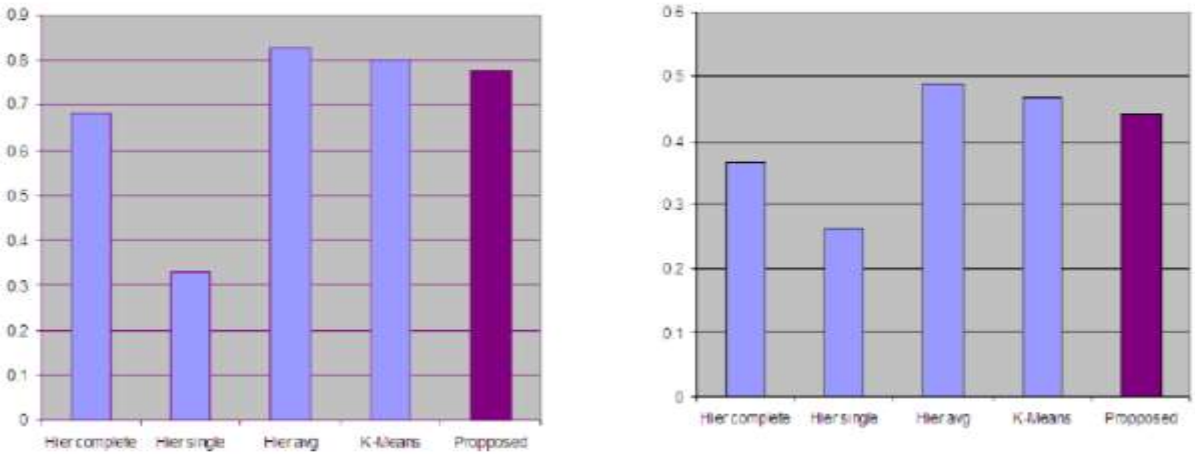

Fig 2: Precision of dataset seeds

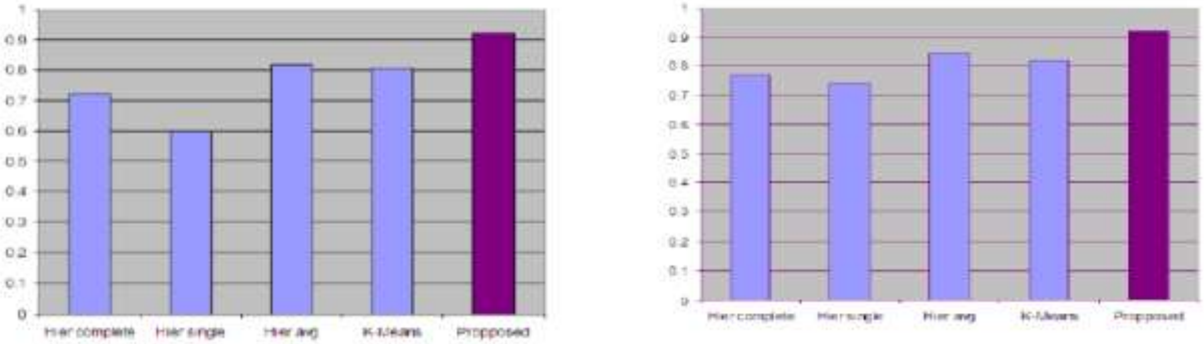

Fig 3: Precision of Iris dataset

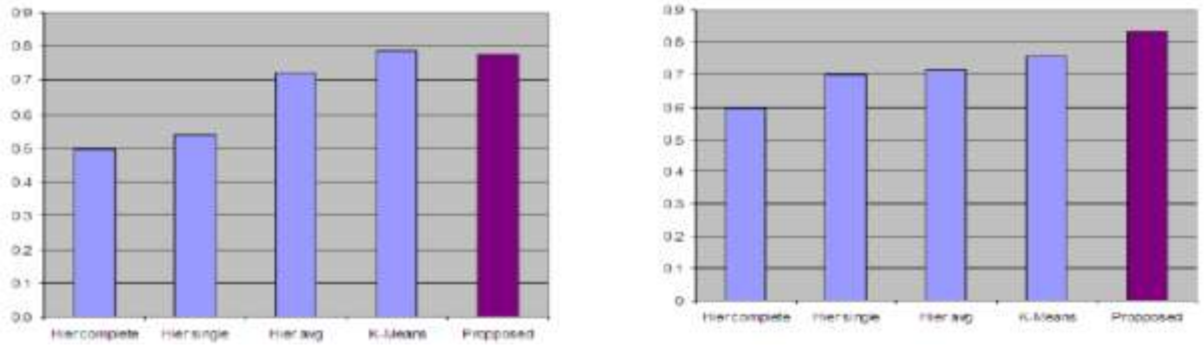

Fig 4: Precision of Flame dataset

V. Final Results and Future Decision

This study presented a novel genetic algorithm-based data clustering method for a bipartite graph's singular matrix. On common datasets, the model performs admirably. The elitist GA model brigs the outcome out of the local optima, which is another benefit of the methodology. Even though the model produces good results, its primary drawback is how computationally expensive it is. Multi-objective evolutionary algorithms are another way to enhance the model. To further fine-tune performance, more soft-computing technologies may be employed.

References

[1] Mohan Kumar and Swetha, Innovation of neural networks in Computer, Sprnger Proceedings *on Fuzzy Logic systems*. 2012; 23(5): 312–322.

[2] Sindhura, Krishan Mohan, Swathi and Tiwari, Implementation of fuzzy and neural in Cluster Analysis, 8th ed. Tata Mc Graw Hill 2017.

[3] R Srikanth, Hima Bindu and Niharika, Algebraic Graph for islanded grid system for neural networks, *IEEE Transactions*. 2012; 210

[4] G Rothermel, R Untch, C Chu, MJ Harrold. Prioritizing Test Cases for Regression Testing. *IEEE Trans. Software Eng*. 2001; 27(10): 929-948.

[5] P. C. Chang and C. H. Liu, "A TSK type fuzzy rule based system for stock price prediction," *Expert Syst. Appl*., vol.34, pp. 135–144, Jan 2008.

[6] SK Abdul Rehaman, John Vangli and Shanli, "Neural network system combined with Fuzzy-rough data reduction with ant colony optimization," *Fuzzy Set Syst*., vol. 231, pp. 56-65, March 2010.

[7] Chen Chen-Hung "A unctional-Link-Based Neurofuzzy Network for Nonlinear System Control"- *IEEE Transaction on Fuzzy Systems*, Vol 16 No 5, October 2008.

[8] Manju Bargavi, Siddharda Roy and Mohit Reddy, "ANN and FLANN Based Forecasting for Conceptual S&P 500 Index"*Information Technology Journal*, 6 (1): 121-132, 2010 Asian Network System for Artificial Scientific Information.