

## Optimizing Node Centrality by Using Machine Learning

Azad Shojaei <sup>a</sup>, Javad Mohammadzadeh <sup>b\*</sup>, Keyhan Khamforoosh <sup>c</sup>

<sup>a</sup> Department of Computer Engineering, Tehran North Branch, Islamic Azad University, Tehran, Iran.

<sup>b</sup> Department of Computer Engineering, Karaj Branch, Islamic Azad University, Karaj, Iran.

<sup>c</sup> Department of Computer Engineering, Sanandaj Branch, Islamic Azad University, Sanandaj, Iran.

**Article History:** Received: 5 April 2021; Accepted: 14 May 2021; Published online: 22 June 2021

**Abstract:** Data is usually described by a large number of features. Many of these features may be unrelated and redundant for desired data mining application. The presence of many of these unrelated and redundant features in a dataset negatively affects the performance of the machine learning algorithm and also increases the computational complexity. Therefore, reducing the dimension of a dataset is a fundamental task in data mining and machine learning applications. The main objective of this study is to combine the node centrality criterion and differential evolution algorithm to increase the accuracy of feature selection. The proposed method as well as the performance dataset preparation of the proposed method was compared with the most recent and well-known feature selection methods. Different criteria such as classification accuracy, number of selected features, as well as implementation time were used to compare different methods. The comparison results of the different methods were presented in various forms and tables and the results were completely analyzed. From the statistical point of view and using different statistical tests like Friedman different methods were compared with each other. The results showed that the selected evolutionary differential algorithm for clustering, instead of finding all the elements of the cluster centers present in the data set, found only a limited number of DCT coefficients of these centers and then by using the same limited coefficients, cluster centers reconstructed.

**Keywords:** Machine Learning, Node Centrality Criterion, High Dimensional Data

### 1. Introduction

Data is usually described by a large number of features. Many of these features may be unrelated and redundant for desired data mining application. The presence of many of these unrelated and redundant features in a dataset negatively affects the performance of the machine learning algorithm and also increases the computational complexity. Therefore, reducing the dimension of a dataset is a fundamental task in data mining and machine learning applications [6].

The high dimension of data in the machine learning algorithm also creates important problems. High-dimensional images in machine learning include features for unrelated and redundant learning that reduce the performance of the learning algorithm and consider processing these data that requires a great deal of time and computational resources based on selecting the appropriate feature to improve machine learning performance in terms of reducing time for the learning model and increasing the accuracy in the learning process. Because selecting the appropriate feature for machine learning with high dimensional data is necessary to demonstrate the proper performance of machine learning by selecting the appropriate feature (Hamidi et al., 2016). Feature selection has an important role in machine learning and pattern recognition. Many of the selected features are selected with a learning algorithm to produce complete classification properties. However, many of the features selected for the learning task are unrelated and reduce the efficiency of the learning algorithm and face the algorithm fitness with difficulty. The accuracy of learning and the learning fastness may be significantly worse with redundant features. Therefore, the selection of related and essential features at the pre-processing stage is crucial and important [9].

High dimensions of data are used in different areas of machine learning, artificial intelligence and data mining. As an example, its applications can be in the categorizing documents, which results in the retrieval of appropriate information, but high dimensionality of the feature spaces for selecting feature is conducted by feature dimensionality reduction approach. Not all features are important for text extraction in document categorizing, and some features need to be selected according to the desired criteria. Reducing the dimension of feature and time of text categorization, enhancing the accuracy of text categorization, and how to search for the sub-set space of possible features in categorization of documents is of particular important. Feature selection increases the learning algorithm and improves the accuracy of the data results. Selecting a feature simplifies the model and reduces computational cost because a model with low inputs has less evaluation for new features [10].

Also categorize texts in machine learning due to the high volume of words cause problems in extracting data and appropriate features. This problem can be solved by word modeling and selection of appropriate feature and improved word classification efficiency in the text. This modeling enables the features to be selected appropriately and simple classifications can produce good results. Because by text categorization can select the high dimensional features, which it reduces dimensions in text categorization [3].

In the dataset, time complexity is one of the fundamental problems in data mining. The time complexity of high-dimensional data is created when the data is stored in a database to extract features. Because the feature is

used to measure the observable processes. To select the appropriate features of the data, can classify its machine learning algorithms. By selecting a feature in this data, the computation equipment as well as dimensional effect of data and its efficiency increase. Therefore, the time complexity of this type of data can be very effective with machine learning algorithms to focus on the features selected from the dataset [4]. By selecting a feature, a subset of features is selected based on the optimization criteria. Since the performance of traditional statistical methods has decreased, this decrease in data dimension is due to the increase in the number of observations and the number of features related to the observation. Because the feature selection is used to improve performance prediction, better understanding of datasets related data to machine learning and so on. Also by removing unrelated and redundant features can reduce the redundant network performance and increase the accuracy and reduce the complexity of the data. Time series analysis in high dimensional data is considerable because it uses the clustering method to select the feature in different time series. This performance is based on data similarity criteria at similar or near time intervals, which simplifies data approaches and plays an important role in data mining and data discovery in dataset. Due to the expansion of search space at high dimensions, the ability of evolutionary algorithms to deal with complex and multi-objective problems, optimization and lack of structural information in dataset clustering leading to lack of need to modeling and their understand in data clustering for complex datasets. The efficiency and performance of the differential evolutionary optimization algorithm can be used as a fast and efficient search method for clustering data to select the appropriate feature. Also, due to the high volume of data in these networks, the complexity in the nodes also occurs and cuts off the main data in network, but removing each node does not affect the entire network. In these networks, the importance of nodes depends on the centrality of the node. In fact, dataset is the essential principles of an indicator for predictable behavior that the centrality criteria are depend on the graph structure. Accordingly, combination of differential algorithm and node centrality criterion can be used to increase the accuracy of feature selection in high-dimensional networks to improve their performance.

In this research, an attempt is made to present a feature selection method based on feature clustering and evolutionary differential algorithm. In the proposed method after feature clustering by using graph clustering and community detection algorithms, each cluster is identified by evolutionary differential algorithm based on centrality criterion of appropriate features. Feature selection based on the relationships between features is an old idea for decreasing dimensions of problem that has long been considered. Features clustering can conduct primary analysis on data features and remove many redundancies in the primary features. As a result, the selected features will have the highest correlation with the target class and the least redundancy. The advantage of this method is that it first results in reduced computational complexity. Because it is costs less than thousands of features when classifying on a small subset of features. Secondly, due to the reduction in dimension, the execution time is reduced.

### 1.1. Machine learning

In machine learning, imaging techniques contain precise information that process challenges large amounts of data, and develops new concepts and technologies in remote sensing as well as machine learning methods in parallel. By fully understanding the basic algorithms and methods in machine learning from deep connection to system development solves the problem, which by solving this problem can optimally extract the information and classify machine learning algorithm. In high-dimensional images, many features of unrelated redundant and noisy learning task reduce and the processing of this data requires computational time and resources. Therefore, feature selection plays an important role in improving the performance of machine learning algorithms in terms of reducing time, creating a learning model, and increasing accuracy in the learning process [11].

Deep learning is a kind of deep neural network of machine learning and a set of algorithms. This learning can model high-level concepts by learning at different levels and layers. This learning helps a kind of link establishes among different classes and datasets, which is a common paradigm feature, so it can be applied to the designed system equations after defining paradigm and the model to be trained model because the system itself discovers paradigm [7].

As one of the broad and widely used branches of artificial intelligence, machine learning is adjustment and discovering methods and algorithms that enable computers and systems to learn. The objective of machine learning is that the computer (in its most general concept) can gradually and with increasing data have higher efficiency in the desired task.

### 1.2. Clustering

Data clustering is an important data mining technique in data analysis. It is similar to a single dataset in a database and is necessary in wide range of research fields and their applications. In paradigm detection, data clustering is often used as unsupervised classification of unlabeled data. Data clustering is also one of the most popular tools for extracting a subset of data that is an interesting view of large databases. Data clustering techniques are suitable for data compression and are known as quantization vectors. Clustering algorithms are iterative refinement. Generally it is added to find appropriate clusters from a dataset. However, the computational cost of

the algorithms increases with the size and dimensions of the dataset, and is a major problem. Therefore, in order to reduce the cost for clustering high-dimensional data, there are many methods for clustering parallel data [8].

In classification of nodes or clustering all the nodes in the network are first subdivided into clusters, and in each cluster, one node is selected as the head, and the rest of the nodes are called normal nodes. The method of selecting head in each method takes into account different criteria. In cluster-based methods, its main objective is to distribute energy consumption across all nodes. In fact, the methods that are working based on nodes classification they might have different energy consumption. These methods are among the best routing algorithms in these networks and are inspired by new methods to enhance the lifespan of the network [13].

Clustering can be considered as the most important issue in unsupervised learning. Clustering attempts to divide the data into different clusters so that the similarity between the data within each cluster is maximized and the similarity between the data within the different clusters is minimized.

## 2. Research Methodology

### 2.1. Clustering Nodes Based on the Nodes Centrality Profiles

The value of centrality measurement can depend on clusters of nodes and subsets with separate centrality profiles and network topology performances. Hierarchical clustering used for a network with centrality measurement values across all nodes firstly by using normalization sigmoid function.

$$f(x) = \left[ 1 + \exp\left(\frac{-(x-\mu)}{\sigma}\right) \right]^{-1}$$

Which  $\mu$  is the mean and  $\sigma$  is the standard deviation of the values for measuring data centrality across the node of a network and then it is calculated linearly in unit distance. Hierarchical clustering is performed by using the least ward variance method for the Euclidean distance between pairs of centrality (normalization) criteria. Assuming that the Davies-Bouldin index is used to determine the specific resolution of the dendrogram slice and it is investigate the resulting clusters. Davies-Bouldin index is a different inter-cluster similarity ratio for a clustering solution. Low values of Davies-Bouldin represent an effective clustering solution.

### 2.2. Additional Methods Definitions of Centrality

Each network as (A) adjacent matrix  $N \times N$  indicates that  $A_{ij} = 1$ . If the nodes of  $i$  and  $j$  are connected  $A_{ij} = 0$ . Adjacent matrix indicates the weighted network, which encodes weight element of  $W_{ij}$  from the edge between the nodes of  $i$  and  $j$ . These definitions of weight networks were simply replaced for  $A_{ij}$  by  $W_{ij}$ .

#### 2.2.1. Degree/Strength (DC)

The simplest measurement of centrality is called the degree of centrality, which is defined as the number of edges connected to a node:

$$DC_i = k = \sum_{j \neq i} A_{ij}$$

For weighting networks, similar degrees of weighting degree are used; otherwise it is used as  $s$  node strength, which is the sum of all edge weights connected to the node.

$$DC_i = s = \sum_{j \neq i} W_{ij}$$

#### 2.2.2. Harmonic Index Centrality (HC)

While it is commonly used to measure the productivity and impact of scientists' work, the index of  $h$  has also recently been used as a centralized matrix for analyzing complex networks. If so, a set of  $i$  node neighbors that having value equal to or greater than  $h$ , index of  $h$  from  $i$  node can be defined as follows:

$$HC_i = \max_{1 \leq h \leq k_i} \min(|N_{\geq h}(i)|, h)$$

Where  $h$  is the value between 1 and the degree of the  $i$  node. Therefore, the  $h$  index of a node defines the maximum value of  $h$  for the  $h$  neighbors of the  $i$  node at least degree of  $h$ .

### 2.2.3. Leverage Centrality (LC)

It is the another measure of centrality, which is considered in the leverage centrality relationship of a neighbor node, in addition to the measurement of the other centrality, is impossible; leverage centrality can allocate negative values to a node. The node has less connection than its neighbors. A node is affected by its neighbors. A node of positive values can have more connections than its neighbors, that is, it affects the influence of its neighbors. Leverage centrality is defined as follows:

$$LC_i = \frac{1}{k_i} \sum_{j \in N(i)} \frac{k_i - k_j}{k_i - k_j}$$

Which  $N(i)$  is the set of neighbors of  $i$  node. In the weighting networks strength used instead of degree.

### 2.2.4. Eigenvector Centrality (EC)

Eigenvector centrality has high value among the nodes and determines with the neighbors that are has high degree. This measurement as Eigenvector,  $v$ , with the greatest value of Eigenvector  $\lambda_1$  is related to the adjacent matrix and can be expressed as follows

$$EC_i = v_i = \frac{1}{\lambda_1} \sum_j A_{ji} v_j$$

### 2.2.5. Katz Centrality (KC)

In a network connected with a large and flexible module, the exclusive main center of high score for the nodes within the module and a low score (if not zero) is allocates for the nodes outside the module. Therefore, it is not appropriate to measure node detection outside the module. Two parameters Katz centrality  $\alpha$  and  $\beta$  are added as Eigenvector centrality definition.  $\alpha$  Parameter is the contribution of distant dependencies (i.e., neighboring nodes of neighbors) intervenes to the node centrality stage. The parameter allocates a certain amount of centrality to each node, so confidence of each node as the centrality value is non-zero. As Katz centrality allocates each node a small amount of centrality, it also ensures that highly connected nodes in other clusters have high centrality score. Katz centrality can be expressed as follows:

$$KC_i = \alpha \sum_j A_{ji} v_j + \beta$$

Or it is defined in the matrix form as follows:

$$KC = \vec{\beta}(I - \alpha A)^{-1}$$

Which  $\vec{\beta}$  is a vector of  $N$  size with each element of  $\beta$  and  $I$  is the identity matrix of  $A$ . In all analyzes,  $\alpha$  is the set is less than 10% inverse of the largest eigenvalue (usually a value close to the largest eigenvalue usable) from the network and  $\beta$  is the set of 1.

### 2.2.6. PageRank Centrality (PR)

With the Eigenvector and Katz centrality, low-degree nodes may gain a high score, only because with their low-degree are connected to more nodes. Correct PageRank centrality for this behavior with the scale of neighbors nod contribution of  $i$  and  $j$  to node centrality of  $i$  by  $i$  degree is expressed as follows:

$$PR_i = \alpha \sum_j A_{ji} \frac{v_j}{k_j} + \beta$$

Matrix of this definition can be written as follows:

$$PR = \vec{\beta}(I - \alpha D^{-1}A)^{-1}$$

That the D is an oblique matrix and  $D_{ii}$  is the degree of i node (in the weighting network of S instead of oblique  $S_{ii}$  the power of node is i). Parameters of  $\alpha$  and  $\beta$  are the same functions in Katz centrality. Or all of analyzes  $\beta$  is the set of 1 to 0.85 of set.

### 2.2.7. Closeness Centrality (CC)

Closeness centrality defines a node as the center, if it has at least the mean of minimum path length to any other node in the network. It is assumed that nodes with a mean short path length to other nodes can transmit or receive information in a relatively short period of time. Since the mean of a shorter path is introduced as a reversed central node, most central nodes have higher values and define as below:

$$CC_i = \frac{N}{\sum_j l_{ij}}$$

This  $l_{ij}$  is shortest topology distance between the nodes of i and j. Weighting networks of computed  $l_{ij}$  is as the shortest weighting path (the path with the shortest set of weighting edges) is determined by using weighting inverse matrix.

### 2.2.8. Information Centrality (IC)

This criterion is also known as Closeness centrality of the main stream can investigate all possible paths between two nodes as well as overlap in these paths and their weight in each information value that includes the path. Information in the path identifies as the inverse topology of that path.

To estimate the Information centrality, firstly  $C = (L + J)^{-1}$  matrix is identifies, which L laplacian of A and J is the  $N \times N$  matrix by computing all elements to an element. Information centrality is defines as below:

$$IC_i = \left( C_{ii} + \frac{\sum_j C_{jj} - 2\sum_j C_{ij}}{N} \right)^{-1}$$

In the weighting networks L is the laplacian of W.

### 2.2.9. Random Walk Closeness Centrality (RWCC)

Random Walk Closeness centrality computes the mean time to a random step is taken at the start of each node in the network to reach the i node and inverse of the first average time mean to a particular node. The first average time mean can be computed from the main Z matrix.

$$Z = (I - P + \Pi)^{-1}$$

That the (I) is the identity matrix, the transfer matrix  $P = D^{-1}A$  (or  $P = S^{-1}W$  in the weighting networks), and  $\Pi$  is a  $N \times N$  matrix that each column, the  $\pi$  vector of stable distributed probable state is from the transfer matrix (like  $\Pi_{ij} = \pi_j$ ).  $\pi$  Vector can be obtained by solving the system of linear equations  $\pi P = \pi$  and  $\sum_i \pi_i = 1$ . The matrix of H first average time mean can be defined as follows:

$$H_{ij} = \frac{Z_{jj} - Z_{ij}}{\pi_j}, i \neq j$$

Element of  $H_{ij}$  is the mean matrix of first average time from the  $i$  node to  $j$  node. Random Walk Closeness centrality simply computed as follows:

$$RWCC_i = \frac{N}{\sum_j H_{ji}}$$

### 2.2.10. Sub Graph Centrality (SC)

Like other criteria, Sub graph centrality counts a number of steps, but instead of counting steps with other nodes, considers the method of steps closeness. Thus, the criterion of Sub graph centrality consists of several sub graphs, which are determined by the close steps, to which the node belongs, to the smaller sub graphs. Longer steps (and thus larger sub graphs) computed by the weight of each step with  $\frac{1}{n!}$  coefficient that the  $n$  is the length of step. Therefore Sub graph centrality can be computed as follows:

$$SC_i = \sum_{n=0}^{\infty} \frac{[A^n]_{ii}}{n!} = [e^A]_{ii}$$

In the weighted networks reduction of adjacent matrix  $S^{-1/2}wS^{-1/2}$  is used instead of  $A$ .

## 3. Data Analysis

### 3.1. Clustering

Clustering is one of the unsupervised learning branches and is an automated process in which the samples are divided to the groups that its members are similar to each other, referred to as cluster. Cluster is therefore a set of objects in which the objects are similar to each other and dissimilar to the objects in the other clusters. For similarity, a different criterion can be considered, for example, the distance criterion can be used for clustering, and objects that are closer to each other can be considered as a cluster that this type of clustering is called distance-based clustering.

### 3.2. Centrality

Obviously, the importance of features is not the same among available data. Some of them are more important because of the different positions in the type of subject. This importance allows for greater access to information or a greater role in its transmission for individuals. That is why we consider certain data to be influential and determinative in the dataset. The importance and popularity of these data in different contexts are determined by different criteria. In this study we use the centrality of eigenvalues:

This method computes the importance of nodes based on adjacent nodes. The computation happens on graphs with strong connectivity. If a node is connected to the nodes that are of high importance, its importance is increased by their influence. This method considers the same importance iteratively for the computation of node. First, all nodes are given an initial score. This concession goes on as long as the chain that achieves stability continues. The scoring in this method is based on this concept that high-connectivity nodes help the nodes that follow them in terms of score.

This part of the simulation is used as an innovation by combining the differential evolutionary algorithm and the node centrality criterion to reduce the dimensions of problem and feature selection. The laplacian centrality criterion algorithm is used to cluster features. And some kind of social networking algorithms coupled with differential evolution algorithm for feature clustering are first investigated in this research. The advantages of using these algorithms in combination are as follows. In these algorithms the number of clusters is automatically determined and the number of clusters is determined by the user. On the other hand, in most clustering methods the features are predetermined, in which the dispersion of the features in each cluster is not taken into account in the clustering process. Therefore this method will not be able to detect the optimal clusters. In the population detection algorithm used in this study, both the dispersion of features within each cluster as well as the degree of features connection across different clusters is considered. Therefore, this algorithm will be able to find optimal clusters.

### 3.3. Datasets used

The UCI reference is used to evaluate the method presented in the previous sections from the datasets in the real world. This library contains datasets for evaluating machine learning algorithm.

In order to evaluate the proposed model in this simulation, in this part of the study, first the neural network is used to increase the accuracy of the data and the support vector machine is used as well.

### 3.4. Artificial Neural Network

This classification algorithm is presented based on non-parametric approaches and is widely used in scoring problems. ANN can be used for nonlinear problems. In the present study, the ANN has three layers, the inner layer consisting of neurons related to the input variables and the output layer having one neuron.

### 3.5. Support Vector Machines

Support vector machines are a supervised classification method based on the theory of statistical learning theory. The basic idea behind this classifier is to find an optimal page cloud as a decision level to maximize the margin between the two classes. If the data is not linearly separated, the data is transferred to a higher dimensional space with a nonlinear kernel data and the optimal page cloud is determined in that space.

### 3.6. Compare and Evaluation of Previous Articles

Abolghasemi and Momtazi (2018) in a research used a machine learning algorithm to select text features and improve text mining, which resulted in accessing to increased feature selection in the Persian text categorization in this regard. The method presented in this section demonstrates the ability to identify the feature of the connections classifications in the dataset with high accuracy and error reduction compared to the presented research method in this paper.

Ismaili and Abbasi (2017) in a research by using closest neighboring classification method and genetic algorithm have investigated the reduction of feature dimensions which have resulted in attaining increased accuracy in data classification. In the method considered in the Ismaili and Abbasi paper, the data of one cluster is similar to the data close to it as possible and is different from the data of the other clusters. Similarity between records is measured by the distance function. The distance function receives two input records and returns a value that shows the similarities of the two, one of the disadvantages of this method is that usually the output of a clustering algorithm consists among each cluster. To implement the clustering algorithm, it is necessary to determine in advance the number of clusters that require prior information of the existing data because without them, the prerequisite for determining the clusters is difficult. Each data sample is assigned to a cluster of data that is the least distance to the center of that cluster. At the end, the centers of each cluster are determined as the output of the algorithm, by using the genetic algorithm to determine the most optimal value, and the feature is selected based on the selection of the best value. From disadvantages of this method can mention the time of accessing to best value that is it is long and expensive in some cases, but the algorithm proposed in this study has a higher accuracy than the algorithm presented in Ismaili and Abbasi article in addition to being optimally low in time because the differential evolutionary algorithm is presented as the multi objective optimization algorithms, which have the ability to find near-optimal solutions for math problems. In fact, the accuracy of selecting the differential evolutionary algorithm is higher than the genetic algorithm and the error rate in the proposed algorithm is lower than that of the genetic algorithm.

Chadhari and Agarawal (2018) in a research by using the EBQPSO algorithm concluded that its effect is resulted in improved accuracy and invocation of particle and quantum optimization algorithms. In this study, EBQPSO algorithm was used in gene dataset to classify cancer. One of the main problems of the algorithm used in this study is the lack of local search algorithms. All data may not be examined and the accuracy can be reduced to some extent, resulting in an increased error rate in the simulation, but the differential evolutionary algorithm used combined with the node centrality criterion algorithm eliminates this defection and resulting in high accuracy and reduced error rate in the simulation.

Hiwa et al (2018) in a study by using support vector classification machine and genetic algorithm optimization, proposed a new method for automatic extraction of specific brain network and graph theoretical properties. The method used in this study has limitations such as the fact that, for example, it is not yet clear how parameters can be determined for a mapping function. Support vector-based machines are requiring complex and time consuming computations and due to the computational complexities consume a lot of memory. Discrete and non-numeric data are also incompatible with this method and must be converted. But the proposed algorithm in this study solves these problems and is able to select the high data feature in the less time.

## 4. Conclusion

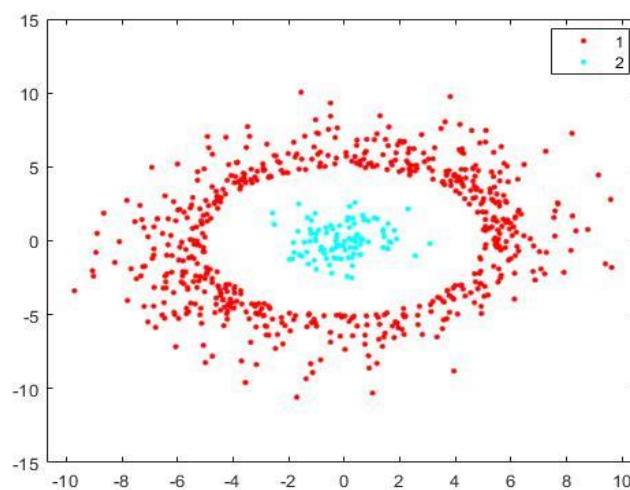
After extracting the feature, the clustering is conducted on the obtained features. Given the number of clusters with the clustering algorithm, the features in each cluster are ranked. From each cluster numbers of highest scored features are selected. The features selected from all clusters are sorted by the score at which the cluster is assigned

and the obtained features are evaluated. The use of feature values in clusters or classes causes local feature behavior to be applied to different categories.

The clustering process consists of two stages. In the first step, the optimal number of clusters is determined based on the validation indices. The clustering algorithm depends on several factors such as the number of clusters and the distance between the clusters. After determining the optimal number of clusters, the mean-k algorithm is used for cluster categorization. The main objective of the optimization head clusters phase is to select each cluster based on a series of defined criteria of a cluster center. The following three criteria are used to select cluster centers:

- 1. The energy sum of selected cluster centers
- 2. The distance sum of the nodes of that cluster from the center of the cluster
- 3. The distance sum of the selected node centers from each other
- 4. The node centrality sum of the selected cluster centers

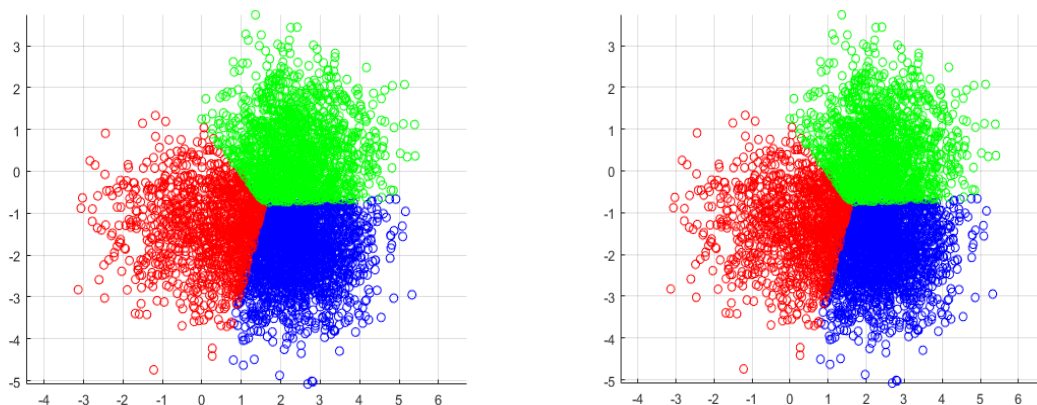
As can be seen in this respect, the higher the first and the fourth criteria and the two other criteria are lower is indicate better cluster centers.



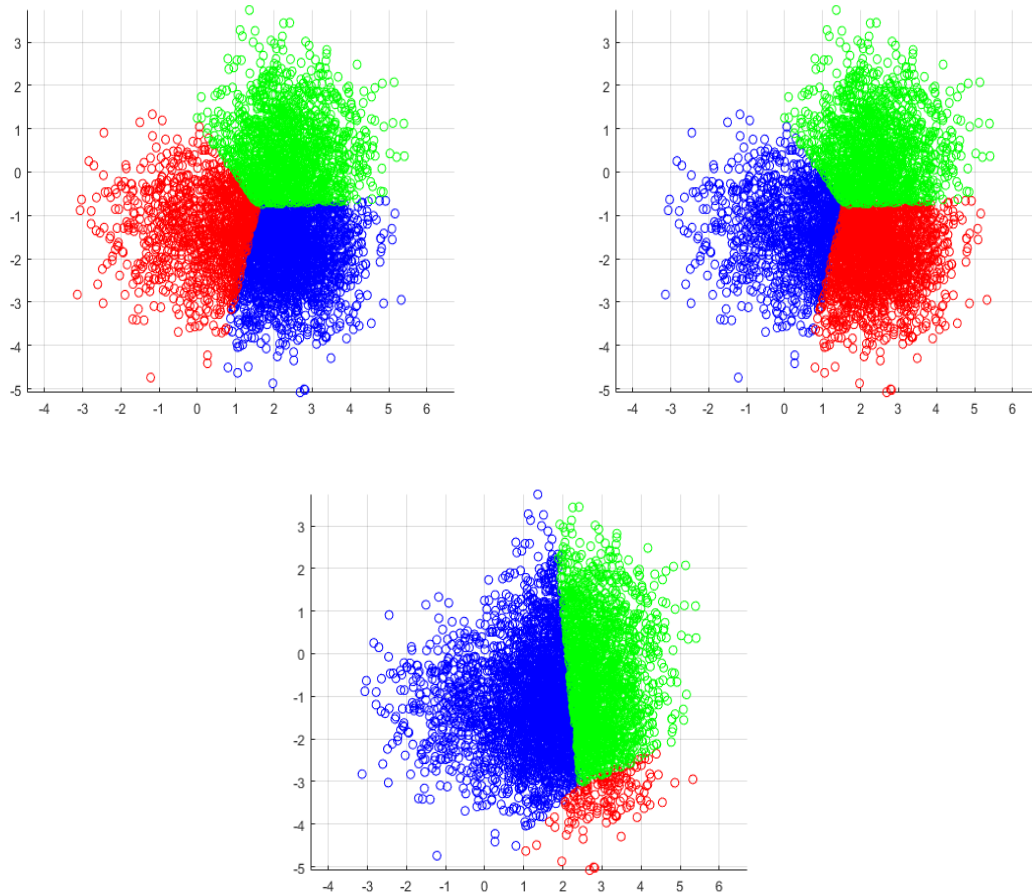
**Figure 1.** Clustering

Feature clustering can perform a basic analysis on the features and can eliminate many redundancies in the primary features. As a result, the features selected will have the most relevance to the target class and the least redundancy.

Solve of clustering problem in general and automated clustering issue specifically can be out of common clustering algorithm power. One of the solutions that considered for this subject is that converting clustering issue to an optimization issue and its resolve can be carried out by intelligence and revolutionary optimization algorithms.







**Figure 2.** Diagrams of differential evolutionary clustering

As a result, the proposed method is a multi-objective algorithm that needs to be optimized.

In this algorithm, each solution has the same number of dimension to the cluster centers and must be select the best cluster centers. Each particle of the hybrid algorithm has a  $k$  dimension that  $k$  denotes the number of clusters. In this algorithm the  $(i)$  dimension of each particle represents the center of the selected cluster for the nodes in that cluster. The algorithm employs a differential operator to generate new solutions that exchange information between population members. One of the benefits of this algorithm is having a memory that stores the information for the right answers in the current population. Another advantage of this algorithm is its selection operator. In this algorithm, all members of a population have an equal chance of being selected as a parent. That is, the infant's generation is compared to the parent's generation in terms of the extent of competence measured by the objective function. Then the best members move on to the next stage as the next generation.

The most important features of the DE algorithm are its high speed, simplicity and robustness. This method only starts by setting three parameters. NP parameter, population extent, parameter F mutation weight and parameter C are multiplied to the difference of two vectors and added to the third vector. The F parameter is usually set to 0 to 2,  $r$  and the C parameter is set to 0 to 1.

The differential evolution algorithm is presented to overcome the main problem of genetic algorithms, namely the lack of local search in these algorithms. The main difference between genetic algorithms and DE algorithm is in the order of mutation and coupling operators as well as in the mode of selection operator. The selected evolutionary differential algorithm for clustering, instead of finding all the elements of the cluster centers in the dataset, finds only a finite number of DCT coefficients of these centers and then reconstructs them by using the same finite coefficients of the cluster centers. After clustering features by using graph clustering and community detection algorithms, from each cluster, appropriate features is identified by evolutionary differential algorithm based on centrality criteria.

---

**References**

1. Abolghasemi, M., Momtazi, S. 2018, Text Mining Improvement by Selecting Feature Words, Fourth International Web Research Conference, Tehran.
2. Bendechange, M., & Kechadi, M. (2018). Distributed clustering algorithm for spatial data mining. arXiv preprint arXiv:1802.00304.
3. Bharara, S., Sabitha, S., & Bansal, A. (2018). Application of learning analytics using clustering data Mining for Students' disposition analysis. *Education and Information Technologies*, 23(2), 957-984.
4. Chaudhari, P., & Agarwal, H. (2018). Improving Feature Selection Using Elite Breeding QPSO on Gene Data set for Cancer Classification. In *Intelligent Engineering Informatics* (pp. 209-219). Springer, Singapore.
5. Esfandiarpour, S. 2015, Feature Selection by Using Colonial Competition Algorithm, Master Thesis of Computer Science, Shahid Bahonar University of Kerman.
6. Fraiwan, L., & Lweesy, K. (2017, March). Neonatal sleep state identification using deep learning autoencoders. In *Signal Processing & its Applications (CSPA), 2017 IEEE 13th International Colloquium on* (pp. 228-231). IEEE.
7. Hamidi, M., Ebadi, H., Kiani, A. 2016, A Comprehensive Review on Machine Learning and Feature Selection Methods with Emphasis on Classification in Remote Sensing Applications, 2nd National Conference on Geospatial Information Technology Engineering, Khajeh Nasireddin Tusi University of Mapping Engineering
8. Hiwa, S., Obuchi, S., & Hiroyasu, T. (2018). Automated Extraction of Human Functional Brain Network Properties Associated with Working Memory Load through a Machine Learning-Based Feature Selection Algorithm. *Computational intelligence and neuroscience*.
9. Ismaili, Z., Abbasi, E. 2017, Application of Hybrid Algorithm Feature Selection Method in Predicting Short Term Performance of Corporate Initial Public Offerings in Securities Exchange of Shiraz University Central Conference.
10. Liu, Y., J.-W. Bi, and Z.-P. Fan, Multi-class sentiment classification: The experimental comparisons of feature selection and machine learning algorithms. *Expert Systems with Applications*, 2017. 80: p. 323-339.
11. Marsan, G. A., Bellomo, N., & Gibelli, L. (2016). Stochastic evolutionary differential games toward a systems theory of behavioral social dynamics. *Mathematical Models and Methods in Applied Sciences*, 26(06), 1051-1093.
12. Masoudian, S., Derharmi, V., Zarifzadeh, S. 2015, Investigation of Feature Selection Methods and Text Subject Classification Methods by Using Persian News Data, 46th Iranian Mathematical Conference, Yazd University.
13. Oldham, S., Fulcher, B., Parkes, L., Arnatkeviciute, A., Suo, C., & Fornito, A. (2018). Consistency and differences between centrality metrics across distinct classes of networks. arXiv preprint arXiv:1805.02375.
14. Wu, H., & Prasad, S. (2018). Semi-Supervised Deep Learning Using Pseudo Labels for Hyperspectral Image Classification. *IEEE Transactions on Image Processing*, 27(3), 1259-1270.