International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7241
Article Received: 26 April 2023 Revised: 20 June 2023 Accepted: 01 July 2023

Attribute Selection Algorithm with Clustering based Optimization Approach based on Mean and Similarity Distance

[1]Rajasekhar Kaseebhotla, [2]Dr. K. Raghava Rao, [3]Dr. Mallikarjuna Rao
[1]Research Scholar,
Koneru Lakshmaiah Education Foundation,
Guntur, India
Email Id: rs.kaseebhotla@gmail.com
[2]Professor in CSE, Dept.of ECM,
Koneru Lakshmaiah Education Foundation,
Guntur, India,
Email Id: raghavarao@kluniversity.in
[3]Professor of CSE, GRIET,
Hyderabad, India
Email Id: professorcmrao@gmail.com

Abstract— With hundreds or thousands of attributes in high-dimensional data, the computational workload is challenging. Attributes that have no meaningful influence on class predictions throughout the classification process increase the computing load. This article's goal is to use attribute selection to reduce the size of high-dimensional data, which will lessen the computational load. Considering selected attribute subsets that cover all attributes. As a result, there are two stages to the process: filtering out superfluous information and settling on a single attribute to stand in for a group of similar but otherwise meaningless characteristics. Numerous studies on attribute selection, including backward and forward selection, have been undertaken. This experiment and the accuracy of the categorization result recommend a k-means based PSO clustering-based attribute selection. It is likely that related attributes are present in the same cluster while irrelevant attributes are not identified in any clusters. Datasets for Credit Approval, Ionosphere, Annealing, Madelon, Isolet, and Multiple Attributes are employed alongside two other high-dimensional datasets. Both databases include the class label for each data point. Our test demonstrates that attribute selection using k-means clustering may be done to offer a subset of characteristics and that doing so produces classification outcomes that are more accurate than 80%.

Keywords- PSO; Attribute selection; High dimensional; Classification; Cluster.

I. INTRODUCTION

With the benefits that machine learning offers, research is swiftly spreading to all of the key sectors where manually processing data was previously a barrier. Machine learning studies the process by which computers may glean information from data and then anticipate future data based on knowledge already amassed. The data collected from the preceding data are technically used to create an objective function. The objective function is frequently referred to as a model. In supervised learning, the objective function assigns each input data to the proper class [1]. The process of assigning the correct class label (aiming for class) to information which greatly contributes to that classification is known as classification. The objective function used for this mapping is also known as a classifier [2]. Data containing one or more attributes or variables make up the objective function's input. High-dimensional data has a large number of attributes, sometimes thousands. In reality, not every aspect of high-dimensional data is relevant or essential to the intended function. In other words, eliminating these unused characteristics from the input data will not affect how the objective function behaves [3].

High-dimensional data may also contain redundant attributes, which are many components that have an identical effect on the results of the objective function, in addition to irrelevant attributes. Such traits could be encapsulated in a single attribute to reduce the computational burden of high-dimensional data [4]. Finding the most crucial characteristics in high dimensional data is the goal of an attribute selection strategy [5]. Attributes that are judged extraneous or superfluous will be removed. This deletion attempts to reduce the processing load of high-dimensional information in order to speed up the calculation of the optimization problem in the classification process [6]. The most important attributes

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7241
***Article Received**: 26 April 2023 **Revised:** 20 June 2023 **Accepted:** 01 July 2023*

should adequately reflect the whole attribute set of high dimensional data and be able to be used to predict class labels, which is a goal function output. An attribute selection technique has the advantage of making high-dimensional data computations simpler while keeping the precision of the resulting class labels as represented in Figure 1.

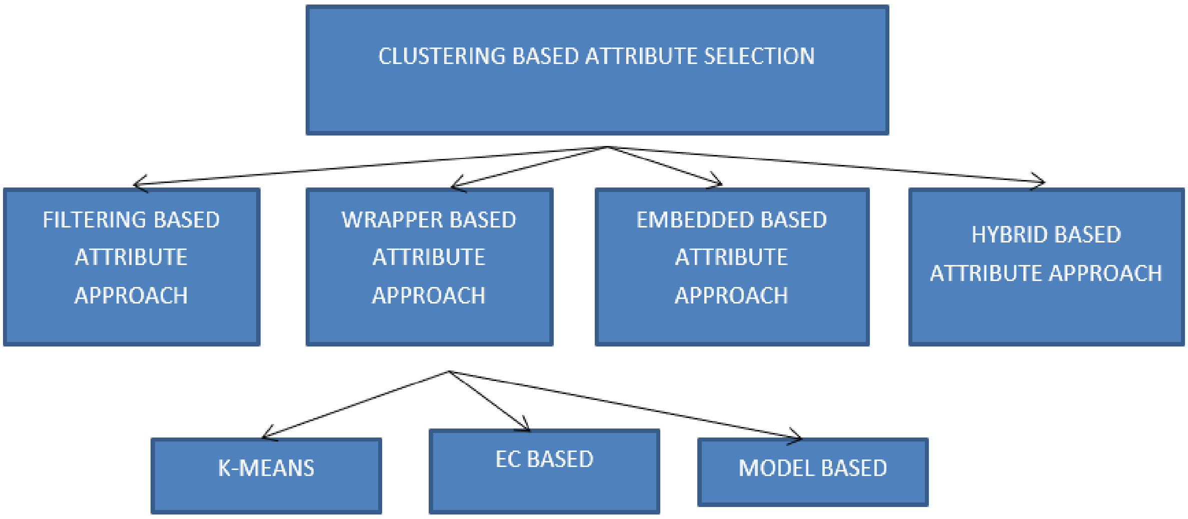

Figure. 1. Clustering based attribute selection approaches

Finding the most crucial characteristics in high dimensional data is the goal of an attribute selection strategy [7] and [8]. Attributes that are judged extraneous or superfluous will be removed. This deletion attempts to reduce the processing load of high-dimensional information in order to speed up the calculation of the optimization problem in the classification process [9]. The most important attributes should adequately reflect the whole attribute set of high dimensional data and be able to be used to predict class labels, which is a goal function output. An attribute selection technique has the advantage of making high-dimensional data computations simpler while keeping the precision of the resulting class labels. The attribute selection algorithms currently discussed in the literature can thus be divided into three categories: algorithms that focus entirely on either eliminating unnecessary and redundant attributes or algorithms that focus entirely on both.

The work contribution is summarized as follows,

II. LITERATURE SURVEY

The evolutionary computing strategy based ACBEA (Automatic Clustering with BEA) algorithm was implemented by Das et al. By using the Bacterial Evolutionary Algorithm (BEA), an automatic grouping challenge was solved. BEA's chromosome population can be increased by applying two specific operations in it: transfer operation and Bacterial Mutation. Bacterial mutation, like bacterial genetics, restores chromosomal sections and their mimic. A transfer operation carries out the population's chromosomes' informational exchange. As a result, the genetic information of the bacteria spreads fast to other cells in the absence of a crossover operation. Initially, chromosomes are encoded with different numbers of classes due to their varying lengths as the number of clusters is unknown. Using Evolutionary Algorithms (EAs), [10] suggested an automatic grouping method. Here, the primary goal of the suggested approach is to transform the clustering problem into a global optimization problem and offer an EA solution. The main goal of the suggested GA method, which balances intra-cluster consistency and inter-cluster consistency, is to provide new validity indices. Three adaptive coding approaches have been designed for the automatic detection of cluster numbers. These coding systems were created in a way that fixed-length chromosomes were used to address the issue of variable length optimization.

A coefficient of variation-based K-Means algorithm (CV-k-means) was put forth by [11]. The devised approach uses a variation coefficient weight vector to lessen the impact of irrelevant features. A feature learning choice-based clustering approach was put into practice by [12]. The automatic indirect feature weight learning method improved the clustering performance. This method is also employed to safeguard against quick convergence. k-means and Generalized Fashion method (GFA) based hybrid optimization method were presented by [13] to achieve optimal clustering. The suggested algorithm's performance was evaluated by comparing its findings to those of other algorithms such as k-means, GA, imperialist competitive

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7241
***Article Received**: 26 April 2023 **Revised:** 20 June 2023 **Accepted:** 01 July 2023*

algorithm, particle swarm optimization, and GFA. In terms of convergence speed and solution quality, the suggested method outperforms the alternatives. K-means clustering, and GA were used to create the genetic K-means clustering algorithm by Lu et al. The outcome confirmed that the proposed GA-based strategy performed better than the conventional k-means algorithm. [14] designed a K-means-ACO algorithm to address concerns such as sluggish ACO convergence and K-means misclassification. The suggested approach takes into account the ACO elicitation data, which is the k-means result, and adds illumination pixels and illumination probability in ACO ants seeking rules. As a result, it allows direct node selection that disregards chance in favor of pheromone concentrations. likewise takes into account all information that was elicited without modifying the ACO random search's quality.

In order to address premature convergence and quick data clustering, [15] developed a two-phase GAI-PSO-based K-Means data clustering algorithm. This newly created technique provides quick data grouping while also preventing premature convergence. First, the GAI-PSO upgraded PSO algorithm was used. The population-based heuristic search technique known as GAI-PSO was created using a combination of social and cultural rules that were discovered through the use of GA, natural selection, and PSO analysis. The position updating rules and standard velocity of PSOs were merged in GAI-PSO using crossover, mutation, and selection from Gas. GAI-PSO's detection of the ideal initial cluster centroids from the search solution space preceded the nesting phase. The second stage, sometimes referred to as the localized refinement stage, used the k-means algorithm to arrive at the best answer. The created method merged the K-Means approach's capability for quick convergence with that of the global search's GA and avoided both of their limitations. A method for determining a species from its genomic sequence was put out by [16]. Here, the MapReduce framework was used to find feature descriptors for a genome sequence. They included A, T, C, and G nucleotide bases in their strategy to create three-letter keyword unique feature descriptors. The feature descriptor count was used to group Genome sequences from related species. Finally, they created a MapReduce-based clustering system using ACO, Differential Evolution (DE), and K-Means.

[16] used ACO and PSO with improvements to suggest a mixed clustering method. In this instance, the data set partition was optimized with ACO and adaptive ACO results were refined with enhanced PSO. [17] suggested a clustering method based on PSO (PSOBC). This procedure was created to regulate the ideal cluster center in cluster analysis. In [18], the author clusters using PSO and K-means. Additionally, [19] presented a PSO-based k-means algorithm taking into account its capability of a global search to conduct the best selection of the first clustering center. This was accomplished through the dynamic adjustment of specific parameters and inertia weight. The suggested methods confirmed superior performance in the estimate of the centroid's ideal solution, which is used in the clustering process. Additionally, author [20] did a comparable attempt in which they integrated a user-specific number of clusters with the provided collection of data. Similar efforts were made by Tran et al., who presented a k-means and PSO-based technique [21] and used a neighborhood search strategy to enhance PSO. This improved PSO accelerates the algorithm's convergence and provides assistance in overcoming local optima.

A hybrid dynamic data clustering technique called KCPSO, which combines k-means with combinatorial PSO, was developed by [22]. The proposed KCPSO does not require a specific number of clusters throughout the clustering process. The benefit of the created KCPSO is that this is obviously necessary for the conventional k-means approach [23]. As each iteration of the approach suggested to optimize the number of clusters, a discrete PSO was utilized. This gives better results for grouping with the K-means method. [24] devised a clustering algorithm that combines the Bee algorithm (BA) and the PSO to tackle concerns with local minima and used it for data clustering [25] Some of the literature also focused on the clustering of data using an ant colony on a digraph, which is covered in the next section.

1. Problem statement

These days, every field is producing datasets with a higher dimension. This results in a decline in machine learning performance. This problem can be solved by using a Dimensionality Reduction (DR) approach. In the DR method, feature selection is one method that may be used to help minimize the upper dimension by picking the most prominent features. Three alternative strategies, namely filter, wrapper, and embedding, are used in the feature selection procedures. A few recent literary works have demonstrated that the clustering technique is also utilized to choose the most pertinent qualities.

III. PROPOSED METHODOLOGY

A. Clustering Techniques:

Clustering is the technique of automatically grouping unlabeled occurrences in a given dataset using established similarity measures such as Euclidean distance, point symmetry, and hamming coding. A cluster of similar instances is established [26]. Each cluster's centroid serves as a specific representation of that cluster. Clustering has been widely applied to a wide range of data mining and machine

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7241
***Article Received**: 26 April 2023 **Revised:** 20 June 2023 **Accepted:** 01 July 2023*

learning applications, including image analysis, pattern recognition, information retrieval, and wireless networks, to mention a few as in Figure. 2.

A large number of clustering approaches have been proposed in the literature. These methodologies can essentially be divided into partitioned, hierarchical, density-based, and model-based methodologies [27]. Utilizing distance-based metrics, partitioned techniques iteratively distribute instances to clusters. These methods produce one-level, non-overlapping spherical forms. The methods K-means, K-medoids, and fuzzy C-means (FCM) are the most representative of in this area. The goal of hierarchical techniques is to create a hierarchy by dividing instances into several levels. By merging or breaking clusters, the hierarchy can be created [28].

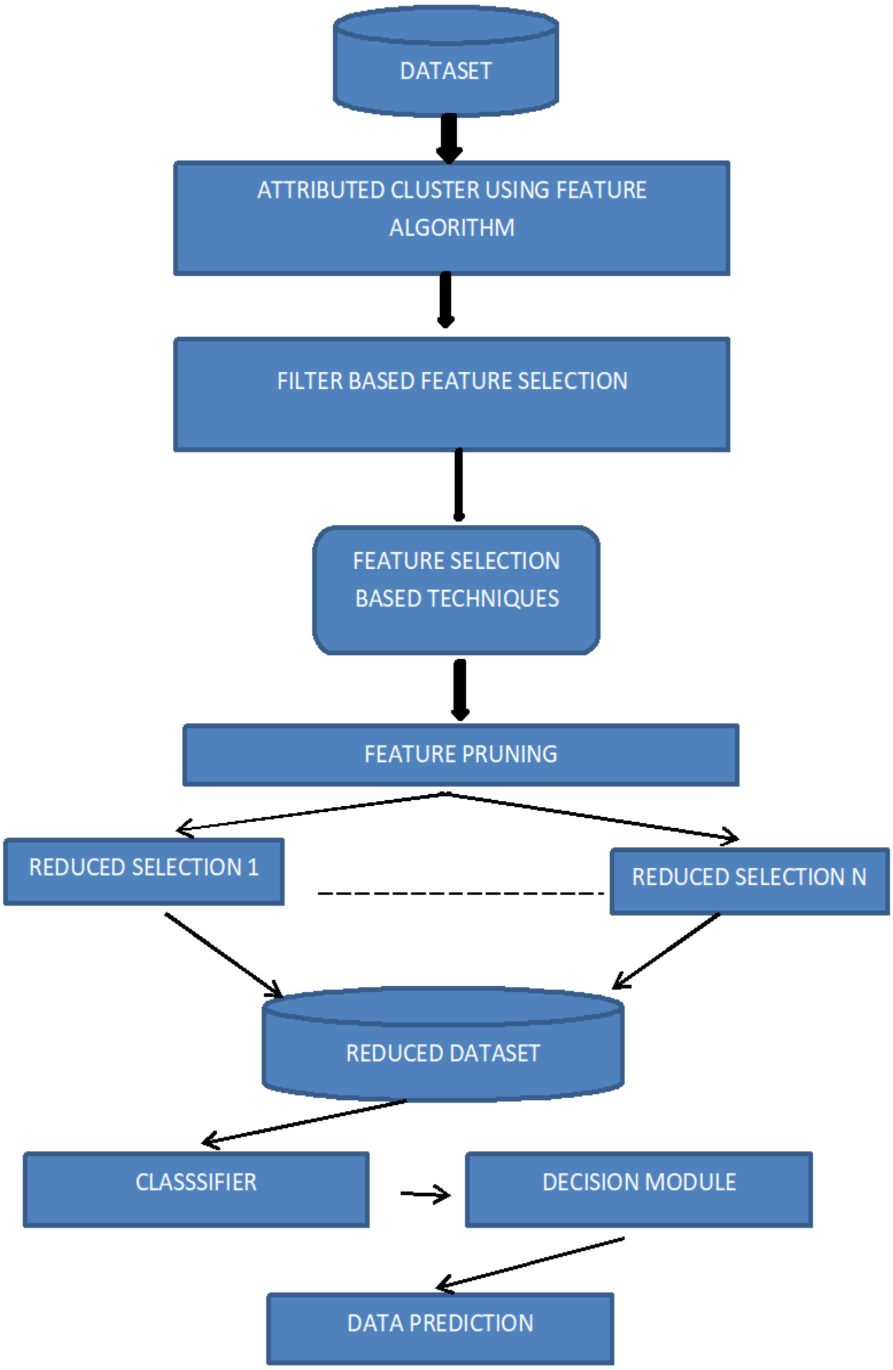

Figure. 2. Feature Selection Model

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7241
Article Received: 26 April 2023 Revised: 20 June 2023 Accepted: 01 July 2023

By taking the concentration of the data's areas into account, density-based techniques attempt to find clusters. Low-density areas separate instances from various clusters. The three most well-liked methods in this group are DBSCAN, OPTICS, and DENCLUE. Model-based clustering attempts to develop a mathematical model for the data and then employs a probability resulting from this model as the grouping criterion [29-31]. Expectation-maximization and the Gaussian mixture modeling serve as exemplary models for model-based clustering. Hence the proposed modified PSO uses the canopy ***k***-Means clustering technique is used to identify the similarity attribute.

B. Optimization Method:

The traditional partition clustering method (K-means algorithm) requires the value of K (number of clusters) to initiate the clustering procedure. However, the proposed approach does not require the precise value of K as input; rather, it takes a range of possibilities and finds the right value of K. We also test the suggested approach for the quality of produced clusters using several cluster validity criteria. The suggested method combines partition clustering (K-means) and swarm intelligence (PSO) techniques. In clustering, we must optimize the values of inter-cluster distance and intra-cluster distance in order to produce high-quality clusters. Therefore, we view partition clustering as an optimization problem and employ PSO to address this optimization issue.
The user must enter the values for the range of the number of clusters in the suggested strategy. The proposed method iteratively evaluates every potential value in the interval before providing the numerical value (number of clusters) for which the best answer was identified. The method additionally provides the centroids of these clusters. The K-means algorithm is run once for each value of the number of clusters in the suggested method. The K-means algorithm's output, the centroids' values, is fed into PSO to set the starting location of a particle. Random initialization is applied to the PSO's leftover particles. The proposed method optimizes the distance between clusters and quantization error using PSO. The suggested approach makes use of multiple cluster validity measures to assess the effectiveness of the generated clusters in order to determine the ideal number of clusters as in Algorithm1.

<u>Algorithm 1: Optimization Algorithm based on Particles.</u>

Input: Number of Clusters ‘C’

Output: Optimal Number of Cluster and Centroid Value

1. Based on the number of clusters, maximum and minimum values are assigned
2. Deploy K-Means Algorithm
3. {
4. Assign value based on the number of clusters.
5. Determine the centroid of the cluster.
6. }
7. Initialize the PSO particles based on K-means.
8. {
9. Calculate the Particle Swarm Optimization based on number of iteration.
10. Optimize the inter and intra cluster.
11. Determine the optimal centroid value based on ‘C’.
12. Validate the optimal value.
13. }
14. If (Clustering = Values ‘C’)
15. {
16. Number of Clusters ‘C’ = Optimal Value validation
17. Optimal number of Cluster
18. Centroid location
19. Else
20. Terminate
21. }

C. KM- PSO Algorithm

The proposed KM-PSO algorithm is used to find the similarity attribute by using the clustering approach. In this algorithm, feature clustering and the filter method are combined to select the similarity attribute. In the first stage, canopy ***k***-Means clustering is used to identify the nature of the attribute and it will group the attribute based on the Euclidean distance measure. In the second stage, a correlation measure is used to remove the dissimilarity attribute from each cluster. Finally, PFA is used to select the similarity attribute.

Based on Figure. 3, training datasets and it classify the data into similarity and dissimilarity information in the data clusters. Then the input contains datasets as it contains data cluster ‘D’ = {CD1, CD2, …. CLDN} and then choose an arbitrary value for a cluster in ‘P’. Based on the arbitrary value, distance between the attribute can be calculated using Euclidean Distance. Then assign the data mean and similarity distance for each cluster to determine the mean value. Classify the similarity and dissimilarity information in the data cluster. Based on the relevant information in the data cluster, PFA is calculated using K-means algorithm. Then based on the dissimilarity information in the data cluster, Correlation variance is calculated as represented in Algorithm 2.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7241
***Article Received**: 26 April 2023 **Revised:** 20 June 2023 **Accepted:** 01 July 2023*

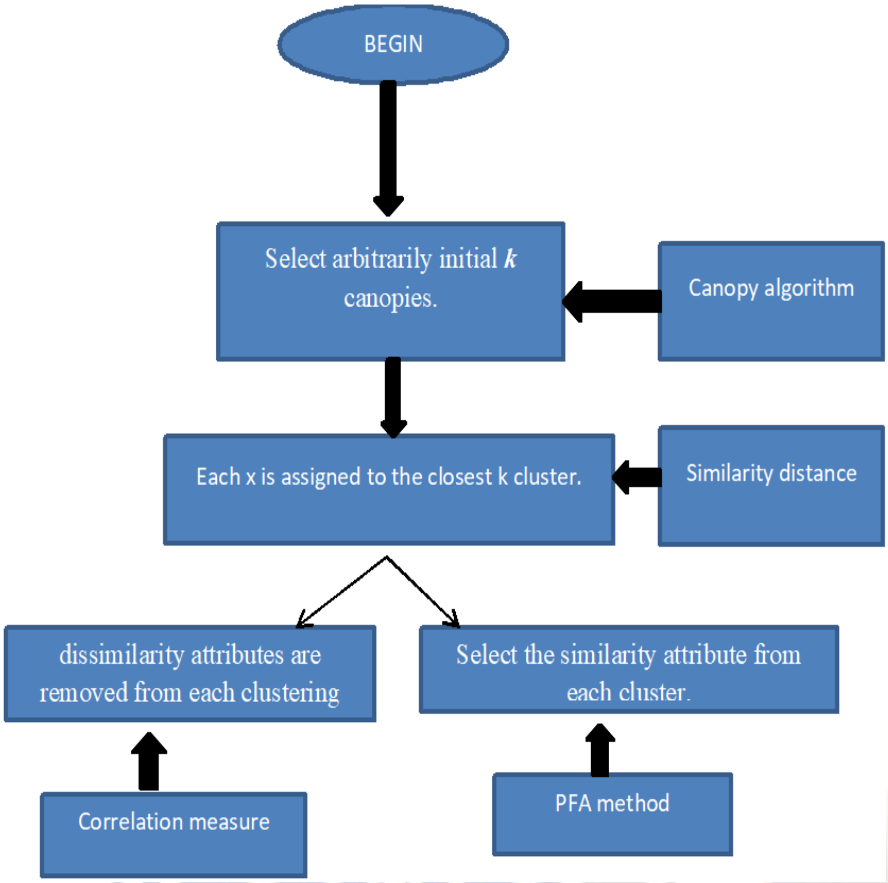

Figure. 3. Proposed KM- PSO Flowchart

Algorithm 1: Proposed KM-PSO Algorithm

Input: Training Datasets

Output: Optimal feature subset

1. Data cluster 'D' as {CD_1, CD_2, …. CD_n}
2. Choose an arbitrary value for a cluster selected initially 'P'
3. Determine the distance from each attribute in each cluster 'P'
4. {
5. Euclidean Distance $d^2_{xy} = (x_1 - y_1)^2$ - $(x_2 - y_2)^2$
6. $d^2_{xy} = \sqrt{(x_1 - y_1)^2 - (x_2 - y_2)^2}$
7. }
8. Assign mean and similarity distance based on data cluster for each attribute.
9. {
10. Compute the new mean value for each cluster.
11. Repeat 2 to 4
12. Stop clustering.
13. Remove the dissimilarity data in the clusters.
14. }
15. If (similar data in the clusters)
16. Calculate the PFA among the cluster attribute.
17. {
18. For (i=1; i<=P; i++)
19. {
20. Compute the covariance / Correlation matrix.
21. $P_{ij} = E[X_i X_j] / E[X_i^2] E[X_j^2]$
22. Compute Principal Component and Eigen Values
23. Choose the dimensional Subspace 'q' and matrix construction 'A_q' belong A
24. Construct Variability Record
25. Calculate the vector cluster using K-Means Algorithm
26. {
27. Euclidean Distance $d^2_{xy} = \sqrt{(x_1 - y_1)^2 - (x_2 - y_2)^2}$
28. }
29. }
30. Determine the corresponding vector 'V' = principal vector
31. Else If (dissimilarity data in the clusters)

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7241
***Article Received**: 26 April 2023 **Revised:** 20 June 2023 **Accepted:** 01 July 2023*

{
Correlation between the attributes
For (i=1; i<=P; i++)
{

$$C = \frac{E(XC)-E(X)E(C)}{\sqrt{\sigma^2(X)\,\sigma^2(X)}}$$

}

IV. PERFORMANCE ANALYSIS

The relevant features are selected using FSALS algorithm. The classification accuracy of the FSALS is analyzed with k-NN classifier.

A. Comparison Analysis of Proposed KM-PSO Algorithm

The proposed KM-PSO algorithm is used to find the relevant features. In this algorithm, clustering technique and is integrating with correlation measure to select the most relevant features. Comparison is made between the proposed KM-PSO and with the existing two feature selection algorithms and the result are shown in Table 1 and the same are graphically displayed in figure 4 and 5 by considering the Credit Approval, Ionosphere, Annealing, Madelon, Isolet and Multiple features datasets.

TABLE I. COMPARISON OF CFS, RELIEFF AND KM-PSO ALGORITHMs

Datasets	Actual Features	CFS	ReliefF	KM-PSO
Credit Approval	16	14	12	10
Ionosphere	35	25	22	18
Annealing	39	33	29	22
Madelon	500	430	405	385
Isolet	617	603	575	534
Multiple features	649	625	605	574

B. Estimating the Classification Accuracy using Classifier

In order to show that the proposed KM-PSO algorithm efficiently improves the classification accuracy, an experiment is conducted. For this experiment, the k-NN classifier is used to find the classification accuracy after the selection of relevant features using KM-PSO.

THE results from the analysis reveal that the proposed KM-PSO algorithm shows the improved classification accuracy than the existing algorithms. Table 2 describes the results of the classification accuracy of all the algorithms it is graphically shown in Figure 5.

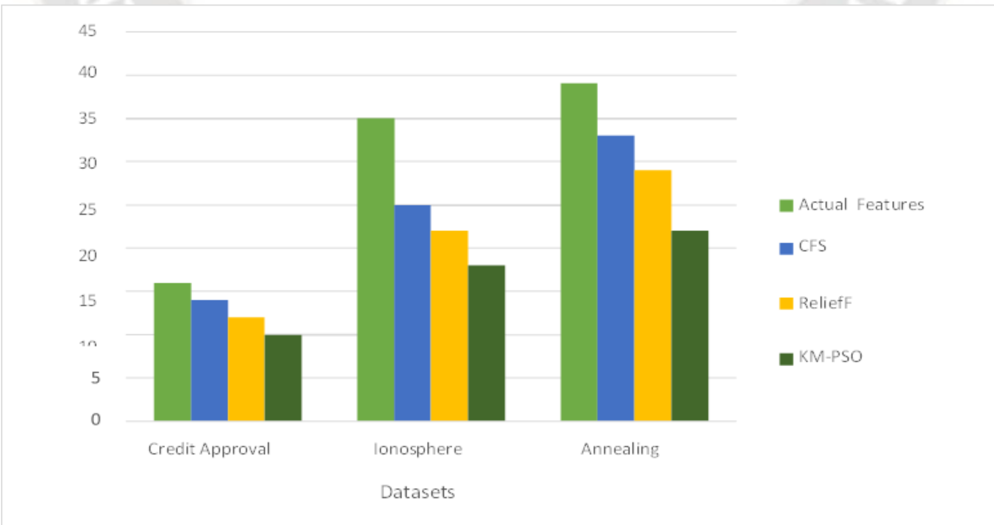

Figure 4. Comparison of CFS, ReliefF and KM-PSO for the First Three Datasets

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7241
***Article Received**: 26 April 2023 **Revised:** 20 June 2023 **Accepted:** 01 July 2023*

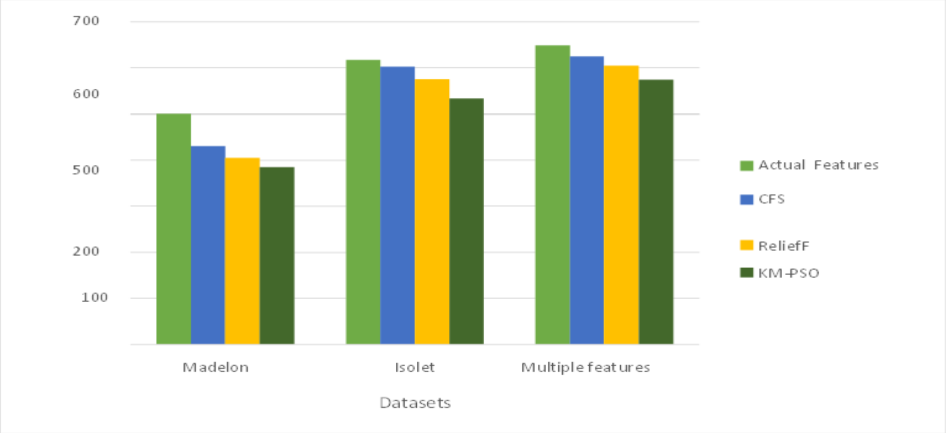

Figure 5. Comparison of CFS, ReliefF and KM-PSO for the Next Three Datasets

TABLE II. CLASSIFICATION ACCURACY OF CFS, RELIEFF AND KM-PSO ALGORITHMS USING K-NN CLASSIFIER

Datasets	**Actual Features**	**CFS**	**ReliefF**	**KM-PSO**
Credit Approval	16	14	12	10
Ionosphere	35	25	22	18
Annealing	39	33	29	22
Madelon	500	430	405	385
Isolet	617	603	575	534
Multiple features	649	625	605	574

In recent days few algorithms were developed to select the relevant features using the clustering approach. Hence KM-PSO algorithm is proposed using a clustering approach with standard UCI datasets shown in table 2. In this analysis, the proposed methods are compared with the CFS (Correlation based Feature Selection) and ReliefF feature selection algorithms.

V. CONCLUSION

This study provides a new attribute subset selection strategy to address the dimensionality issue. Using the k-means clustering algorithm, dissimilarity characteristics are identified and eliminated, and redundant characteristics are removed using the correlation measure from each cluster. Every last feature has been given priority. The suggested method's accuracy is evaluated using selection criteria that consider a wide range of relevant features. The proposed method was tested on Text and Microarray datasets using the Nave Bayes classifier, and the results were compared to representative feature selection methods. Experimental results showed that the proposed technique is fairly effective and noticeably speedy. Furthermore, we propose to study the method of using a filter measure as a measure of similarity for selecting just representative attributes with the goal to enhance the efficiency for learning algorithms.

It is interesting to see that using cluster centroids as representative examples for the full feature set produces decent class prediction accuracy. Class prediction automatically disregards features that are unrelated to class prediction since they are situated on the periphery of clusters and are not cluster centroids. To learn more about the sensitivity and make sense of the data being looked at, this experiment can be widened to include more datasets. However, given the growing amount of data and the need for online classifiers, feature selection ought to take precedence. Big data analysis will be one of the most important tools in all facets of life. It may be used to identify trends and patterns.

REFERENCES

[1] S. Tiruveedhula, V. Narayana, "A Survey on Clustering Techniques for Big Data Mining", Indian Journal of Science and Technology Vol 9(3), DOI: 10.17485/ijst/2016/v9i3/75971, January 2016, pp. 1-12.

[2] Yadav C, Wang S, Kumar M. Algorithms and approaches to handle large data sets - A survey. International Journal of Computer Science and Network. 2013; 2(3):1–5.

[3] Fahad A, Alshatri N, Tari Z, Alamri A. A survey of clustering algorithms for Big Data: Taxonomy and empirical analysis. IEEE Transactions on Emerging Topics in Computing. 2014 Sep; 2(3):267–79.

[4] Berkhin P. Survey of clustering data mining techniques in grouping multidimensional data. Springer. 2006; 25–71.

[5] Xu R, Wunsch D. Survey of clustering algorithms. IEEE Trans¬actions on Neural Networks. 2005 May; 16(3):645–78.

[6] Kailing K, Kriegel HP, Kroger P. Density-connected subspace clustering for high- dimensionality data. Proceedings of the 2004 SIAM International Conference on Data Mining; 2010. p. 246–57.

[7] Liu, C., Zhou, A., & Zhang, G. Automatic clustering method based on evolutionary optimisation. IET Computer Vision, (2013). 7(4), 258–271.

[8] S. Ren and A. Fan, "K-means clustering algorithm based on coefficient of variation," Image and Signal Processing (CISP), 2011 4th International Congress on, Shanghai,2011, pp.2076-2079.

[9] Bin Zeng, Chao Luo, Wei Zhao and Benyue Chen, "The optimization arithmetic of K-means clustering based on Indirect Feature Weight Learning," 2010 International Conference on Computer and Communication Technologies in Agriculture Engineering, Chengdu,2010,pp.243-246.

[10] A. Aghamohseni and R. Ramezanian, "An efficient hybrid approach based on K-means and generalized fashion algorithms for cluster analysis," AI & Robotics (IRANOPEN),2015, Qazvin,2015,pp.1-7.

[11] Dwarkanath Pande, S. ., & Hasane Ahammad, D. S. . (2022). Cognitive Computing-Based Network Access Control System in Secure Physical Layer. Research Journal of Computer Systems and Engineering, 3(1), 14–20. Retrieved from https://technicaljournals.org/RJCSE/index.php/journal/article/view/36

[12] Bin Lu and Fangyuan Ju, "An optimized genetic K-means clustering algorithm," Computer Science and Information Processing (CSIP), 2012 International Conference on, Xi'an, Shaanxi, 2012, pp. 1296-1299.

[13] Sun Xu, Zhang Bing, Yang Lina, Li Shanshan and Gao Lianru, "Hyperspectal image clustering using ant colony optimization (ACO) improved by K-means algorithm," 2010 3rd International Conference on Advanced Computer Theory and Engineering(ICACTE), Chengdu, 2010, pp. V2-474-V2-478.

[14] R. F. Abdel-Kader, "Genetically Improved PSO Algorithm for Efficient Data Clustering," Machine Learning and Computing (ICMLC), 2010 Second International Conference on, Bangalore, 2010, pp. 71-75.

[15] R. Bhavani, G. Sudha Sadasivam and R. Kumaran, "A novel parallel hybrid K-means-DE-ACO clustering approach for genomic clustering using MapReduce," Information and Communication Technologies (WICT), 2011 World Congress on, Mumbai, 2011, pp.132-137.

[16] Wen Xiong and Cong Wang, "A novel hybrid clustering based on adaptive ACO and PSO," Computer Science and Service System (CSSS), 2011 International Conference on, Nanjing,2011,pp.1960-1963.

[17] B. Naik, S. Swetanisha, D. K. Behera, S. Mahapatra and B. K. Padhi, "Cooperative swarm based clustering algorithm based on PSO and k-means to find optimal cluster centroids," Computing and Communication Systems (NCCCS), 2012 National Conferenceon, Durgapur,2012,pp.1-5.

[18] H. A. Atabay, M. J. Sheikhzadeh and M. Torshizi, "A clustering algorithm based on integration of K-Means and PSO," 2016 1st Conference on Swarm Intelligence and Evolutionary Computation (CSIEC), Bam, Iran, 2016, pp. 59-63.

[19] L. Tan, "A Clustering K-means Algorithm Based on Improved PSO Algorithm," Communication Systems and Network Technologies (CSNT), 2015 Fifth International Conference on, Gwalior, 2015, pp. 940-944.

[20] A. Ahmadyfard and H. Modares, "Combining PSO and k-means to enhance data clustering," Telecommunications, 2008. IST 2008. International Symposium on, Tehran, 2008, pp.688-691.

[21] D. C. Tran, Z. Wu and V. X. Nguyen, "A new approach based on enhanced PSO with neighborhood search for data clustering," Soft Computing and Pattern Recognition (SoCPaR), 2013 International Conference of, Hanoi, 2013, pp. 98-104.

[22] D. Yazdani, S. Golyari and M. R. Meybodi, "A new hybrid approach for data clustering," Telecommunications (IST), 2010 5th International Symposium on, Tehran,2010, pp.914-919.

[23] A. Abraham, S. Das and A. Konar, "Document Clustering Using Differential Evolution," 2006 IEEE International Conference on Evolutionary Computation, Vancouver, BC, 2006, pp. 1784-1791.

[24] K. Lu, K. Fang and G. Xie, "A Hybrid Quantum-Behaved Particle Swarm Optimization Algorithm for Clustering Analysis," Fuzzy Systems and Knowledge Discovery, 2008. FSKD '08. Fifth International Conference on, Shandong, 2008, pp. 21-25.

[25] Y. Kao and S. Y. Lee, "Combining K-means and particle swarm optimization for dynamic data clustering problems," Intelligent Computing and Intelligent Systems, 2009. ICIS 2009. IEEE International Conference on, Shanghai, 2009, pp. 757-761.

[26] A. D. Thakare, C. A. Dhote and S. M. Chaudhari, "Intelligent hybrid approach for data clustering," Communication and Computing (ARTCom 2013), Fifth International Conference on Advances in Recent Technologies in, Bangalore, 2013, pp.102-107.

[27] Ling Chen, Li Tu and Hong-Jian Chen, "Data clustering by ant colony on a digraph," 2005 International Conference on Machine Learning and Cybernetics, Guangzhou, China, 2005, pp. 1686-1692Vol.3.

[28] Madhu Y., Srinivasa R. P., Srinivasa T M, "Enhancing K-means Clustering Algorithm with Improved Initial Center", International Journal of Computer Science and Information Technologies (2010), Vol. 1 (2), pp. 121-125.

[29] Abdul N. and Sebastian M. P., "Improving the Accuracy and Efficiency of the k-means Clustering Algorithm", Proceedings of the World Congress on Engineering (2009), Vol I.

[30] Romero-Zaliz, R. C., Rubio-Escudero, C., Cobb, J. P., Herrera, F., Cordon, O., & Zwir, I. A multiobjective evolutionary conceptual clustering methodology for gene annotation within structural databases: A case of study on the gene ontology

database. IEEE Transactions on Evolutionary Computation, (2008, December). 12(6), 679–701.
[31] Drown, D. J., Khoshgoftaar, T. M., & Seliya, N. Evolutionary sampling and software quality modelling of high-assurance systems. IEEE Transactions on Systems, Man, and Cybernetics— Part a. Systems and Humans, (2009, September). 39(5), 1097–1107.
[32] Das, S., Abraham, A., & Konar, A. Automatic clustering using an improved differential evolution algorithm. IEEE Transactions on Systems, Man, and Cybernetics—Part a. Systems and Humans, (2008, January). 38(1), 218–237.