

Converting Ensemble Clustering Problem to a Mathematical Optimization Problem and Providing an Approach to Solve Based on Optimization Toolbox

Simintaj Salehpour*¹, Hamid Parvin²

¹Faculty of Engineering, Department of Computer, Yasouj branch, Islamic Azad University, Yasouj, Iran

²Faculty of Engineering, Department of Computer, Nourabad Mamasani Branch, Islamic Azad University, NourabadMamasani, Iran

*Email: salehpour.simintaj@gmail.com

Abstract

Nowadays, we live in a world in which people are facing with a lot of data that should be stored or displayed. One of the key methods to control and manage this data refers to grouping and classifying them in clusters. Today, clustering has a critical role in information retrieval methods for organizing large collections inside a few significant clusters. One of the main motivations for the use of clustering is to determine and reveal the hidden and inherent structure of a set of data. Ensemble clustering algorithms combine multiple clustering algorithms to finally reach an overall clustering system. Ensemble clustering methods by lack of information fusing utilize several primary partitions of data to find better ways. Since various clustering algorithms look at the different data points, they can produce various partitions from such data. It is possible to create a partition with high performance by combining the partitions obtained from different algorithms, even if the clusters to be very dense from each other. Most studies in this area have examined all the initial clusters. In this study, a new method is used in which the most sustainable clusters are utilized instead of all primary produced clusters. Consensus function based on co-association matrixes used to select more stable clusters. The most stable clusters selection method is done by cluster stability criterion based on F-measure. Optimization functions are used to optimize the obtained final clusters. The genetic algorithm is the optimizer used in this article to find the ultimate clusters

participated in a consensus. Experimental results on several datasets show that the output of proposed method is various clusters with high stability.

Keywords: Ensemble clustering; cluster stability; F-measure; co-association matrix; genetic algorithm

Introduction

Data mining is known as non-trivial, unknown and potential information extraction from data with large amounts (Frawley et al., 1992; Hand et al., 2001). Data mining aims to extract concepts or available knowledge in the data so that this knowledge to be accessible and understandable and can also be used for future decisions. Data mining purposes is generally divided into several groups:

Forecast: the group aims to build a model to predict the values of certain properties. Input data for predictive modeling includes two types of attributes or variables: a. explanatory variables, the attributes that are used to predict and b. target variable, the attribute whose value should be predicted. This task can be divided into two smaller categories: classification and regression. Classification is used to predict the value of a discrete attribute. In contrast, regression is used to predict the value of a continuous attribute.

Association analysis: the aim is to produce a set of rules that describes the set of attributes associated with each other. For example, it can be used to identify the products that are purchased often by customers.

Clustering: it aims to identify the similar data groups so that the same data to be in a category and data point to be in a category and compared to the points in other categories to be noticeably more like.

Anomaly detection: The purpose of this group is to detect anomalies (namely, the number of data points that are very different compared to the rest of the data). Detect attacks on computer networks is an example of this task.

In general, data mining algorithms are divided into several categories that two main categories are presented below.

Classification: It means the prediction of labels for data based on the previous labeled data (Han et al., 2011).

Classification is a process to find a model (function) that describes the data and recognize their classes. To achieve this goal, the model is used that can obtain the unspecified classes label. The expressed model is achieved based on the analysis of a set of training data (data that their classes label is determined).

Clustering

Unlike classification that analyzes the label of data, clustering analyzes the without label classes data (Witten et al., 2005; Han et al., 2011). In general, class labels on test data does not exist simply because they may not be known. Clustering can be used to produce such labels. Objects that are clustered and grouped based on maximizing similarity within classes and minimizing similarity between classes. Namely, clusters of objects are formed in a way that objects within a cluster have highest similarity compared with each other, but the most differences with objects of other classes. Each formed cluster can be recognized as a class of those objects that the laws can be gained by them. Among clustering algorithms it can be pointed to hierarchical clustering algorithms, self-organizing map, K-Means, k-Medoid and so on.

Indeed, clustering refers to finding the structure of the data sets that are not grouped and categorized. In other words, it can be said that clustering is to place the data in groups in which members of each group are similar from a certain angle. The similarity of

members of a cluster according to a criterion is more than a threshold and the similarity of a cluster with other clusters is less than this threshold. Calculating the distance between two data is very important in clustering. Various criteria have been set for considering the clustering and evaluating its performance and many researchers through different ways have assessed them by using these criteria. By calculating the distance between two data it can be found that how much these data are close to each other and accordingly we put them in a cluster. The similarity criterion can be distance, mutual Information, covariance, etc. (Jain et al., 1988; Frigui et al., 1999; Theodoridis et al., 2003).

Significance of study

Clustering is a way to classify patterns into the various groups. These patterns can be observations, data items or features vector and the mentioned groups will be the same clusters. As noted in previous sections, clustering issues are considered as important tools in exploration and data analysis such as exploratory patterns analysis, decision-making, machine learning and so on which includes topics such as data mining, retrieval of documents, image segmentation and pattern classification (Jain et al., 1999).

Sometimes this division into several groups would be extremely difficult. One reason for this complexity is that a partition without ambiguity does not exist for some data sets or cannot be made by the man himself (Saha et al., 2008). But even in some datasets that this problem does not exist, some clustering algorithms fail. This is because in most of the existing algorithms clustering are based on only a function of internal evaluation. In fact, this function is the objective that measures the properties of the inner partition. These properties can be the separation between clusters or density within each cluster. By presenting different clustering algorithms on datasets, different partitions of data are obtained that the most appropriate partition should be selected based on the function of purpose. Totally, this study aims to provide a method to optimize ensemble clustering

on data set.

Objectives of the study

- Promoting the clustering process
- Increasing the accuracy in clustering
- Reducing the effect of noise and outlier in clustering
- Improving clustering in common data sets in data mining
- Offering a comprehensive approach to use in all applications

Literature review

For the first time Jardinand Van Rijsbergen in 1971 through several tests have proved that the efficiency of information retrieval systems can be improved by using documents 'clustering and it is also expected that the efficiency of these systems to be increased by using clustering. Because research results of Willett (1985) showed that the clustering considers the relationship between the documents of set and the relevant documents that by searching for the best fit with questions are placed at the end of list constitute a group with other related documents and this lead to the retrieval of them at a higher level, then the performance is increased. Clustering hypothesis is based on the efficiency improvement. This hypothesis states that the documents related to the question compared with non-relevant documents have more tendencies for being similar. Therefore, they are placed in a cluster (Jardin et al., 1971). If this hypothesis issued to a set of specific documents, it can separate the relevant and non-relevant documents well.

Tombrosin 2002 obtained a formula to calculate the similarities between documents and called it sensitive to questions similarity criterion. The criterion is improved by Valizadeh and colleagues in 2004. Valizadeh by using N nearest neighbor has proved

that his proposed criterion act better compared to the other criteria.

Clustering methods

In general, single-objective clustering algorithms can be divided in several respects including Exclusive or Hard Clustering and Overlapping or Soft Clustering

At exclusive clustering after clustering, each data belongs to a cluster such as K-Means clustering. But in overlapping clustering after clustering a degree is attributed to each data for each cluster. In other words, a data with different ratios can belong to several clusters. Fuzzy clustering is an example of that.

Hierarchical clustering: in hierarchical algorithms, data are placed in different categories and groups based on the criterion of similarity.

Clustering by partitioning method: partitioning clustering algorithms unlike the hierarchical methods that offer a clustering structure only show one division of data. Partitioning methods usually through optimization of an objective function produce clusters locally (on a subset of data) or globally (on all data).

Clustering based on density: These methods are based on the fact that clusters are areas of data space with high-density separated by areas with lower density. These algorithms have the ability to manage noise and often do the search once on the data.

Grid-based clustering: In this method, at first data space is divided into smaller units called Grid. Then partitioning is done on the resulting spaces and ultimately, clustered spaces lead to clustered data.

Kernel-based clustering: Kernel-based clustering methods are a conversion of dataset to a new space with high-dimension so that non-linear relationships between points may

convert the simpler relation.

Ensemble clustering methods

Ensemble clustering algorithms combine several clustering algorithms to finally reach an overall clustering system. These algorithms can be used in multifunctional clustering to combine several algorithms. Ensemble clustering includes two steps of producing and combining results to create final clusters. One of the common methods of synthesizing the results refers to using the co-association matrix. Since the initial clusters do not have appropriate sustainability, therefore, they should first be assessed by one of the clustering evaluation criteria and the most stable clusters should be involved in the co-association matrix. In this study, the Improved F-measures used to assess the stability of the clusters.

Different methods of Validation

A wide range of validation methods are presented in study of the clusters analysis (Rezaee et al., 1998; Abul et al., 2003; Tbshirani et al., 2005). A general classification divides these methods into two categories:

Internal validation

In this method the clustering algorithm results are evaluated in terms of quantities that include vectors of data (proximity criterion). The features of this validation include:

- Various solutions with regard to the goodness evaluate the match between each clustering and data.
- Approaches include: intra-cluster similarity, inter-clustering separation
- Criteria of CHI Index, Silhouette Score Dunn's Index are examples of this validation.
- The comparisons only possible between clusters that are made of a model or a metric.

- Often make assumptions about the structure of clusters.

These methods are performed based on an evaluation function. The assessments are based on the criteria and properties used in the process of clustering without any external supervision and additional information and are usually used after the completion of the clustering.

External validation

The result of clustering algorithm is estimated based on a predetermined structure that indicates the matching rate between output of clustering algorithm and the structure. Specifications of validation include:

- It evaluates the rate of conformity between the current solution of clustering and predetermined reference clustering.
- Approaches include counting matching pairs of clusters, sorting such clusters, using methods of information theory
- Among examples of this validation, we can refer to the criteria of Jaccard, Rand, NMI, F-measure. These criteria are not applicable in the real world, which is without supervision, because they need the real label of data in their calculations.

Generally, the results of applying clustering algorithms on a data set according to the selection of the parameters of algorithms can be very different from each other. Since the clusters generated by clustering algorithms are very different, the optimized clusters should be chosen for the final combination. In order to estimate the optimal clusters, several validation indexes are presented which aim to find clusters, which have the best fit with desired data. In this regard, different optimization algorithms can be used to find optimized clusters. In this way, at first fitness function should be created and then it should be optimized by optimization algorithms. The genetic algorithm is used as

optimization algorithm in this study. The genetic algorithm is a replication-based popular technique that most of its sectors are selected by random processes.

Consensus Clustering

In consensus clustering the results of several clustering are combined to finally achieve a single clustering. Consensus clustering algorithms often produce better clustering. Consensus clustering tries to find an ensemble clustering that cannot be produced by any other clustering algorithm. These methods lead to the production of stable clusters.

Computing cluster stability

Sustainable clusters are clusters that in different clustering on subsets obtained from various sampling have had the highest repeat. A stable cluster refers to the cluster that if several other clustering methods to be implemented on the data sets, more likely this cluster will not be seen.

Proposed method

In this proposed method, instead of using all clusters in the ensemble clustering, the set of clusters that are more stable are used. Choosing the most stable clusters is done through the F-measure stability standard. Therefore, first the primary clustering is done using different algorithms.

This can be done through data sampling, using different clustering algorithms, with a subcategory of the characteristics or choosing different parameters for a clustering algorithm.

In the next stage, the resulted clusters are evaluated to determine each cluster's quality. To evaluate each cluster, the F-measure stability criteria are used. After calculating the stability of the clusters, the most stable among them are chosen. To choose the most stable, first a threshold amount must be defined and each cluster with stability higher

than the amount of the threshold for the next stage should be considered to be combined with each other and the final clusters are reached.

Evaluating the clusters

A stable cluster is the one that if several other clustering methods were to be implemented on that set of data, that cluster is most probably to be seen again. In other words, stable clusters are the ones that have the most frequency of appearance in different clustering with the reached subcategories from different samplings.

To achieve consensus, the dispersion factor must be heeded. To have the most dispersed consensus, the sampling method is utilized. Here, the sampling is done on 80 percent of the data. 80 percent of the data is randomly chosen and the algorithm is applied on that 80 percent. Some of the data in this partition has the "-" quantity. The "-" quantity means that this data's cluster is not clear. To find their cluster, first the cluster's centers must be found and then, the data that are not labeled are appointed to the closest cluster. Now to calculate the cluster's stability, the clustering algorithm must be applied on the same set of data once again and by applying clustering algorithms, create as many new clusters as the reference set and then calculate the stability of the primary clusters. Each cluster from the primary clusters is divided to its partitions and then the stability of each of the partitions from the first set is calculated.

Then, the sampling of new sets of data is created and different clustering is done on that set to reach as many clusters as the number of the reference sets. To reach stability (C_i) (meaning the stability of cluster C_i), first the amount of the F-measure for this cluster

with all the other clusters in this clustering and their average must be calculated.

Input:

D - a dataset $\{x_1, x_2, \dots, x_n\}$

K - maximum number of Ensemble Clusters

$|RS|$ - numbers of clusters in new partition (Reference Set)

C_i - cluster i from Ensemble

RS_j - cluster j from ReferenceSetcluster i

Output:

Stability(C_i)- Stability of C_i

Require:

Resample D to obtain the perturbed data set D' ;

Run K-Means or other clustering algorithms ($\text{cluster}(D, K)$)

over D to obtain $P'(D)$;

Re-labeling $P'(D)$ to $P(D)$;

1. $\text{stability}(C_i)=0$;
2. For $j:=1$ to RS do
3. $\text{stability}(C_i)=\text{Fmeasure}(c_i, RS_j)+ \text{stability}(C_i)$;
4. end for

$$5. \text{ stability}(C_i) = \frac{\text{stability}(C_i)}{|RS|};$$

The algorithm for calculating the stability of cluster C_i as propriety function

There are three types of F-measures fir clusters:

1. Simple F-measure
2. Maximum F-measure
3. Modified F-measure

Simple F-measure

One of the main criteria to evaluate a clustering algorithm is the amount of its validity in the test set. This criterion is a logical efficiency criterion because it states the percentage of a model's success in comparison with the examples whose classes are not accessible.

The amount of the similarity of two partitions is calculated through the famous equation of F-measure criterion. Now the way the F-measure criterion is calculated must be described. This criterion, which in this thesis is considered to evaluate one partition, is the F-measure criterion 1.

$$FM(P, L) = \max_{\tau} \sum_{i=1}^{K_P} \frac{2 \times N_i^P \times \left(\frac{N_{i\tau(i)}^{PL}}{N_i^P} \times \frac{N_{i\tau(i)}^{PL}}{N_{\tau(i)}^L} \right)}{N \times \left(\frac{N_{i\tau(i)}^{PL}}{N_i^P} + \frac{N_{i\tau(i)}^{PL}}{N_{\tau(i)}^L} \right)} \quad (1 - 1)$$

Here, K_P is the number of the clusters of the partition P , N_i^P is an indication of the number of the existing data in cluster i of the partition P , N_{ij}^{PL} shows the number of existing data in cluster j from the partition L , N_{ij}^{PL} indicates the number of the data that are both in cluster i of the partition P and cluster j of the partition L . N shows the number of the total data and τ is a permutation of numbers one to N . If the two partitions P and the label L were to be completely similar, then the amount of the FM, meaning one is showed and if the two partitions were completely different, the amount is zero.

Maximum F-measure

To calculate the amount of the maximum F-measure between two different clustering, first the cluster that has the maximum amount of similarity with the primary cluster must

be found in the cluster that is related to the reference set. After finding that cluster, it must be considered as C^* and their data must be considered as complements. In the end, the F-measure between the two clusters is reached.

Modified F-measure

The modified F-measure is similar to maximum F-measure. To calculate the modified F-measure between two different clustering, first the cluster that has the maximum amount of similarity with the primary cluster must be found in the cluster that is related to the reference set. After finding that cluster, it is considered as C^* and the other clusters and data are each appointed to an independent cluster.

Choosing the stable clusters

Now the different clustering are put in order of the amount of their stability. Based on previous researches, the best results are reached by using 50 percent of the most stable clusters (Parvin, 2013). Therefore, the best 50 percent of the clustering is reached by putting the clusters in order.

After calculating the amount of each cluster's stability and the stable clusters were found, in the next step, the clusters are given weights. Instead of having all the clusters participating in the final result to a probability, each cluster is appointed with a weight (probability), which is the same as the amount of that cluster's stability.

Goal and fitness functions

The goal function is used to present a criterion of the way the subjects and their efficiency work. The optimization issues are in the form of minimizing or maximizing. In the optimization problems of minimizing, the most proper members must have the least numeral amount of the goal function. Also, in optimization problems of the maximizing kind, the most proper members must have the highest numeral amount of

the goal function. The other function is the fitness function, which converts the amounts evaluated by the members of the goal function using conversion functions into amounts that are called chromosomes' amount of fitness. In most meta-heuristic algorithms, including the genetic algorithm, the primary answers are chosen randomly.

In this case, it is an ensemble clustering in which the primary answers of a binary string are as many as the stable clusters. The amount of one in a gene means that this cluster takes part in the fitness function and the amount of zero means that that cluster is not participating in the final ensemble.

Here, the act of choosing clustering is done in two phases. In the first stage, an evolutionary algorithm tries to find the subcategory of the clustering with the most amount of stability. This goal is reached at the side with choosing the semi-stables in the step before the evolutionary algorithm. The second stage is when the most diverse clustering is chosen and it is evidently shown in the efficiency function of the evolutionary algorithms.

These evolutionary algorithms have a bit string chromosome as long as the number of the existing clustering in the final consensus in the part where different clustering is produced. Each of the genes in this chromosome can be one or zero. The chromosomes are like bits. Number one indicates that there is a clustering with the number of that gene among the chosen clusters and zero means that there is no clustering with the number of that gene among the chosen clusters.

To calculate the amount of the fitness function in this evolutionary algorithm, the amount of the chosen clustering's dispersion is calculated. For instance, if the number of the stable clusters was to be 6, the length of this chromosome equals 6. Using the genetic

algorithm, the following chromosome is reached.

1	1	1	1	0	1
---	---	---	---	---	---

Figure 1: A candidate solution (chromosome)

In this chromosome, the clustering with the amount of 1, are given the right to participate in the final correlation. Therefore, only the fifth clustering is not participating.

Then, the correlation matrix related to the final clustering is created. The amount of the correlation of the two entries x and y is calculated as follows.

$$Co(x, y) = \frac{Co_{min}(x, y)}{Co_{max}(x, y)}$$

$$Co_{min}(x, y) = \frac{1}{B} \sum_{i=1}^B \sum_{j=1}^{k_i} \min(p_j(x), p_j(y))$$

$$Co_{max}(x, y) = \frac{1}{B} \sum_{i=1}^B \sum_{j=1}^{k_i} \max(p_j(x), p_j(y))$$

In the last step, the reached correlation matrix from the optimal third consensus is considered as a similarity matrix. Therefore, a hierarchic clustering algorithm is considered as the final consensus function and the takes the reached correlation matrix as entry and gives the final adaptive clustering. The amount of the fitness function for this matrix is reached through the following equation:

$$FitnessFunction = 0.5 - \frac{\sum_x \sum_y abs(Co(x, y) - 0.5)}{N^2}$$

In this paper, the primary clusters for each set of data include 50 clusters. Each of these

50 clusters were reached by the six methods of K-Means clustering method, single link, average link, full link, the FCM algorithm and the EAC algorithm in Matlab software. Each of these six methods was used with the primary parameters to produce 50 clustering. In presenting the results, the methods were always implemented for an average of 20 times. Therefore, each result is an average of 20 times of different implementations.

The data generation

The proposed algorithms are evaluated on four data sets which including iris, Half ring, wine and Nglass. Wine contains data on the type of wine (grape juice) which were made with three different types. Models are described by thirteen properties that wine exists in all three types. Iris data set includes 150 samples in 3 categories of 50 divisions in which each category is related to a type of Iris plantain this data set, each sample has 4 features that make data set to be converted into 3 clusters including the data about Glass, data on the chemical components constituting 6 types of glass and each sample has 9 features. These data sets are from UCI Machine Learning Repository. The features of these four data sets are given in table 1.

Table 1. Features of data sets

Name of sets	Number of data	Number of Properties	Number of clusters
Iris	150	4	3
Wine	178	13	3
Half ring	400	2	2
Nglass	214	9	6

Providing data and results is as a function of the parameters.

Results

Community of initial clustering for each dataset includes 50 clustering. Each of the 50 clustering is obtained by one of the six clustering methods of K-Means, single link,

average link, complete link, FCM algorithm and EAC algorithm in software of Matlab. Each of these six methods is used with different initial parameters to produce 50 clustering. The average of 20 times implementation is always used to present results. This means that every result has been an average of 20 times of different implementation. An Improved F-measure issued for evaluation of the final clustering. The final table is as follows:

Table 2.The obtained experimental results

	data	Iris	Wine	Half ring	Nglass
	logarithm				
Simple F-measure	K-means	31.93	35.15	11.02	44.98
	Single Linkage	39.25	46.07	41.18	44.21
	Average Linkage	29.06	42.49	42.76	40.06
	Complete Linkage	30.03	40.25	29.47	42.11
	FCM	26.73	33.26	11.02	41.74
	EAC	45.52	45.94	43.24	39.27
Max F-measure	K-means	30.8	31.22	11.4	44.5
	Single Linkage	39.34	46.28	42.11	46.76
	Average Linkage	28.42	42.26	44.18	48.19
	Complete Linkage	28.63	39.08	31.86	47.99
	FCM	26.81	31.76	11.14	43.87
	EAC	45.6	46.17	44.25	<u>48.29</u>
Improved F-measure	K-means	29.63	27.6	11.64	33.35
	Single Linkage	22.96	<u>46.3</u>	42.24	42.31
	Average Linkage	28.71	27.9	15.55	15.42
	Complete Linkage	27.87	31.23	26.07	26.8
	FCM	28.85	26.91	11.62	11.44
	EAC	<u>45.58</u>	46.17	<u>44.37</u>	44.34

According to the results of applying various algorithms on multiple data sets the following results are achieved:

The following figure is obtained by applying 6 different algorithms on data set of Iris.

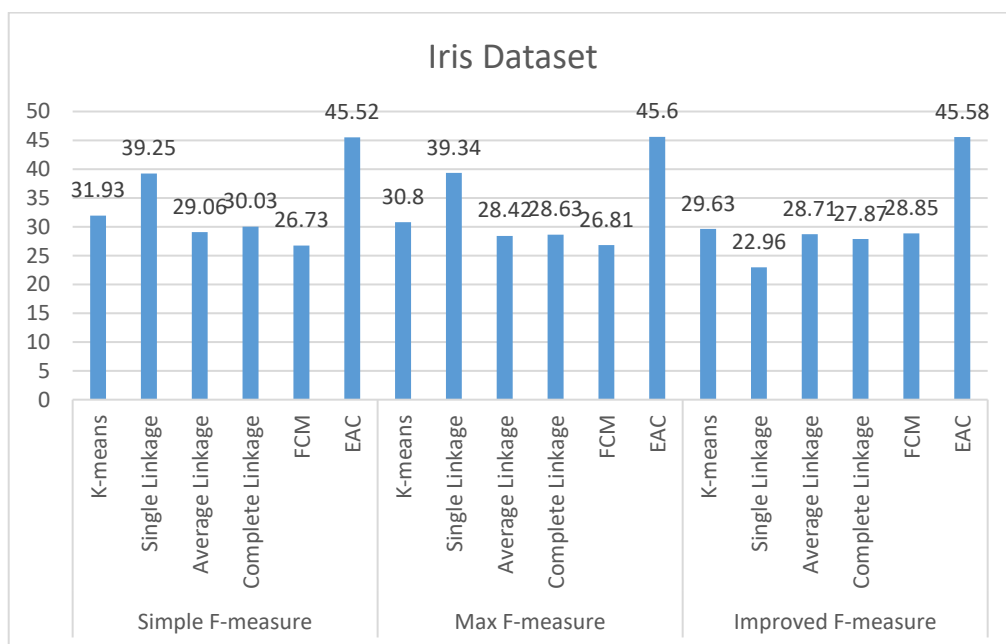


Figure 2.Applying different algorithms on data set of Iris

According to the above figure, it can be understood that the utilization of the evaluation criteria of modified F-measure, EAC base algorithm on Iris set leads to better result compared to other algorithms.

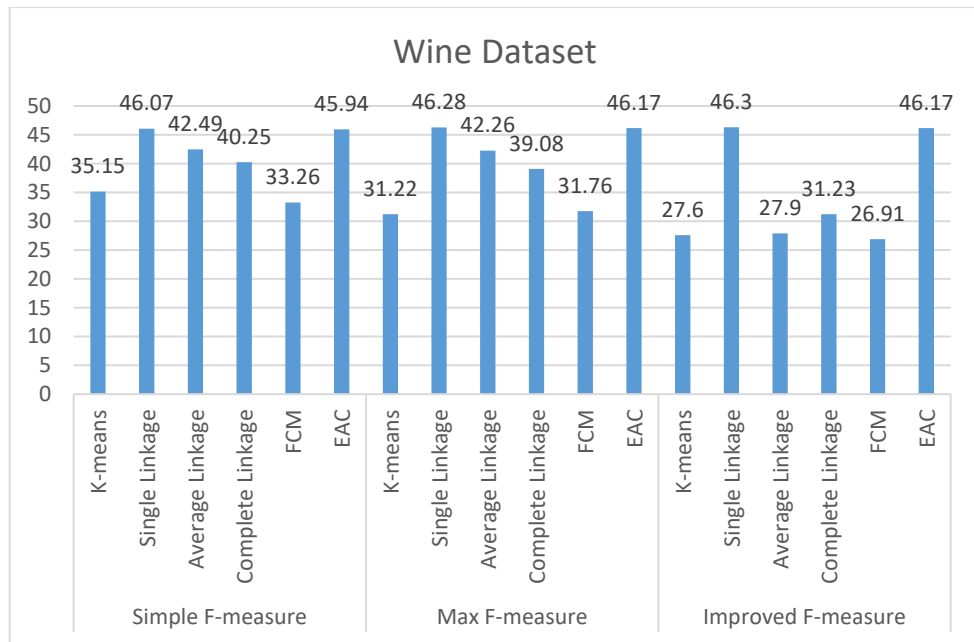


Figure 3.Applying different algorithms on data set of wine

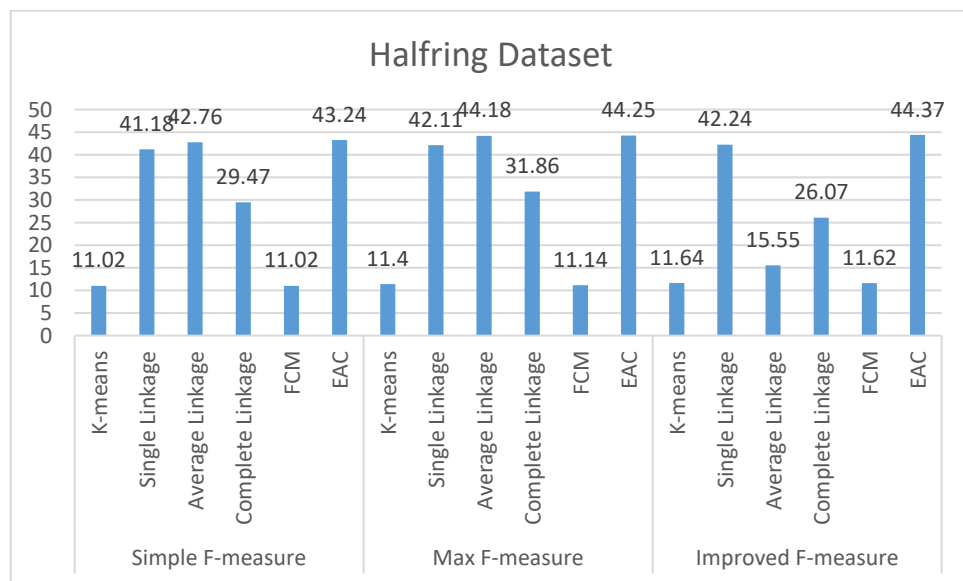


Figure 4.Applying different algorithms on data set of Half ring

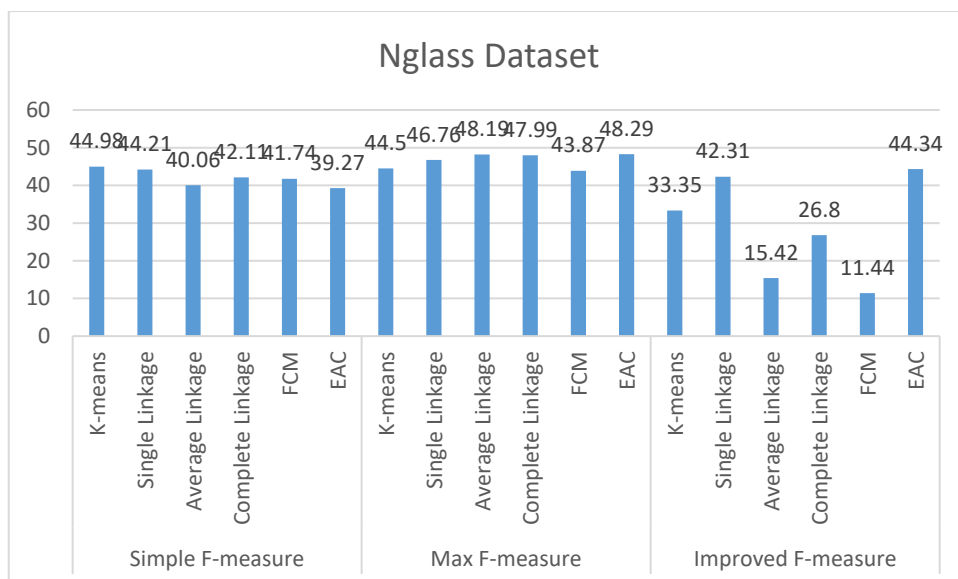


Figure 6. Applying different algorithms on data set of Nglass

Figure 3 shows that by using evaluation criteria of modified F-measure, Single Linkage base algorithm on wine set indicates better result compared to other algorithms. Figure 4 demonstrates the result of applying various algorithms on data sets of half ring. In this dataset the basic algorithm of EAC with evaluation criteria of modified F-measure have better results than the other data sets. In these three sets it can be concluded that better result is obtained by using evaluation criteria of modified F-measure compared to the two other criteria.

The result of applying above algorithms on Nglass set unexpectedly, with maximum F-measure and EAC algorithm has the best result.

This study examined the quality of various clustering and in this way at first, by using one of the clustering logarithms an ensemble clustering was created based on data set and then with respect to the F-measure it was tried to choose the most stable clusters among the results of primary clustering algorithms. Then the final clustering was done on the selected subset to obtain the final clusters. The proposed method is a dynamic one because of entering a combination of performed primary clustering into the final ensemble clustering based on each dataset. The obtained experimental results indicate the efficiency and ability of the proposed method in clustering information.

Conclusion

Due to the weakness of traditional ensemble clustering methods in creating stable clusters, this study considered the clusters to select the stable clusters based on Improved F-measure. As it was shown, Improved F-measure compared to the two other criteria had higher efficiency on most data sets. By considering the different algorithms it was concluded that better result was obtained on various dataset by using evaluation criteria of modified F-measure compared to the two other criteria and also the basic algorithm of EAC indicated better result on data.

Recommendations

Normalization of data is one of the necessary actions in using Euclidean distance. Since there is no guarantee for improving the quality of clustering when using the data normalization algorithms, usually proposed clustering methods provide their reports on the raw and abnormal data. Thus, another idea that could be investigated in future studies is to find a dynamic method for assigning a normalization method to each data set (Parvin, 2013).

Since the obtained final clusters may have weaknesses, a series of limitations and constraints should be applied on them. For example, in final clustering a data may not be placed in any cluster or there is not any data in a cluster. These restrictions should

be imposed on the clustering to be sure about the accuracy of performed work.

References

Abul, O., Lo, A., Alhajj, R., Polat, F., & Barker, K. (2003). Cluster validity analysis using subsampling. *IEEE International Conference on Systems, Man and Cybernetics*, 2:1435-1440.

Frigui, H., & Krishnapuram, R. (1999). A Robust Competitive Clustering Algorithm with Applications in Computer Vision: *IEEE Tran. Pattern Analysis and Machine Intelligence (PAMI)*, 21, 455-465.

Han, J., Kamber, M., & Pei, J. (2011). *Data mining: concepts and techniques*. Morgan Kaufman Pub.

Hand, D.J., Mannila, H., & Smyth, P. (2001). *Principles of data mining*, The MIT press.

Jain, A. k., & Dubes, R.C. (1988). *Algorithm for Clustering*, Englewood Cliffs, NJ: Prentice Hall.

Jain, A.K., Murty, M. N., & Flynn, P. J., (1999). Data Clustering: A Review, *ACM Computing Surveys*, 31(3).

Jardin, N., & Sibson, R. (1971). *Mathematical Taxonomy*, New York: Wiley.

Jardin, N., & Van Rijsbergen, C. J. (1971). The use of hierarchical clustering in information retrieval, *Information storage and retrieval*, 7, 217-240.

Parvin, H. (2013). *ensemble clustering with weighting along with clusters and attributes* (doctoral dissertation).

RazmeRezaee, M., Leleiveldt, B.P.F., & Reiber, J.H.C. (1998). A New Cluster Validity Index for the Fuzzy c-Mean, *Pattern Recognition Letters*, 19(3-4), 237-246.

Saha, S., & Bandyopadhyay, S., (2008). Application of Multiobjective Optimization for Data Clustering, Seminar (Machine Intelligence Unit, Indian Statistical Institute, 203, B.T. Road, Kolkata 700108, India)

Tibshirani, R., & Walther, G. (2005). Cluster validation by prediction strength, *Journal of Computational and Graphical Statistics*, 14(3), 511-528.

Tombros, A. (2002). The effectiveness of query-based hierarchic clustering of documents for information retrieval (Ph.D. Thesis), University of Glasgow.

Witten, I.H., & Frank, E. (2005). Data mining: Practical machine learning tools and techniques, Morgan Kaufman Pub.