Indonesian Journal of Electrical Engineering and Computer Science
Vol. 4, No. 3, December 2016, pp. 617 ~ 628
DOI: 10.11591/ijeecs.v4.i3.pp617-628

Clustering Large Data with Mixed Values Using Extended Fuzzy Adaptive Resonance Theory

Asadi Srinivasulu*, Gadupudi Dakshayani
Sree Vidyanikethan Engineering College, Tirupati, India
*Corresponding author, e-mail: srinu.asadi@gmail.com

Abstract

Clustering is one of the technique or approach in content mining and it is used for grouping similar items. Clustering software datasets with mixed values is a major challenge in clustering applications. The previous work deals with unsupervised feature learning techniques such as k-Means and C-Means which cannot be able to process the mixed type of data. There are several drawbacks in the previous work such as cluster tendency, partitioning, less accuracy and less performance. To overcome all those problems the extended fuzzy adaptive resonance theory (EFART) came into existence which indicates that the usage of fuzzy ART with some traditional approach. This work deals with mixed type of data by applying unsupervised feature learning for achieving the sparse representation to make it easier for clustering algorithms to separate the data. The advantages of extended fuzzy adaptive resonance theory are high accuracy, high performance, good partitioning, and good cluster tendency. This EFART adopts unsupervised feature learning which helps to cluster the large data sets like the teaching assistant evaluation, iris and the wine datasets. Finally, the obtained results may consist of clusters which are formed based on the similarity of their attribute type and values.

Keywords: *clustering, k-Means, C-Means, Fuzzy ART, unsupervised feature learning, extended fuzzy adaptive resonance theory*

1. Introduction

The existing approaches like Fuzzy ART and visual assessment of cluster tendency approaches can better able to deal with both numerical and categorical but it takes some more time to compute, to reduce the time i.e, to improve the performance the extended Fuzzy ART came into existence and brief discussion about these approaches are mentioned in the next sections.

1.1. Brief Description about FART and ART

The problem of clustering[1] mixed type of data can be resolved by using the fuzzy ART in conjunction with the unsupervised feature learning where the fuzzy ART consists of three parameters like category choice, vigilance test and the learning rate. To select the closest prototype vector the parameter category choice was used and to compare this selected prototype with the winning prototype the vigilance test was used and the learning rate is an optional parameter that is used to measure the speed of learning. Clustering [1] is the problem of grouping the unlabelled data items into classes based on the similarity of the items, here the unlabelled data items are nothing but the data which is in the form of images, audios and videos.

Fuzzy ART was used to provide the higher or sparse feature demonstration to abtain the unknown structure and by using this Fuzzy ART the distinction among the numerical and non-numerical features will become less, whereas by using some of the traditional clustering [1] algorithms the distinction among them will make the failure of the mechanism.

Stephen Grossberg and Gail Carpenter developed the concept of network called the adaptive resonance theory which is used to process the information. Here the resonance theory describes about the neural networks by using the methods like supervised and unsupervised learning. It is also used to resolve the problems related to the pattern recognition[7].

The unsupervised learning[5] model is the basic ART system and it is the neural network which consists of two fields like comparison and recognition. It also consists of the two

Received September 23, 2016; Revised November 2, 2016; Accepted November 19, 2016

modules like vigilance test and the reset. The comparison and the recognition fields are used to perform the vigilance test where the comparison field is used to select the most similar input vector and the recognition filed is used to compare this selected input vector with the prototypes that already exists. The most similar prototype will be selected based on the weights of the vectors for the input and the matched prototype. The adaptive resonance theory neural network was used to form the clusters by using this matched prototype, the input vectors are assigned into any one of the clusters based on the similarity of the features in the matched prototype. It can be known that the features which are clustered onto one group in the prototype will be useful to form the clusters to this new input pattern.

By using the ART network the input data was first trained to form the clusters. The training was given to the data based on the comparison among the input vector, vigilance parameter and the recognition field. The prototype which passes the vigilance parameter is taken to train the input data and it is grouped into classes as per the prototype which is considered to train the data.

The weights of the successful acceptance neuron are balanced towards the elements of the data vector. Something else, if the match level is beneath the carefulness parameter (i.e. the data vector's match is outside the ordinary expected reach for that neuron) the successful acceptance neuron is reserved and a search for system is completed. In this search methodology, acceptance neurons are verified one by one using the reset capacity until the vigilance parameter is overcome by an acceptance match. Specifically, at every cycle of the inquiry method the most dynamic acknowledgment neuron is chosen and afterward exchanged off if its performance is under the vigilance parameter (take note of that it along these lines discharges the remaining acceptance neurons from its difficulty). Slow and fast are the two fundamental methods that are used for data training. The difference among the slow and fast learning methods can be known based on the time. The slow learning methods will take more time to train the input data recognized pattern and the fast learning method will take less time to train the input data to form the clusters. The fast learning method is an effective method that is used for several tasks where as the slow learning method is only used only if there is a continuous change in the input vector? there are different types of ART networks like: ART1, ART2, ART3, Fuzzy ART, ARTMAP and the Fuzzy ARTMAP

1.2. Basic Techniques

The general clustering[1] methods what we are using previously can only deal with either numerical or non numerical data but not with the both, however the real world data sets are the combination of both numerical and non numerical data and also the images (heterogeneous data) where as the conventional clustering techniques cannot deal with heterogeneous data. One of the mostly used conventional clustering algorithms is k-Meansv [2] which cannot be able to deal with the heterogeneous data because the distance between the vectors of both numerical and non numerical data cannot be measured directly. Several algorithms have been proposed to deal with the heterogeneous data and those are of two types. They are:

1. Those that cluster heterogeneous data directly and
2. Those that cluster heterogeneous data based on feature transformation.

The unsupervised feature learning algorithm with k-Meansv [2] has three advantages and they are:

1. Only the k number of clusters is required, not any other parameters are needed;
2. Unsupervised feature transformation for the non numerical data is Robust and it is mutual information based.
3. The UFT Technique can be used to convert the numerical values into the non numerical values and it will not use the distance vector for clustering.

K-prototype and improved k-prototype results will have problem in differentiating the classes by the usage of algorithm which cannot provide the different distances between different numerical values, here the distances are nothing but hamming distance. The unsupervised feature learning [1] for aerial scene classification [3] may consist of various steps for clustering the areas such as forest, schools, and temples and so on. The four steps that are used is feature extraction, feature learning, feature encoding and the feature pooling.

To combine this hamming distance with Euclidean distance the parameter used is not optimal for the mixed dissimilarity measurement. Gauss-multinomial distribution is not appropriate for the numerical features shown and the KL-FCM-GM relies in this distribution. For the Latest technologies such as Clustering of high dimensional satellite images which exploit spatial context this kind of clustering technologies such as k-Means[2] and other fuzzy ART[4] techniques will be used. The feature extraction strategy for classifying the high resolution images was proposed by the authors Bruzzone and Carlin.

The image is divided into several segments and the obtained segments were used as spatial context for pixel. The hidden information and pattern from the segments have been collected and those are used as features of the image and it is used to compare it with other images and to identify the location based on the image. In the same way, the authors Shackelford and Davis generated the features based on the combination of both object and pixel based features for providing the image classification based on the object of an image. At first, for generating the classification labels textual and the spectral features were used. The data Extracted from the labels of soft classification, Statistics computed over the soft classification labels, spectral dimensions, and geometrical attributes related to the segments of the image are used as the features of classification. In both the cases of classification success depends only on the quality of the classification. Bellens performed the opening and closing morphological operations on image for generating the morphological profiles. For the generation of the pixel features the morphological attributes associated with geometrical attributes are combined with spectral measurement. To classify the images into various classes such as buildings, roads and plants all the above mentioned classification approaches were used.

In contrast with all the approaches mentioned above Unsalan and Boyer proved that the intermediate representation of the scenes based on single local line parameter was an efficient way to show the difference between the geo spatial locations. The different scene classification was based on the measurement of length, breadth, contrast and the distribution of the orientation provides the unique representation of the dimensions for different scene classification categories. In the same way Huang proposed an idea which is similar to the previous ones but it is based on the generation of direction lines for the pixel features. The similarities among the pixels at some distances were measured and also the calculations were performed on the orientation of the image to determine the classification lines. The measurements obtained from the length of the direction line related to the every pixel of an image were used as the vector of the feature. The thresholds which are determined heuristically are used to recognize the direction lines that are passes through the pixels of an image. For modelling the various sets of neighbour class's line based approaches are used effectively.

The BoVW-based approach has been used for various scene classifications. This BoVW approach is of two types and they are feature learning and feature encoding. In the process of feature learning, the features of an image as clustered and also the center of the clusters has been identified and they will form the visual words. Next in the process of feature encoding the features i.e, the low level features of an image are extracted and are mapped with the visual word that are formed by the identification of center of the cluster. The identified visual words of an image will form the new features of the image. To form the visual words in case of an image which is simple the mean and variance of the pixel intensity of an image was clustered. Some extra low level features such as orientation of an image, filter oriented responses, parameters of the line and the color of an image is used for generating the visual words. In both the cases to model the distribution of visual word the technique such as LDA which is nothing but Latent Dirichlet Allocation and an unsupervised learning[1][5][5] framework was used.

To group the visual words spatial pyramid matching kernel (SPMK) approach has been used and it is proposed by Lazebnik. Visual words are defined by the spatial pyramid representation of an image and they are computed at various scales and are combined to produce better scene classification based on the image. Yang and Newsam identified visual words for generating the distribution model of higher order of visual words.

2. Literature Survey

Data with numerical and non-numerical features can be unified to be purely numerical and then clustered effectively. The k-Means [2] can better deal with numerical data rather than the non-numerical, so that the k-Means [2], [6] is used with the unsupervised, [1], [5] feature transformation. Heterogeneous data is the data which consist of both numerical and categorical features [7], [8].

To perform heterogeneous data clustering several algorithms have been proposed in and those techniques are of two types i) those that cluster heterogeneous data directly. ii) Those that cluster heterogeneous data based on feature transformation. To deal with the first type the hierarchical clustering**Error! Reference source not found.** and similarity based agglomerative clustering algorithms are used where SBAC uses the dissimilarity measurement which measures distance and density for numerical and non numerical data. By using UFL with k-Means [2], [6] the mixed type of which consist of both numerical and non numerical features are divided into two separate categories and then the unsupervised [1], [5] feature transformation was applied on the non numerical data to convert them into numerical. Finally the k-Meansb [2], [6] was applied on the numerical data to form the clusters.

The task of assigning a sentiment polarity to user-generated short documents to determine whether each of them communicates a positive or negative judgment about a subject. The method presented in this work is able to generate a sparse encoding of short documents in an unsupervised [1], [5] manner, without using any prior knowledge related to the context of the problem. Here the input documents are represented in a vector space model. The weight assigned to each term of the dictionary is computed using a weighted function. The weighting function produces a vector representation of the document and it is provided as training data to the GHSOM (growing hierarchical self organizing map) that learns a new representation for the input data. GHSOM generates maps that hierarchically represent the distribution of the training data. The new set of feature vectors along with their respective labels constitutes the training data for an SVM classifier. Once the training phase ends, the classifier is able to assign a positive or negative label to each of the input vectors.

The following are the some other authors who worked in the approaches that are similar to the one which are mentioned above:

Pang and Vaithyanathan (2002) deals with corpus based methodologies by using the techniques in machine learning instead of depending o the previous approaches. The framework that is used for encoding the document is known as bag-of-words and the task of classifying the sentiment is treated as the task of categorization of task based on binary data values. They have proven that the SVM classification technique can work effectively compared to the other techniques and also better results can be obtained compared to others. Maas (2011) deals with an unsupervised feature learning [1], [5] approach and it is based on the Latent Dirichlet Allocation for generating the vector representations of the document. The supervised classifier deals with similar words and those are having similar representation. The supervised feature learning technique can be useful for clustering and the number of clusters may have to be known before the implementation of data set for the formation of clusters. Better results can be obtained by including the sentiment information in vector space representation approaches. Socher (2011) used an auto encoder which is semi supervised recursive for obtaining the vector representation of the documents. Vector representation was performed by softmax layers of neurons during the task of classification. It does not use any special lexicon and the representation of bag of word. Glorot (2011) constructed a profound neural network [9] for the input vectors to learn new representations**.** This network is pretrained by a heap of denoising encoder's i.e, auto encoders and it uses the linear rectified units. The data cases with the usage of presence or absence binary vectors are encoded for every word that is present in the dictionary. The network is used for mapping each and every input vector into a different feature space in that each input data is finally classified by using a linear support vector machine. This framework is used for area adaptation; its channel is fundamentally identical. Instead of using a complex architecture a simple model is used with both linear and non linear classifiers.

3. Proposed Approaches

3.1. Extended Fuzzy Adaptive Resonance Theory(EFART)

The EFART is the technique that is used to form the clusters in any kind of data like binary, analog and the numerical. To cluster the data and to validate the obtained number of clusters the fuzzy ART [4] is used in combination with the co-VAT[11] which is referred as Extended Fuzzy ART [4] and those approaches are discussed.

3.1.1. Fuzzy ART

Fuzzy ART is the approach developed by carpenter in the year 1991 which consists of three parameters such as category choice and the vigilance parameter. Generally the fuzzy ART [4] is the extension of ART1 which may consist of two operators like intersection and union; these two operators are better able to deal binary data rather than analog and the mixed type of data. To overcome the draw backs that are present in this ART1 the fuzzy ART [4] was developed in which the ∩ and ᵛ operators in ART1 are replaced with min and max operators.
There are several differences in fuzzy ART [4] compared to the ART1 and they are: (1) not only binary but also several other kinds of data can be processed by using FART (2) consist of only one connection of weight vector and (3) it consist of not only the vigilance parameter (ρ) but also the category choice (α) and the learning rate (β).

Fuzzy ART [4] is the knowledge theory which may consist of two layers like F1 and F2 where the F1 is the input layer and F2 is the output layer. The basic Fuzzy ART [4] Algorithm consists of three input parameters such as: category choice (β>0) and the vigilance parameter ($0 \leq \rho \leq 1$).

The following are the steps involved in fuzzy ART algorithm:

Step 1: Initialization

Initialize all the parameters

Step 2: Apply Input Pattern

Let I: = [next input vector]

Let P: =be the set of candidate prototype vectors

Step 3: Category Choice

Select the training points to train the data and form clusters of the input data according to the training points.

Step 4: Vigilance Test

The prototype which is selected in the previous step is compared with the one which is already present in the database by using the below mentioned formulae:

norm(min(a, wij(J_{max},:)),1)>=norm(a,1)

If the prototype passes this vigilance test then it is selected to form the clusters, otherwise this prototype is removed and another one is selected to compare it with the winning prototype.
If no prototype passes the test then the new prototype will be created to form the clusters and the process will be repeated from the step 2.

3.1.2. co-VAT

Let us assume an $m \times n$ matrix **D**, where the elements of **D** are pair-wise dissimilarities between m row objects *or* and n column objects *Oc*. The union of these disjoint sets are ($N = m + n$) objects *O*. Clustering tendency assessment is the process by which a data set is analyzed to determine the number(s) of clusters present. In 2007, the *co-Visual Assessment of Tendency* (co-VAT) algorithm was proposed for rectangular data such as these. co-VAT is a visual approach that addresses four clustering tendency questions: i) How many clusters are in the row objects *Or*? ii) How many clusters are in the column objects *Oc*? iii) How many clusters are in the union of the row and column objects *Or*∪*Oc*? And, iv) How many (co)-clusters are there that contain at least one of each type? co-VAT first imputes pair-wise dissimilarity values among the row objects, the square relational matrix **D***r*, and the column objects, the square relational matrix **D***c*, and then builds a larger square dissimilarity matrix **D***ruc*. The clustering questions can then be addressed by using the VAT [10] algorithm on **D***r*, **D***c*, and **D***ruc*; **D** is reordered by shuffling the reordering indices of **D***ruc*.
Steps for co-VAT [11] algorithm:

1) Input the rectangular matrix D(m*n).
2) Build estimates of D_r and $D_{c.}$
3) Build estimate of DrUc.
4) Run VAT on $D_{rUc.}$
5) Form the co-VAT ordered Rectangular dissimilarity matrix D*.
6) Forms a Reordered dissimilarity matrices images of D^*, D^*_r , D^*_c , and $D^*_{rUc.}$

4. Results of Proposed Approaches

The following are the sequence of steps involved in implementing the extended Fuzzy Adaptive Resonance Theory:

1. Generate Training Points
2. Fuzzy Adaptive Resonance Theory
3. Co-VAT(Visual Assessment of cluster tendency)
4. VAT(Visual Assessment of cluster tendency)
5. Comparing Results

4.1. Generating Training Points

The data points will be represented as np and the clusters as nc, the random function takes nc as input argument and it randomly generates the training points. To generate the number of clusters the rand () function is used as follows:

Clusx=rand (nc, 1)

The rand () function mentioned above will generate the matrix of random values; if nc value is equal to 3 then it generates the Clusx consisting of three rows and one column.
The points will be generated by using the logic mentioned:

Punt (point, 1) =Clusx (cc, 1) + dcx

By using the equation as mentioned above the training points will be generated and are stored in the file “points.mat” and it can be done by using the command save as mentioned below:

Save points punt –ascii

The dcx value which is used to generate the points can be known by using the equation as follows:

dcx=0.1×randn (1)

The training points, number of clusters will be known by running the genpun in Matlab which uses the function rand to generate the training points randomly. The training points will be generated randomly by using thigenpun as shown in the Figure 1.

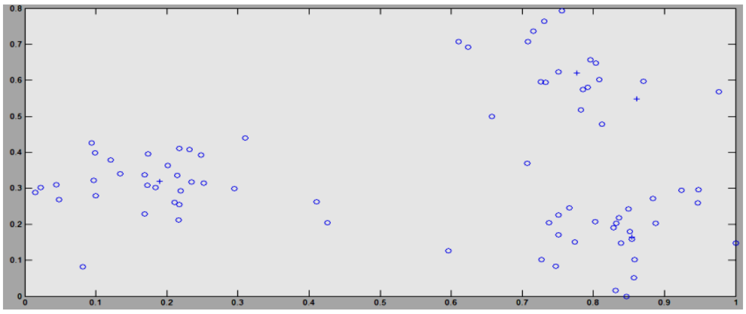

Figure 1. Output of Genpun

In the Figure 1 mentioned above it contains data points and the number of clusters where the training points are indicated with ‘0’ and the clusters are indicated with ‘+’ and are formed according to the value given to np and nc.

4.2. Fuzzy ART

The following are the steps involved in Fuzzy ART[6] algorithm:
Step 1: Initialization

Initialize all the parameters

Step 2: Apply Input Pattern

Let W_{ij}: = input data

Let np: =be the set of training points

Step 3: Category Choice

Select the training points to train the data and form clusters of the input data according to the training points.

$$\frac{\|min(\,a, wij(\,Jmax, :)),1\|}{\alpha + \|w_{ij}(j, :), 1\|}$$

Step 4: Vigilance Test

The prototype which is selected in the previous step is compared with the one which is already present in the database by using the below mentioned formulae:

$$\|\mathrm{min}(\,\mathrm{a}, \mathrm{wij}(\,\mathrm{Jmax}, :)),1\| >= \rho \times \|\mathrm{a}, 1\|$$

If the prototype passes this vigilance test then it is selected to form the clusters, otherwise this prototype is removed and another one is selected to compare it with the winning prototype. If no prototype passes the test then the new prototype will be created to form the clusters and the process will be repeated from the step 2.

Step 5: Update Matched Points

The matched data points will be moved to the matrix map as mentioned:

$\mathrm{map(point)} = \mathrm{J_{max}}$

The output of this Fuzzy ART will be as follows:

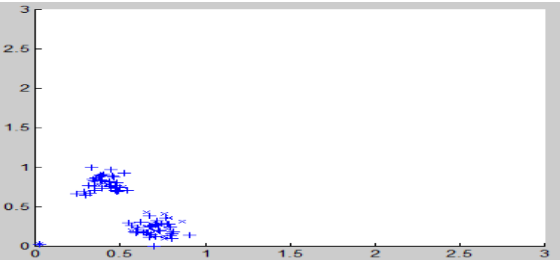

Figure 2. Output of Fuzzy ART

From the Figure 2 mentioned above it can be known that it consist of three different groups which means that the dataset (iris) consist of three numbers of clusters such as iris setosa, iris Virginia, iris versi-color.

The above mentioned image will represent the output of FART which represents the number of clusters as 3 by observing the image, consists of three different groups which are represented in blue color. These clusters are formed in between the range 0 to 1 with alpha as 0.001 and rho as 0.9.

4.3. Co-VAT Steps

1) Input the rectangular matrix D (m×n).
2) Build estimates of D_r and D_c.
3) Build estimate of $D_{r\cup c}$
4) Run VAT on $D_{r\cup c}$.
5) Form the co-VAT ordered Rectangular dissimilarity matrix D^l.
6) Forms a Reordered dissimilarity matrices images of D^l, D^l_r, D^l_c, and $D^l_{r\cup c}$.

4.4. VAT(Visual Assessment of Cluster Tendency) [10], [12]

Input: An n × n matrix of dissimilarities

$D=|d_{ij}|$, with $1\geq d_{ij} \geq 0$; $d_{ij}= d_{ji}$; $d_{ii} = 0$, for $1\leq i, j \leq n$.

Step 1: Set I=ɷ,j={1,2,….n} and P=(0,0,…,0)

Select(i , j) ε argpεj , q εj max{d_{pq}}

Set P(1)=i, I←{i} and J ← J- {i}

Step 2: Repeat for t=2,…..,n

Select(i , j) ε argpεj , q εj min{d_{pq}}

Set P(t)=j, update I←I∪{j} and J ← J- {j}

Step 3: Form the re-ordered dissimilarity matrix $\widetilde{D}$ =| $\widetilde{d_{ij}}$ | = |$d_{p(i)p(j)}$|,for 1≤i, j ≤n.

Output: A scaled gray-scale image I($\widetilde{D}$) so that max{ $\widetilde{d_{ij}}$} corresponds to white and max{ $\widetilde{d_{ij}}$} to black the following Figure 3 will represent the output of Co-VAT and the VAT is called from the Co-VAT.

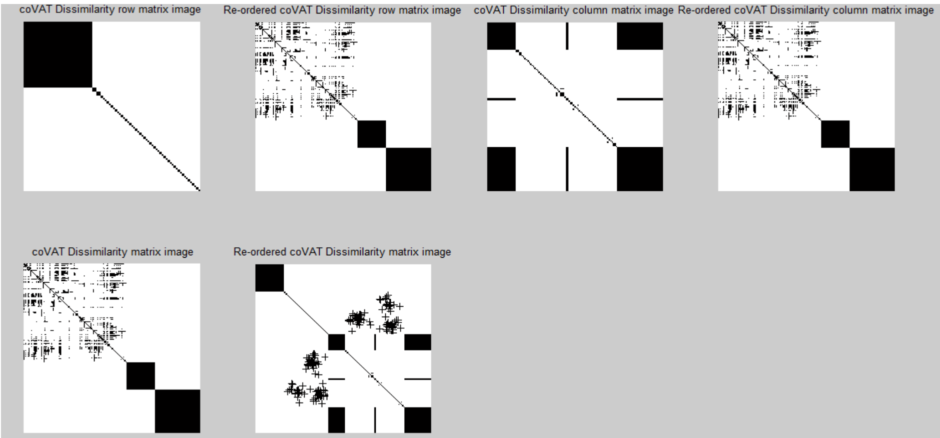

Figure 3. Output of Co-VAT

The number of clusters obtained from FART [13] will be validated by using Co-VAT and VAT [14] and from the Figure 4 above the reordered Co-VAT dissimilarity image consist of three dark blocks along the diagonal axis of an image which indicates that the iris data consist of three numbers of clusters as obtained from the Fuzzy Adaptive Resonance theory.

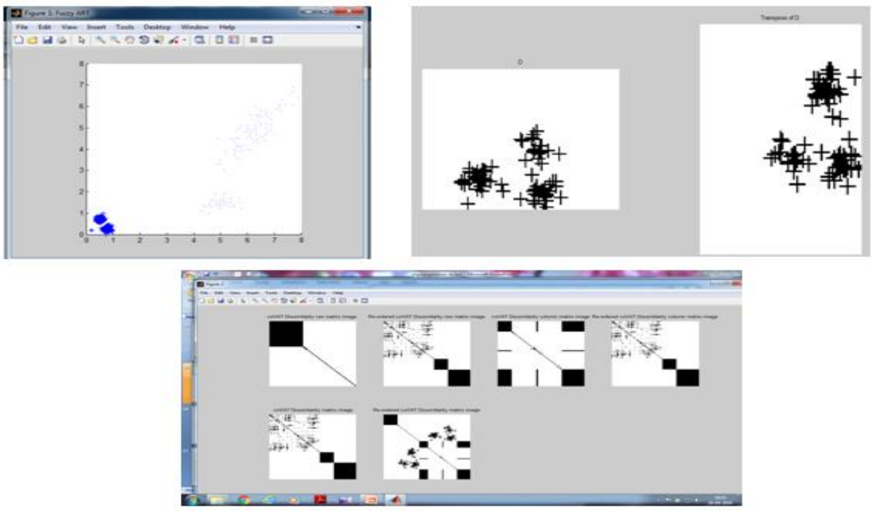

Figure 4. Output of Extended Fuzzy Adaptive Resonance Theory

4.5. Extended Fuzzy Adaptive Resonance Theory(EFART)
Input: Training points, Data set D
1. Apply Fuzzy ART to the data set D.
2. Let 'h' be the image represents output of Fuzzy ART.
3. Apply Co-VAT on Figure 'h'.
4. Apply VAT to obtain reordered dissimilarity matrix.

Output: Reordered dissimilarity image $R_{i.}$

The Figure 4 represents the number of clusters from the FART, VAT and Co-VAT; the EFART is the Extension of FART so that its output represents the number of clusters present in the data by computing all the three algorithms (FART, Co-VAT and VAT).

5. Comparative Study
The performance will be compared by measuring the time taken to compute the algorithms FART, VAT and Co-VAT by using the tic and toc commands in Matlab. The table mentioned below represents the performance of different algorithms like generating training points, fuzzy Adaptive resonance theory, visual assessment of cluster tendency and the co-VAT for different data sets like iris data, wine data and the teaching assistant evaluation data.

a. Iris dataset
Iris dataset may consist of five attributes such as SL, SW, PL, PW and the Class where the class is of three types such as Iris Setosa, Iris Versicolor, and Iris Virginica. The sepal length, sepal width, petal length, and petal width will be determined in centimeters. This dataset may consist of 150 instances where each class will have 50 instances.

This iris dataset was obtained from the UCI website and is available in this link to download *https://archive.ics.uci.edu/ml/datasets/Iris.* This dataset may consist of no missing values. The following may represent the summary statistics of the iris data (Table 1 and 2).

Table 1. Attributes in Iris Data

Attributes	Min	Max	Mean	SD	Class Correlation
sepal length	4.3	7.9	5.84	0.83	0.7826
sepal width	2.0	4.4	3.05	0.43	-0.4194
petal length	1.0	6.9	3.76	1.76	0.9490 (high!)
petal width	0.1	2.5	1.20	0.76	0.9565 (high!)

The class distribution value is 33.3% for each of the class among the three classes. In those five attributes five are both numeric and predictive attributes and the fifth one is class.

Table 2. Details of Iris Data

Data Set Characteristics	Multivariate	Number of Instances	150
Attribute Characteristics	Real	Number of Attributes	4
Associated Tasks	Classification	Missing Values	No
Area	Life	Date Donated	1988-07-01

b. Wine dataset
This wine dataset may gather from the chemical analysis if wine in Italy and is resulting from the three distinct cultivators. This analysis is done based on the information gathered from the three different types of wines. The initial dataset may consist of only 13 attributes.

This dataset consist of 13 attributes (Table 3) such as Alcohol, Malic acid, Ash, Alcalinity of ash, Magnesium, Total phenols, Flavanoids, Nonflavanoid phenols, Proanthocyanins, Color intensity, Hue, OD280/OD315 of diluted wines, Proline. The initial attribute in this wine dataset is class identifier based on which the data was clustered. This dataset consist of 178 instances and the three classes. The instances are divided among the three classes where the class 1, class 2 and class 3 may consist of 59,71 and 48 instances respectively.

This wine dataset is also available in the UCI website and here is the link to download this dataset *https://archive.ics.uci.edu/ml/datasets/Wine.* The class identifier value is in the range of 1-3 where there should be a possibility of forming three different clusters based on class id. All the attributes in this dataset are continuous and the statistical values are not available. The sample dataset is as shown Table 4.

Table 3. Details of Wine Data

Data Set Characteristics:	Multivariate	Number of Instances	178
Attribute Characteristics	Integer, Real	Number of Attributes	13
Associated Tasks	Classification	Missing Values	No
Area	Physical	Date Donated	1991-07-01

Table 4. Five Instances of Wine Data

Class id	Alcohol	Malic acid	Ash	Alcalinity of ash	Magnesium	Total phenols	Flavanoids	Nonflavanoid phenols	Proanthocyanins	Color intensity	Hue	diluted wines	Proline
1	14.23	1.71	2.43	15.6	127	2.8	3.06	.28	2.29	5.64	1.04	3.92	1065
1	13.2	1.78	2.14	11.2	100	2.65	2.76	.26	1.28	4.38	1.05	3.4	1050
1	13.16	2.36	2.67	18.6	101	2.8	3.24	.3	2.81	5.68	1.03	3.17	1185
1	14.37	1.95	2.5	16.8	113	3.85	3.49	.24	2.18	7.8	.86	3.45	1480
1	13.24	2.59	2.87	21	118	2.8	2.69	.39	1.82	4.32	1.04	2.93	735

c. Teaching Assistant Evaluation Dataset

This dataset may consist of 151 instances and five attributes such as size of the class, whether the semesters are summer or regular, type of course and its instructor name and the teaching assistant is whether native speaker or not (Table 5). The information comprise of assessments of showing performance more than three normal semesters and two summer semesters of 151 Teaching Assistant (TA) assignments at the Statistics Department of the University of Wisconsin-Madison. The scores were separated into 3 generally break even with estimated classifications ("low", "medium", and "high") to frame the class variable.

Table 5. Details of TA Data

Data Set Characteristics:	Multivariate	Number of Instances:	151
Attribute Characteristics:	Categorical, Integer	Number of Attributes:	5
Associated Tasks:	Classification	Missing Values?	No

This dataset *https://archive.ics.uci.edu/ml/datasets/Teaching+Assistant+Evaluation.* By observing the figure below it can be known that the time taken to compute fuzzy ART will be less than the time to compute co-VAT the graphical representation of the performance of all the algorithms for the iris dataset is as shown in Figure 5.

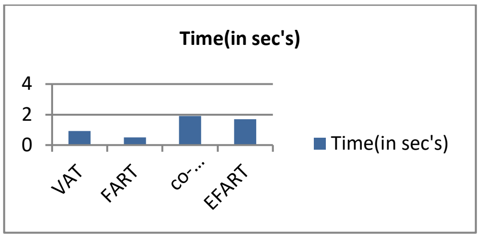

Figure 5. Time Taken to Compute FART, VAT, co-VAT for the Iris Data

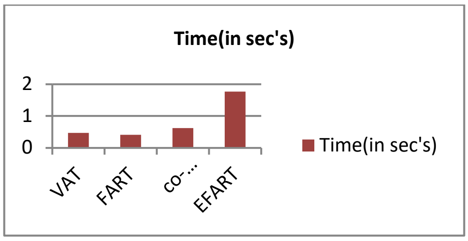

Figure 6. time taken to compute FART, VAT, co-VAT for the Wine Data

The graph mentioned above will represent that the time taken for fart will be lesser compared to the co-VAT and it is same for all the data sets (Figure 6). The graph mentioned below will represent the performance of algorithms for the wine data set. All the three data sets like iris, wine and the teaching assistant (Figure 7 and 8) evaluation data there exists only three number of clusters based on their categories. The number of attributes and instances will differ for the three data sets but the number of clusters will be formed are same.

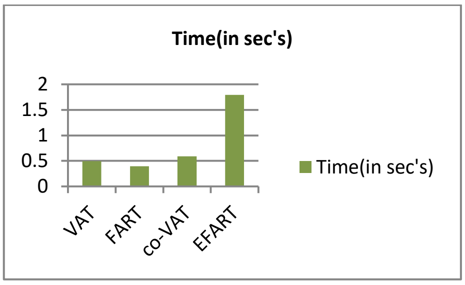

Figure 7. Time Taken to Compute Fart, Fuzzy, co-VAT for the Teaching Assistant Evaluation dataset

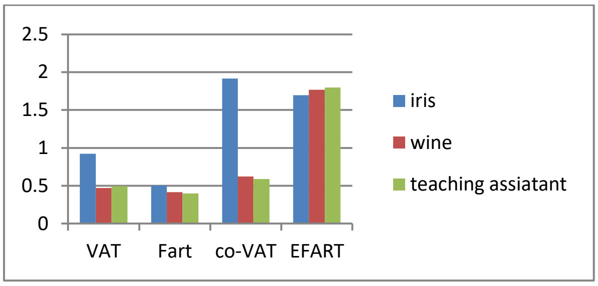

Figure 8. Time Taken To Compute Fart, Fuzzy, co-VAT for the Iris, Wine, Teaching Data

The graph that represents the performance of all the three algorithms for all the three datasets is as shown in Figure 8.

6. Conclusion

A new methodology known as fuzzy ART will be helpful to cluster the Data sets which consist of both numerical and categorical features. The early work deals with the techniques such as k-means, c-means and UFL which are capable to cluster only numerical data but not able to deal with combination of both numerical and categorical data due to which the technique fuzzy ART came into existence. The algorithm often named as FART will consist of three steps such as genpun, fuzzy and fuzzy ART to generate the number of clusters and the number of clusters will be validated by using the visual assessment of cluster tendency algorithms like VAT and the co-VAT. The Extended Fuzzy Adaptive Resonance Theory was obtained by integrating FART, Co-VAT and VAT and the comparative study for all the techniques indicates will be performed based on the time taken to compute the different algorithms for three different datasets like Teaching assistant evaluation dataset, Iris dataset and the wine dataset.

References

[1] Dao Lam, Mingzhen Wei, Donald Wunsch. Clustering Data of Mixed Categorical and Numerical Type with Unsupervised Feature Learning. *The Journal for Rapid Open Access Publishing*. 2015; 3(1).
[2] Min Wei, Tommy WS Chow, Rosa HM Chan. *Clustering Heterogeneous Data with k-Means by Mutual Information-Based Unsupervised Feature Transformation*. Entropy. 2015; 1(1).

[3] Anil M Cheriyadat. Unsupervised Feature Learning for Aerial Scene Classification. *IEEE Transactions on Geoscience and Remote Sensing*. 2014; 52(1).
[4] Gulsen Aydın Keskin, Sevinc Ilhan, Coskun Ozkan. The Fuzzy ART algorithm: A categorization method for supplier evaluation and selection. *Elsevier.* 2010; 1: 1235-1240.
[5] C Li, G Biswas. Unsupervised learning with mixed numeric and nominal data. *IEEE Trans. Knowl. Data Eng.* 2002; 14(4): 673-690.
[6] Z Huang. Extensions to the *k*-means algorithm for clustering large data sets with categorical values. *Data Mining Knowl. Discovery*. 1998; 2(3): 283-304.
[7] CC Hsu, SH Wang. An integrated framework for visualized and exploratory pattern discovery in mixed data. *IEEE Trans. Knowl. Data Eng.* 2006; 18(2): 161-173.
[8] CC Hsu, YC Chen. Mining of mixed data with application to catalog marketing. *Expert Syst. Appl.* 2007; 32(1): 12-3.
[9] TW Ryu, CF Eick. *Similarity measures for multi-valued attributes for database clustering.* Proc. Smart Eng. Syst. Design Neural Netw. Fuzzy Logic, Evol. Program. Data Mining Rough Sets (ANNIE), 1998: 25-29.
[10] JC Bezdek, RJ Hathaway, JM Huband. Visual assessment of clustering tendency for rectangular dissimilarity matrices. *IEEE Trans. Fuzzy Syst.*, 2007; 15(5): 890-903.
[11] Timothy C Havens, James C Bezdek, James M Keller. A New Implementation of the co-VAT Algorithm for Visual Assessment of Clusters in Rectangular Relational Data.
[12] JC Bezdek, RJ Hathaway, JM Huband. Visual assessment of clustering tendency for rectangular dissimilarity matrices. *IEEE Trans. Fuzzy Syst.,* 2007; 15(5): 890-903.
[13] GA Carpenter, S Grossberg, DB Rosen. Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system. *Neural Netw.* 1991; 4(6): 759-771.
[14] JC Bezdek, RJ Hathaway, JM Huband. Visual assessment of clustering tendency for rectangular dissimilarity matrices. *IEEE Trans. Fuzzy Syst.* 2007; 15(5): 890-903.