INTERNATIONAL RESEARCH JOURNAL OF ENGINEERING AND TECHNOLOGY (IRJET) E-ISSN: 2395 -0056
VOLUME: 03 ISSUE: 06 | JUNE-2016 WWW.IRJET.NET P-ISSN: 2395-0072

Study on Clustering Large Data Using Fuzzy Adaptive Resonance Theory

Gadupudi Dakshayani[1], Asadi Srinivasulu[2], Kancharla Haneesh[3]

[1]*Department of IT, Sree Vidyanikethan Engineering College, Tirupati, India, dakshayanigadupudi@gmail.com*
[2]*Department of CSE, JNTUA College of Engineering, Anantapur, India, Srinu.asadi@gmail.com*
[3]*Department of IT, Sree Vidyanikethan Engineering College, Tirupati, India, haneesh.kancharla@gmail.com*

Abstract—*clustering is an approach that is used to form group of similar classes. Here to cluster the mixed type of data the unsupervised feature learning was used to achieve the sparse representation (SR) which makes it easier for the clustering algorithms to separate the data. SR provides simple interpretation of the input data in terms of small number of parts by extracting the unknown composition in the data. Fuzzy Adaptive Resonance Theory (ART) adopts unsupervised feature learning to find out the features from an unlabelled data. The goal of this unsupervised feature learning is to discover the low dimensional data that captures the structure hidden in high dimensional input. The traditional clustering algorithms like k-Means can be better able to deal with numerical data rather than categorical , to deal with mixed type of data which consist of both numerical and categorical features the fuzzy ART was used. The study related to the ART network, Fuzzy ART and the k-Means is mentioned in later sections.*

Keywords: clustering, fuzzy ART, ART1, k-Means, sparse representation

1. INTRODUCTION

The issue of grouping blended sort of information can be determined by utilizing the fuzzy ART as a part of conjunction with the unsupervised element realizing where the fuzzy ART comprises of three parameters like class decision, carefulness test and the learning rate. To choose the nearest model vector the parameter classification decision was utilized and to contrast this chose model and the triumphant model the cautiousness test was utilized and the learning rate is a discretionary parameter that is utilized to quantify the pace of learning.

Bunching is the issue of collection the unlabelled information things into classes in view of the closeness of the things, here the unlabelled information things are only the information which is as pictures, sounds and recordings.

Fuzzy ART was utilized to give the higher or scanty element showing to abtain the obscure structure and by utilizing this Fuzzy ART the qualification among the numerical and non-numerical elements will turn out to be less, while by utilizing a portion of the conventional bunching calculations the refinement among them will make the disappointment of the component.

Stephen Grossberg and Gail Carpenter built up the idea of system called the versatile reverberation hypothesis which is utilized to handle the data. Here the reverberation hypothesis portrays about the neural systems by utilizing the techniques like regulated and unsupervised learning. It is additionally used to determine the issues identified with the example acknowledgment.

The unsupervised learning model is the essential ART framework and it is the neural system which comprises of two fields like examination and acknowledgment. It additionally comprises of the two modules like cautiousness test and the reset. The correlation and the acknowledgment fields are utilized to perform the cautiousness test where the examination field is utilized to choose the most comparative information vector and the acknowledgment recorded is utilized to contrast this chose information vector and the models that as of now exists. The most comparative model will be chosen taking into account the weights of the vectors for the info and the coordinated model. The versatile reverberation hypothesis neural system was utilized to shape the groups by utilizing this coordinated model, the information vectors are allocated into any of the bunches in light of the closeness of the elements in the coordinated model. It can be realized that the components which are bunched onto one gathering in the model will be valuable to frame the groups to this new info design.

By utilizing the ART arrange the information was initially prepared to frame the groups. The preparation was given to the information in view of the correlation among the info vector, carefulness parameter and the acknowledgment field. The model which passes the carefulness parameter is taken to prepare the information and it is gathered into classes according to the model which is considered to prepare the information. The weights of the fruitful acknowledgment neuron are adjusted towards the components of the information vector. Something else, if the match level is underneath the

watchfulness parameter (i.e. the information vector's match is outside the conventional expected span for that neuron) the effective acknowledgment neuron is saved and a quest for framework is finished. In this inquiry technique, acknowledgment neurons are confirmed one by one utilizing the reset limit until the cautiousness parameter is overcome by an acknowledgment match. In particular, at each cycle of the request technique the most element affirmation neuron is picked and a short time later traded off if its execution is under the cautiousness parameter (observe that it thusly releases the remaining acknowledgment neurons from its trouble).

Moderate and quick are the two key techniques that are utilized for information preparing. The distinction among the moderate and quick learning strategies can be known taking into account the time. The moderate learning strategies will take more opportunity to prepare the information perceived example and the quick learning strategy will take less time to prepare the info information to frame the bunches. The quick learning strategy is a successful technique that is utilized for a few assignments where as the moderate learning technique is just utilized just if there is a constant change in the information vector.

There are distinctive sorts of ART systems like: ART1, ART2, ART3, Fuzzy ART, ARTMAP and the Fuzzy ARTMAP

2. LITERATURE SURVEY

The survey is conducted for various approaches that are used to form the clusters and these techniques include the Fuzzy ART, unsupervised feature learning and the k-Means. The explanation for all those approaches are as mentioned below:

2.1. Fuzzy Adaptive Resonance Theory

The Fuzzy ART [4, 7] is the neural network that is used to deal with any kind of data like binary and the analog. This neural network may consist of two layers like f1 and f2 where the f1 layer will consist of input and the f2 is an output layer.
The Fuzzy ART consists of three parameters like category choice, vigilance test and the learning rate.
The steps that are involved in this Fuzzy ART are as follows:
Step1: initialization of all those parameters
Step2: initialization two parameters I and P where I is initialized with the next input vector and the P is initialized with set of prototypes.
Step3: Selecting the prototype P_i from P based on the similarity of the input vector with the set of prototypes that are present in p.
Step4: Here the selected prototype undergoes test by comparing the prototype selected previously with the input prototype if both are matched then the prototype passes the test otherwise the prototype is deactivated and the another prototype is selected to test it with the input vector, if none of the prototype passes the test then the new prototype is created.
Step5: Then the matched prototype is placed in the set of candidate prototypes.

2.2. Unsupervised Learning Feature

This feature was used to extract the hidden structure from an input data. The unsupervised learning feature [1] f can be calculated by knowing the distance from each sample x to the Centroid of the cluster. 'z' is the parameter that is used to represent the distance and the mean of this distance value can be subtracted from the distance of all the samples of that cluster to obtain the unsupervised learning feature.

2.3. K-Means

It is one of the approach that is used to form the clusters but it can only better able to deal with numerical data rather than the categorical and the success of this method will also depends on the analysis of the value k. this method mainly depends on the mean value of the cluster.
The steps that are involved in this method are as shown below:
1. Assume the k cluster centers.
2. Measure the distance from each sample x to the Centroid.
3. If the distance is less between data point to the Centroid of the cluster then the data point is assigned to that cluster.
4. To the newly obtained clusters the new cluster center can be calculated by using the formula sigma of x_j where j value lies between 1 to C_i, where it represents the number of data points that are present in ith cluster.
5. This process repeats until no data point was reassigned into any of the clusters.

Min Wei, Tommy W. S. Chow and Rosa H. M. Chan are the authors tried to cluster the heterogeneous data by using the unsupervised feature learning in conjunction with the k-Means[2]. The process of clustering this heterogeneous data is shown in the figure below:

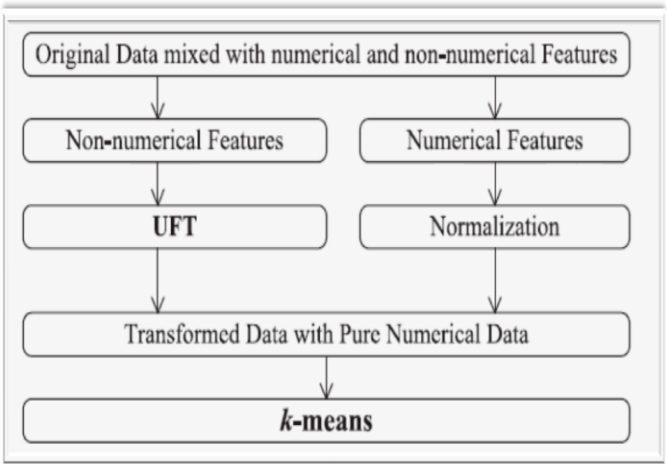

Fig 1: k-Means to deal with heterogeneous data (from [2])

Here the heterogeneous data which consist of both numerical and the non numerical features are divided into two categories such as numerical and the non numerical. The unsupervised feature transformation mechanism was applied to the non numerical data to transform the data into non numerical and to the numerical data the normalization mechanism was applied to extract the pure data and to the transformed data the k-Means algorithm was applied to obtain the refined number of clusters.

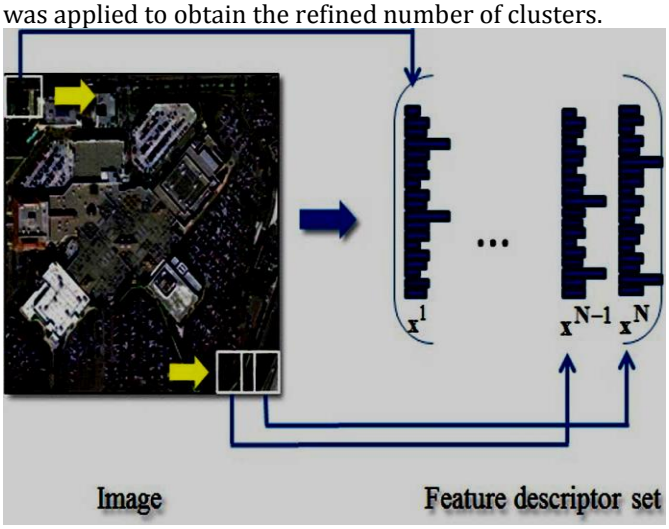

Fig 2: Extracting the features form an image (from [3])

The author Anil M. Cheriyadat has proposed a new technique to classify the scene according to the area[3] by extracting the features of an image. The method that is used by the Anil M. Cheriyadat to identify the scene from an image is the unsupervised feature learning which consist of several steps like the feature extraction, feature encoding and the feature pooling where to identify the scene the features are extracted form an image and are compared with the features that are already present in the database as shown in the figure above based on the similarity among them the area will be known. The steps mentioned above can be seen by using the figure below:

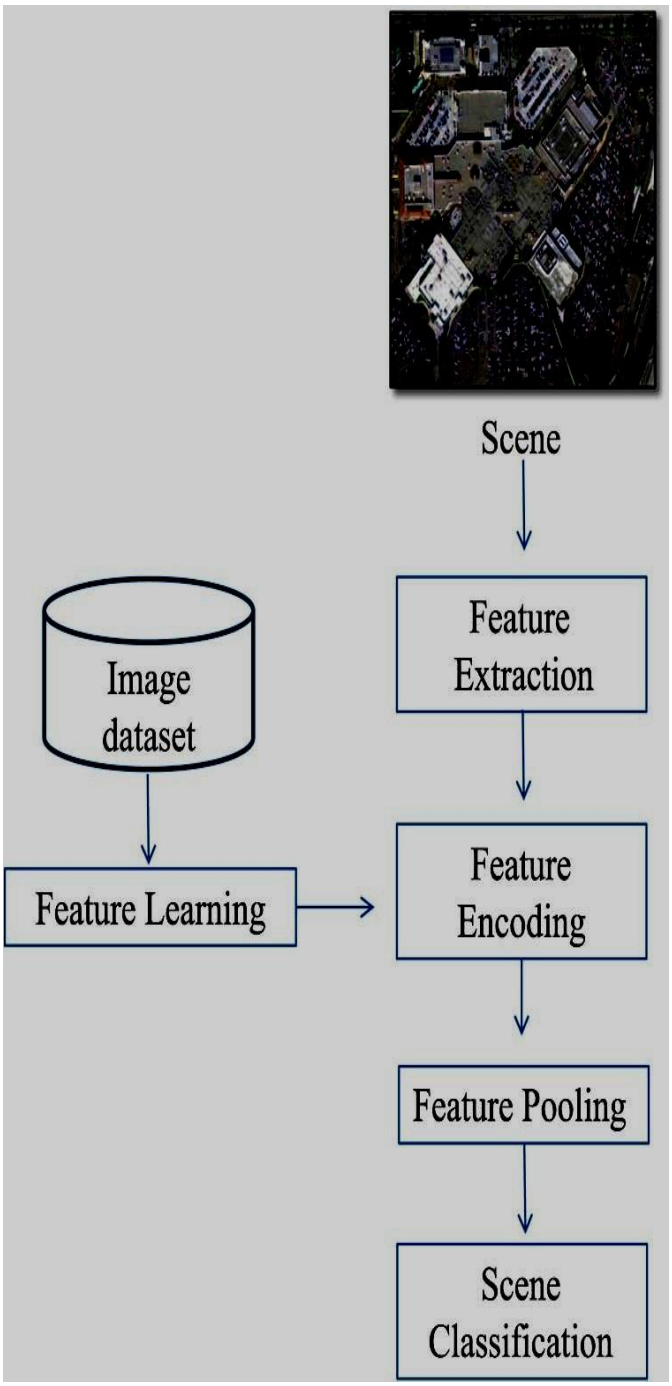

Fig 3: steps in feature mapping (from [3, 7])

Gulsen Aydın Keskin, Sevinc Ilhan, Coskun Ozkan are the authors who have proposed the Fuzzy ART for solving the supplier selection and evaluation problem[4]. Here to categorize the suppliers based on their similarity of the materials the fuzzy ART was used and here the algorithm which is mentioned in the Section A is used to form the categories of the suppliers.

J. C. Bezdek, R. J. Hathaway, and J. M. Huband have used visual assessment of cluster tendency to solve several problems like measuring the cluster tendency for row, column, and the union of this row and column objects[5]. This algorithm will work only by using the n*n reordered

dissimilarity matrix and after obtaining the result as the reordered dissimilarity image(RDI) by converting this matrix into an image the number of clusters can be known by observing the diagonal axis of the RDI. This image consists of white and black blocks the number of dark blocks that are present in the diagonal axis of the image will represent the number of clusters that are present in the data.

The reordered dissimilarity image with bark blocks is as shown in the figure 4 mentioned below. In the figure mentioned below it consist of four dark blocks in the diagonal axis of the image so that it represents that the data consist of four clusters.

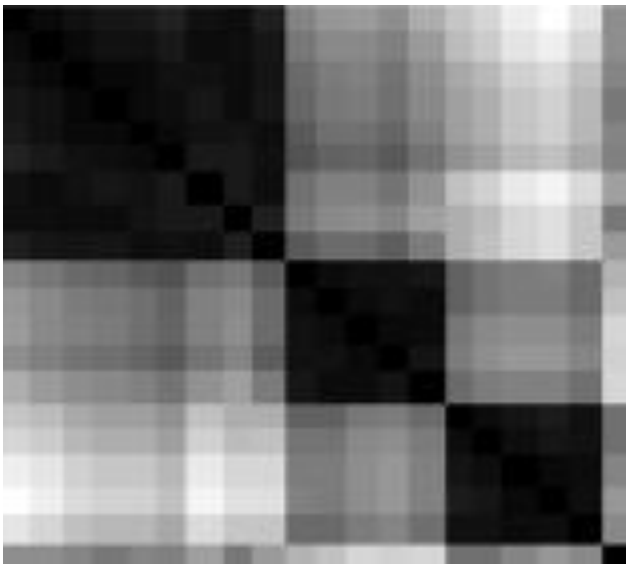

Fig 4: Reordered Dissimilarity image (from [5])

G. A. Carpenter, S. Grossberg, and D. B. Rosen used the Fuzzy ART [6] to categorize the analog pattern. ART1 consists of two parameters like union and intersection so that it can only be able to deal with binary data here to deal with analog data the two operators that are used in ART1 are replaced with the min and the max operators in Fuzzy ART [6].

3. CONCLUSION AND FUTURE WORK

By observing all the approaches that are used to form the clusters one can know that the fuzzy ART can be better used to form the clusters on any kind of data whereas by using traditional clustering algorithms can only be able to deal with numerical data. Compared with the traditional clustering algorithms this Fuzzy ART was used to obtain the better cluster tendency value and it can also be able to deal with large data. The future work should be conducted in obtaining the refined number of clusters by extended this Fuzzy ART by applying some of the techniques like VAT, co-VAT and the iVAT to the results obtained from the Fuzzy ART. It will be helpful to know the number of clusters easily and to validate the clusters by applying VAT algorithms to the results of Fuzzy ART.

REFERENCES

[1] Dao Lam, Mingzhen Wei, and Donald Wunsch, "Clustering Data of Mixed Categorical and Numerical Type With Unsupervised Feature Learning", The Journal for Rapid Open Access Publishing,Vol.3,Issue 1,2015.

[2] Min Wei , Tommy W. S. Chow and Rosa H. M. Chan," Clustering Heterogeneous Data with k-Means by Mutual Information-Based Unsupervised Feature Transformation", Entropy, vol.1,no.1,2015.

[3] Anil M. Cheriyadat," Unsupervised Feature Learning for Aerial Scene Classification", IEEE Transactions on Geoscience and Remote Sensing, vol. 52, no. 1, January 2014.

[4] Gulsen Aydın Keskin , Sevinc Ilhan , Coskun Ozkan , "The Fuzzy ART algorithm: A categorization method for supplier evaluation and selection," Elsevier., no.1, pp. 1235–1240, 2010.

[5] J. C. Bezdek, R. J. Hathaway, and J. M. Huband, ``Visual assessment of clustering tendency for rectangular dissimilarity matrices,'' IEEE Trans. Fuzzy Syst., vol. 15, no. 5, pp. 890903, Oct. 2007.

[6] G. A. Carpenter, S. Grossberg, and D. B. Rosen, ``Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system,'' Neural Netw., vol. 4, no. 6, pp. 759-771, 1991.

[7] Gadupudi Dakshayani, Srinivasulu Asadi, "survey on clustering big mixed data sets using fuzzy adaptive resonance theory," International Journal of Computational Science, Mathematics and Engineering, Vol 3, no.1, January-2016.

BIOGRAPHIES

Gadupudi Dakshayani completed her B.Tech in Audisankara college of engineering for women in the department of computer science during the year 2010-14 and currently pursuing her M.Tech in Sree vidyanikethan engineering college in the department of information technology.

Asadi Srinivasulu is working as an associate professor in Sree vidyanikethan engineering college and he is doing Ph.D in JNTUA in the department of computer science.

Kancharla Haneesh completed her B.Tech in Sree Vidyanikethan engineering college Tirupati in the department of information technology.