A COMPARATIVE STUDY OF FUZZY LOGIC WITH ARTIFICIAL NEURAL NETWORKS ALGORITHMS IN CLUSTERING

Dr. G. M. Nasira
Professor and Head in Computer Applications,
Sasurie College of Engineering, Vijayamangalam,
Erode (Dt), Tamil Nadu.
nasiragm99@yahoo.com

S. Ashok Kumar
Sr. Lecturer in Computer Applications,
Sasurie College of Engineering, Vijayamangalam,
Erode (Dt), Tamil Nadu.
ashokatparks@rediffmail.com

M.Sweety Kiruba
Lecturer in Computer Applications,
Sasurie College of Engineering, Vijayamangalam,
Erode (Dt), Tamil Nadu.

Abstract

Clustering technique challenges to find classes of patterns using some measure of similarity. Fuzzy clustering belongs to a group of soft computing techniques which includes Artificial Neural Networks (ANN) and Fuzzy Systems (FS). The Fuzzy K-means algorithm is one of the simplest unsupervised algorithms that solve a clustering problem. This procedure classifies a given data set through a certain number of cluster priorities. Neural network model attempts to emulate architecture and information representation scheme of the human brain. A class of NN, Self Organizing Feature Map (SOFM) is a clustering algorithm developed by Kohonen. Both Fuzzy Systems and Artificial Neural Networks have advantages when unclear or prior knowledge is required. A comparative study is proposed on clustering algorithm with FS and ANN. The various drawbacks with their performance are analyzed and therefore a combination of FS and ANN has been proposed. Thus this paper brings out the concept referred as Hybrid network or Fuzzy Artificial Neural Networks (FANN) which offers few new features to overcome the individual weakness of both the approaches.

Keywords: Clustering, Fuzzy Artificial Neural Networks, K-means, Self Organizing Feature Map

1. Introduction

Soft computing is a collection of methodologies which work synergistically and not competitively in one form or another reflecting its principle, exploit tolerance for imprecision, uncertainty, approximate reasoning and partial truth to achieve tractability, robustness and close resemblance with human like decision. Soft computing provides flexible information process capability for representation and evaluation of various real life ambiguous and uncertain situations. Without soft computing machine intelligence research remains incomplete.

Fuzzy set theoretic models try to imitate human reasoning and the capability of handling uncertainty whereas neural network models attempt to emulate architecture and information representation scheme of human brain [1]. Hence neuro fuzzy computing acts as more intelligent systems. Artificial neural network is used for learning and adaptation where as fuzzy systems are used to supplement its application domain.

2. Soft Computing

Artificial neural network and fuzzy logic are not scientific curiosities any more. They have been applied in various products. In this paper we describe the various concepts related with soft computing, fuzzy logic, and neural networks and hybrid neuro fuzzy systems [1]. Our purpose is simply to outline the various concepts related with soft computing and comparison on fuzzy logic algorithm with similar neural network algorithm in order to show the algorithms performances. This paper presents and deals only the first two headings i.e. the fuzzy logic and artificial neural networks in the clustering of data and hence considered below.

2.1 Fuzzy Logic

Fuzzy logic is a superset of conventional Boolean logic that has been extended to handle the concept of partial truth values between completely true and completely false. As its name suggests, it is the logic underlying modes of reasoning which are approximate rather than exact.

In the classical set theory a set can be represented by enumerating $A = \{a_1, a_2, a_3, ...a_n\}$ all its elements using the following equation If these elements a_i (i = 1, 2,…n) of A are together a subset of the universal base set x, the set A can be represented for all elements x belongs to x by its characteristic function

$$\mu_A(x) = \begin{Bmatrix} 1 \ if \ x \in A \\ 0 \ otherwise \end{Bmatrix}$$

In classical set theory x has only the values 0 (false) and 1 (true), so two values of truth. Such sets are also called crisp sets [2].

2.2 Artificial Neural Networks

Neural networks take a different approach to problem solving than that of conventional computers. Conventional computers use an algorithmic approach that is the computer follows a set of instructions in order to solve a problem. Unless the specific steps that the computer needs to follow are known the computer cannot solve the problem. That restricts the problem solving capability of conventional computers to problems that we already understand

and know how to solve. But computers would be so much more useful if they could do things that we don't exactly know how to do [2]. Neural networks process information in a similar way the human brain does. The network is composed of a large number of highly interconnected processing elements (neurons) working in parallel to solve a specific problem.

3. Clustering

Grouping of similar observations into separate clusters is one of the fundamental tasks in exploratory data analysis. Depending on the form of the data, the clustering can be done by using central or pair wise clustering techniques. Central clustering techniques minimize the average distance between an observation and its cluster center. Thus, clustering solution can be described by means of cluster centroids [3].

In pair-wise clustering, the clusters are formed by minimizing the average dissimilarity between the observations within the same cluster. Clustering is also called as unsupervised classification and the classes of the available clusters are unknown and sometimes even the number of these classes is unknown.

4. Fuzzy K-means Algorithms

In K-means algorithm, the data set is divided iteratively into k clusters by minimizing the average squared. The algorithm starts with assigning k observations as initial cluster centroids and assigning all the observations to the nearest cluster [3]. This iteration procedure is continued until the centroids stabilize. K-means algorithm is one of the simplest unsupervised learning algorithms that solve the well known clustering problem [4]. The main idea is to define k centroids, one for each cluster. These centroids should be placed in a cunning way because of different location causes different result. The next step is to take each point belonging to a given data set and associate it to the nearest centroid and to re-calculate k new centroids of the clusters resulting from the previous step. After we have these k new centroids, a new binding has to be done between the same data set points and the nearest new centroid. Finally, this algorithm aims at minimizing an objective function, in this case a squared error function. The objective function

$$J = \sum_{j=1}^{k} \sum_{i=1} \left\| x_i^{(j)} - c_j \right\|^2 ,$$

where $\left\| x_i^{(j)} - c_j \right\|^2$ is a chosen distance measure between a data point $x_i^{(j)}$ and the cluster centre c_j, is an indicator of the distance of the *n* data points from their respective cluster centres. The algorithm is composed of the following steps:

i) Place K points into the space represented by the objects that are being clustered. These points represent initial group centroids.
ii) Assign each object to the group that has the closest centroid.
iii) When all objects have been assigned, recalculate the positions of the K centroids.
iv) Repeat Steps 2 and 3 until the centroids no longer move.

Figure 4.1 is an example showing how the means m_1 and m_2 move into the centers of two clusters.

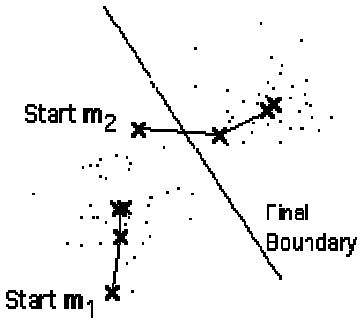

Fig.4.1 Cluster Example

5. SOFM Neural Algorithm

Self organizing map is a clustering algorithm developed by Kohonen. It creates a map of relationships among input patterns. The map is a reduced representation of the original data. A self organizing map net resembles linear vector quantification net. Both have single layer of nodes and uses a distance metric to find the output node closest to a given input pattern. The Self Organizing Feature Maps transforms the input of arbitrary dimension into a one or two dimensional discrete map subject to a topological constraint [4].

This network's key advantage is the clustering produced by the Self Organizing Feature Maps which reduces the input space into representative features using a self-organizing process. Hence the underlying structure of the input space is kept, while the dimensionality of the space is reduced. During training the self organizing map finds the output node that has the least distance from the training pattern. It then changes the nodes weights to increase the similarity to the training pattern and it influences the weights of the neighboring nodes even though they gave only random relationships to the training pattern. Nodes that are close together are going to interact differently than nodes that are far apart.

5.1 The SOFM Algorithm

Assume output nodes are connected in an array usually 1 or 2 dimensional. [4]. Assume that the network is fully connected - all nodes in input layer are connected to all nodes in output layer.

Use the competitive learning algorithm as follows:
i) Randomly choose an input vector x

ii) Determine the "winning" output node i, where w_i is the weight vector connecting the inputs to output node i. and note that above equation is equivalent to $w_i x >= w_k x$ only if the weights are normalized.
iii) Given the winning node i, the weight update is wk new= wk old + x(i ,k) where x(i, k)is called the neighborhood function that has value 1 when i=k and falls off with the distance $|r_k - r_i|$ between units i and k in the output array [5].

6. Results of Work Done

Presently the neuro-fuzzy approach is becoming one of the major areas of interest because it gets the benefits of neural networks as well as of fuzzy logic systems and it removes the individual disadvantages by combining them on the common features. These architectures have been applied in many applications. Neural networks and Fuzzy logic have some common features such as distributed representation of knowledge, model-free estimation, ability to handle data with uncertainty and imprecision etc [6]. Fuzzy logic has tolerance for imprecision of data, while neural networks have tolerance for noisy data. Hence both the areas have their own best and worst cases in their performances.

6.1 Performance Analysis

Experiments were made using three different datasets. First, there was normal distributed dataset with two clusters. These two datasets were generated using random number generator. A comparison on the performance ratio of fuzzy k-means algorithm, which belongs to fuzzy group and self organizing feature maps which belongs to neural network is thus achieved [7].

In 2-dimensional case mean vector for first cluster was m1= [0.3 , 0.3] and for second cluster m2= [0.7 , 0.7]. Deviations in all dimensions were considered to be 0.2. Clustering of each dataset was repeated 100 times so we could see how well algorithms could repeat themselves and what were different partitioning.

In 2-dimensional case SOFM with $\alpha = 0.01$, linear α decrease and net error= 1 was chosen. SOFM performed in the best way. Mean value of error was smallest and standard deviation was zero.

Mean value of error of Self Organizing Maps was not close to Fuzzy K-means algorithm and standard deviation was small hence fuzzy k-means algorithm performed worst. When different clusters are similar and overlapping, mean vectors of clusters move during iteration close to each other and depending on initial values they move to wrong position. The results are discussed in the Table 6.1.1.

Properties/ Algorithms	Fuzzy K-means	SOFM (NN)
Standard Deviation	4.19	0.02
Mean value of Error	10.32%	9.83%

Table 6.1.1 Result of normal distributed dataset

The result of normal distributed dataset of the standard deviation shown in the Table 6.1.1 is depicted as a chart for data comparison and shown in the Chart 6.1.1.

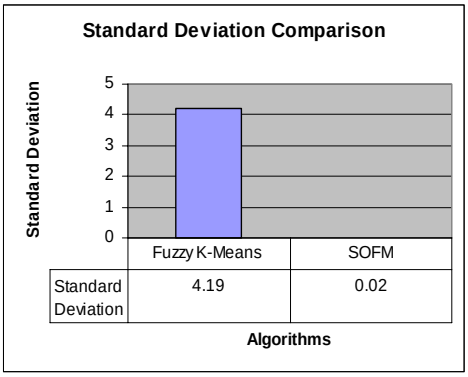

Chart 6.1.1 Standard Deviation Comparison-2D

7. Conclusion

Thus soft computing provides a predictable, greater and better solution in the field of intelligent systems. The contrast of various unsupervised algorithms like Fuzzy K-means and Self Organizing Feature Maps of neural networks depicts that they have their own advantage and disadvantage characteristics and hence a neuro hybrid system is given which takes the advantage of both the methods and hence its an up growing technology in the field of intelligent systems. Integrating fuzzy logic with neural networks has become more and more popular. However this should not be viewed as the deficiency of the methods. Rather it should be viewed as the adaptability of a powerful method to bring the best of both methods into one and to solve the problems.

8. References

[1] Jang and Sun, "Neuro Fuzzy and Soft Computing", Prentice Hall 1998, pp 35-68
[2] Stamatios V. Kartalopoulous "Under- standing neural networks and fuzzy logic", Prentice hall, 2003
[3] Andrew Moore, "K - means and Hierarchical Clustering", 1998 Pearson education, pp 115-186
[4] Abraham A., Jain L. and Kacprzyk, "Studies in Fuzziness and Soft Computing, Springer Verlag" Germany, Chapter 1, pp. 1-35, 2002.
[5] James Buckley and Eslami, "An introduction to Fuzzy logic and fuzzy sets", 2002
[6] www.citeseer.org
[7] www.wseas.com