FUZZY CLUSTERING THROUGH NEURAL NETWORK

Farhat Roohi,
Scientist, Dept of Electronics and IT,
University of Kashmir, Srinagar, J&K, India

Abstract:

Management of data has always been a key concern for scientist, scholars and industry alike. But information explosion due to information and communication technology boom has resulted in a world overloaded with data and information that too growing at an ever increasing exponential rate. Analyzing such big data requires robust techniques to first classify the data so the further analysis reduced to some manageable group of data. This way it becomes systematic and easy. But grouping of data comes at a cost of accuracy and precision of information, as group analysis works on averages and distance between data points within and between clusters is also a cause of concern. Therefore it is imperative to have such data clustering techniques which are fast, able to handle big data and classify data as per the natural logic to facilitate the process of finding knowledge hidden in the data. Fuzzy logic and artificial neural networks are such two concepts which have found increasing application in classification and clustering techniques which have made them more realistic accurate and precise. As such, apart from data clustering, neurofuzzy methodology is widely used in data mining, artificial intelligence, image recognition, knowledge management, control processes, etc. Neurofuzzy methodology performs better in terms finding sequences, associations and patterns in data besides having quick self-learning capability almost comparable to human intelligence. Against this backdrop the current paper attempts to create a conceptual understanding of neurofuzzy methodology and its application to data clustering.

Key Words: Clustering, Fuzzy, Neural Network, Neurofuzzy

Introduction:

Generation of data is not a new phenomenon. It is as old as the universe and may be its use or analysis by humans can be dated back to human civilizations. Data contains information which needs to be understood trough developing various relations and comparisons. This process of developing meaningful relations, data analysis, is at the base of knowledge management and artificial intelligence and all their related applications like, scientific processes, control systems, industrial engineering, forecasting, predicting, image/voice recognition, etc. However, the information hidden in the data, can be revealed by understanding the sequences, associations, and patterns present in the data. Scientists, scholars and knowledge workers have been making continuous efforts to understand these for last many decades. But for last two decades they have been facing the formidable challenge of data storage and analysis as the data generation is going through an explosion and exponential growth. This data explosion has been the fall out of information and communication technology revolution and other scientific and industrial developments. While data management in itself has been a complex process, this explosion has made it more complex. However, classifying and grouping of this huge data into few meaningful clusters makes data storage, analysis and information generation manageable by reducing the infinite number of data points into finite data groups or clusters. This is done in such a manner that there is homogeneity of data within a cluster and heterogeneity between the clusters. So the key for the new data analytics is that as soon as data is generated or captured it should be grouped with different relevant groups to bring an order in the data. While this makes data storage and information extraction practically possible, it also helps to minimize the computational complexities for further data analysis. Therefore, the efficiency and effectiveness of data analysis depends to a great deal on data clustering, apart from data clustering itself being a robust technique of information and knowledge generation. As such, cluster analysis forms an important component in data analysis. Accordingly, an attempt has been made in this paper to navigate through this important concept and also to discuss a clustering framework.

Clustering

At the base of generating information and knowledge lies analyzing relevant and upto date data. Such analysis requires proper storage of data in different categories so that data retrieval and mining becomes easy. Storage of data in different categories requires classification or clustering of data in different categories thus clustering has become the most important or basic step in data analysis. It is being used in other applications like pattern recognition, image, face, voice recognition, prediction, etc. Clustering in simple terms is the organization or grouping or partitioning of a collection of patterns/datum into clusters or sub classes based on some similarity. It works like this that items/data similar based on some attribute/s are clubbed together and those different are clubbed separately in different clusters so that there is within cluster homogeneity and between cluster heterogeneity. While the number of clusters into which data or items are grouped ideally depends on the degree of heterogeneity and concentrations around certain points in the data set, but sometimes number of clusters is also predecided based on certain previous knowledge. It has useful applications in many exploratory pattern-analysis, machine-learning, grouping, and decision-making situations, including image segmentation, pattern classification, data mining, document retrieval, and has been addressed in many contexts and many disciplines by researchers and scholars, that reflects its wide appeal and application (Farhat Roohi, 2013). Although there is an increasing interest in the use of clustering methods in pattern recognition (Anderberg 1973), image processing (Jain and Flynn 1996) and information retrieval (Rasmussen 1992; Salton 1991), clustering has a rich history in other disciplines (Jain and Dubes 1988) such as biology, geography, geology, archaeology, psychology, psychiatry, marketing and finance (Farhat 2013). The availability of huge and vast literature on it also signifies its importance and multidisciplinary nature.

While Farhat (2013) argues that clustering is basically conducted by analyzing the input data for automatic characterization, detection and classification, Jain and Dubes (1988) have put forth that typical pattern clustering activity involves pattern representation (optionally including feature extraction and/or selection), pattern proximity measure definition appropriate to the data domain, grouping or clustering, data abstraction (if needed), and assessment of output (if needed). However, currently clustering is classified into three broad categories viz. statistical, fuzzy and machine learning techniques. While the linear characteristics of data are analyzed in statistical techniques for classification, the machine learning technique captures the nonlinear characteristics of data, like artificial neural network (ANN) for better classification. Working on a different line, the fuzzy set theory technique introduces human like thinking in the classification process and thus makes it robust.

Fuzzy System

While by grouping of data/items usually it is understood that they are partitioned in separate non overlapping clusters known as crisps clusters, but after the emergence of fuzzy set theory and its applications to clustering a new concept of fuzzy clustering emerged in cluster analysis. By fuzzy cluster it is meant that crisp or definite boundaries are not drawn between the clusters. Rather items have a membership function to different clusters, thus creating fuzzy boundaries between clusters. This type of clustering is in conformity to the real world phenomena where the features or attributes of the objects are not sharp but have the tendency to show resemblance to few classes to some degree. Modeling this feature of natural phenomena like human intelligence does is a challenge to data and information scientists and fuzzy clustering offers an opportunity for soft partitioning of the data sets. It has the capability to reflect the real world more objectively and as such has become a very useful data analysis tool and the main area of interest in cluster analysis. It has the advantage of partial commitment to a given class in each of the iterations, therefore, it is preferred over crisp clustering, where total vector commitment is required. It, as compared to crisp approach, is more successful in situations where clusters overlap. It has also better results in avoiding local minima of the cost function and has the great ability to detect hyper volume clusters and those clusters that are thin shells - curves and surfaces.

The principle of crisp clustering is to assign a data point to one and only one cluster, which however, becomes its main deviation from the grouping/classifications found in nature. To overcome this limitation, the fuzzy clustering works by assigning a membership function to each data point which indicates its degree of belongingness to each cluster. Among fuzzy clustering methods, fuzzy c-means (FCM) algorithm, propounded by Bezdek (1974) is one of the best known fuzzy clustering approaches. FCM has found many applications and uses on may data types, including the area of image segmentation. While it works to optimize a specific cost function, it operates well with the compact or isotropic clusters. In view of its wide applications, research has been conducted to improve its accuracy while decreasing the time and number of iterations and rules to create clusters.

As far as number of clusters is concerned, clustering is performed in two ways. Either the number of clusters is decided before hand or it is left to the data and algorithm to find the clusters in the data based on certain criteria or distance measure. The latter is an exploratory method and is generally known as unsupervised method, while as the former represents the supervised case. Fuzzy clustering is capable of working with both the clustering methods.

Artificial Neural Network

Artificial neural networks (ANNs) are being used widely used in cluster analysis. The basic idea of ANN is to imitate the information representation patterns of human brain. It is achieved by copying the architecture pattern of human brain. However, a specific ANN architecture depends on the goal set for the clustering. Data or input patterns are presented at the input layer, which are connected to the output nodes through hidden layers and differential weights. It is followed by an iterative process to adjust the weights between the input nodes and the output nodes until a termination criterion is satisfied and this process of weight adjustment -learning phase- lends continuous learning or artificial learning capability to the system, which can be either supervised or unsupervised learning in ANN (Farhat,2013). She further argues that while the supervised learning phase demands knowing an output class for each of the inputs, the unsupervised learning network itself recognizes the features of the input and self organizes the inputs, and it follows two approaches, the parametric approach and the non-parametric approach. While the former approach involves combination of classification and parameterization, the lateral approach involves partitioning the unclassified data in subsets using adaptive resonance theory (ART). It is based on neurophysiology and involves a wide variety of neural networks (NN), prior knowledge and adaptive to learning. These forms the basis of competitive learning.

Competitive neural networks work like biological neural networks and thus possess competitive learning capability. They work on the principle of winner–takes–all and are often used to cluster input data (Jain and Mao 1996, Farhat 2013). The input data/patterns are grouped together on the basis of similarity, done automatically through data correlations. One such similar group is represented by a single unit or neuron. In such networks learning happens continuously as the updation of weights happens simultaneously with the reception of new data and experiences. Kohonen's learning vector quantization (LVQ) and self-organizing map (SOM) (Kohonen 1984), and adaptive resonance theory models (Carpenter and Grossberg 1990) are some of the artificial neural networks commonly used for clustering. SOM, successfully used for vector quantization and speech recognition, is an intuitively appealing two-dimensional map of the multidimensional data set (Kohonen 1984). However, its requirement is that the initial weights have to be selected properly otherwise it results in suboptimal partitions. Furthermore, its convergence is controlled by various parameters such as learning rate and a neighborhood of the winning node in which learning takes place. One more issue of concern is that a particular input pattern can fire different output units at various iterations. This raises the issue of stability in learning systems. Learning systems are said to be stable if the patterns in the training data do not change their categories even after a finite number of learning iterations. The other related issue is about plasticity, which deals with the ability of an algorithm to adapt to new data. This infers that in learning models the learning rate should tend to approach to zero as iterations progress in order to have stability and plasticity. While Carpenter and Grossberg (1990) state that ART models are stable and plastic, but they are order-dependent that is different partitions are obtained depending on the order in which data is presented to the net. Additionally, the chosen vigilance threshold value determines the size and the number of clusters created by ART net. This value also determines whether a pattern will create a new cluster or will belong to an already created cluster. However, (Hertz et al. (1991) argue that only hyper spherical clusters can be determined with the help of SOM and ART.

Neurofuzzy System- Framework

Among the characteristics of the natural systems is that they have an elements of uncertainty and self-learning, that is they learn with the addition of new data. While the traditional models or systems lack this capability, the humans possess both these capabilities of self-learning and uncertainty mapping. Whereas, the fuzzy systems have the characteristic of mapping the uncertainty in the systems by generating overlapping or fuzzy clusters, they lack the ability to learn from data. Contrary to this neural network have the ability to learn from data, however, they do not possess the feature of explaining in natural language what has been learned. In this sense neural networks and fuzzy systems are complementary to each other as far as these capabilities are concerned. In corollary to this neurofuzzy systems have been developed to mimics the human like learning capability. In such systems fuzzy systems are embed or merged with the neural networks, which offers learning as well as interpretability. This feature of neurofuzzy systems is at the centre stage to attract the attention of industry and researchers, as it imparts natural like capability artificially to the models developed. The fuzzy systems use a priori knowledge in fixing the membership functions and the consequents of the models. However, in cases where a priori knowledge is not available, neural networks are used to determine the parameters of the fuzzy system based on input-output data set and such models are labeled as neurofuzzy models.

The neurofuzzy models by combining the learning capability of neural networks and the uncertainty handling ability of fuzzy systems come closer to the natural learning and intelligence. This has made neurofuzzy (NF) models very popular. They are extensively used for computing and for solving complex problems. They however have a special application in solving clustering problems, which is one of the most important and widely used data analysis techniques. Particularly as an exploratory data analysis technique, where prior information is unavailable, it is used to find the patterns and structures in the data. In cases where the clustering knowledge is available in linguistic form or linguistic rules, the fuzzy inference system (FIS) is used to generate initial clusters. Thus neurofuzzy systems use learning from data capability of neural networks to develop the clusters and use fuzzy characteristics of fuzzy set theory to draw the inferences closer to natural inferences in linguistic form. Fuzzy characteristics are also used to develop rules in qualitative form or to take input available in qualitative form to feed into the neurofuzzy system.

While for building artificial neural networks (ANN) it is necessary to specify its architecture and the learning algorithms, the construction of FIS requires specifying the knowledge base of the fuzzy sets, and fuzzy operators. The neurofuzzy system integrates these two systems and takes advantages of both in a complementary way to have the characteristics of the learning capability and the formation of linguistic rule base in one system. Such systems have the ability to identify the structure and the parameters of the model from data and extract the rules which are interpretable in natural form. Thus neurofuzzy systems have emerged as the most popular techniques to cluster and classify the data under both unsupervised and supervised conditions. The different components and phases of neurofuzzy system have been modeled in fig. 1, adapted, along with description, from Farhat (2013). The model depicts the two phases of the neurofuzzy system with input and output space.

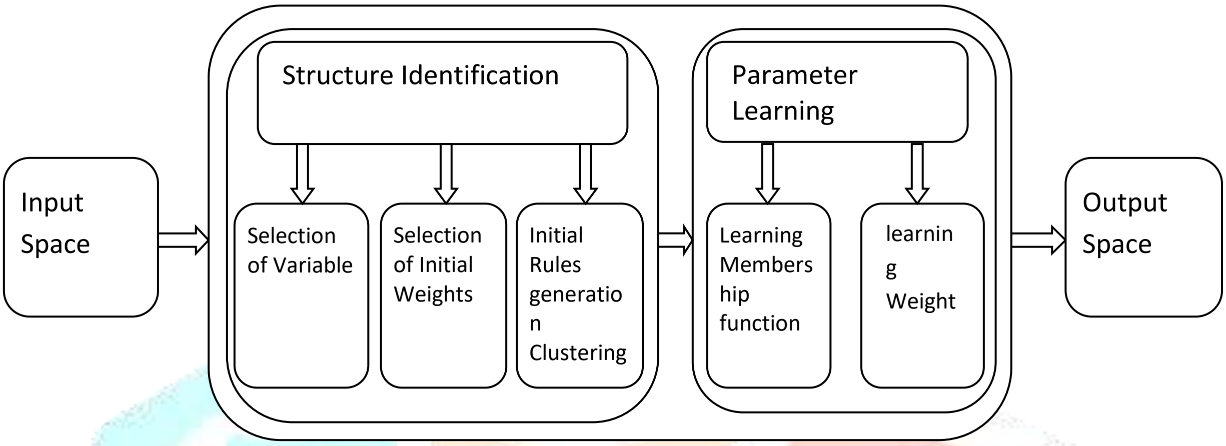

Fig. 1 Neurofuzzy Model (Adapted from Farhat 2013)

The proposed model comprises of structure identification and parameter learning phases. During the first phase sufficient number of rules and membership functions are determined to model the data under study. These rules and membership functions are determined from data to make it data driven and the essence of this phase is to properly model a system on the basis of identified input and output variables. During this structure identification phase an integrated neurofuzzy network structure is developed, that includes: sufficient input and output variables; appropriate membership functions or antecedents and consequents description; and the proper size of hidden layers of network or the numbers and the expression of the rule (Farhat2013). She further argues that while the input variable selection describes /determines the whole operating region, the output variable selection explains/determines, in certain operating regions, the behavior of the system. While the structure of the system is determined by the antecedents and consequents of the model, the rule generation is usually carried out through fuzzy clustering Fuzzy clustering works by creating fuzzy partitions, which have been classified in three diverse forms: input-space clustering (E.Kim, et al. 1997), output-space clustering (Emami, et al, 2000) and input-output-space clustering (Hellendoom and Driankov 1997). The data in the input-space clustering based on the data correlation are classified into various sub-groups through clustering which is followed by fitting of output data every sub-groups by functions like Gaussian, linear, Sigmoid functions, etc. The output-space clustering follows a different procedure. In it, first the output space is clustered, followed by the fuzzy partitioning of input space which is done by projecting the output space partition separately onto each input space. In contrast to these, the input-output-space clustering is carried out by combining the input and the output data which is followed by fuzzy clustering of the whole space. In this method, also known as product-space clustering, data are combined on the basis of data causality. Generally unsupervised type of learning is followed in this type of clustering and either direct method based on the estimation of the probability function or indirect method based on the similarity metric of the data is used (Farhat2013). The quality and representativeness of the output of fuzzy clustering depends on the clustering criteria, distance metrics, similarity metrics, and the training data. Therefore, selection of clustering criterion is of prime importance in clustering. Researchers have used different criteria for clustering like, the correlation among components of sample data (Kim et al, 1998), the similarity matrix which is composed of feature weights in the objective function (Yeung and Wang, 2000), and target function using entropy(Qiu et al, 2002), etc. After the structure is identified in the first phase, the learning or fine tuning takes place in the second phase.

The parameter learning phase, in complement to the first phase, deals with the tuning of rule coefficients for better model fit. Comparatively, better estimated solutions and lesser ambiguity is associated with this phase. Usually learning happens in two steps. Whereas the first step deals with the determination of the network structure or the number of rule nodes and the initial rule parameters or weights, the second step involves adjustment of parameters by using parameter identification. It is usually done with the help of the family of gradient algorithm or the family of least squares estimation. Of the many gradient identification algorithms that have been proposed, E.Kim, et al. (1997), in place of nonlinear optimization methods have used the gradient descent algorithm for the precise adjustment of parameters of the fuzzy model, whereas Wong and Chen, (1999) used a multi-input fuzzy model constructed automatically by using gradient descent approach and clustering method. Similarly least square method is also used for clustering. It works on the principle of minimizing the squared distances or error terms between data points or clusters. This method due to its wide successful practical applicability has been used and recommended by many researchers as a basic identification technique. As this system generates the initial structure and parameters from the data, thus it requires little tuning which makes it fast. But, over fitting the data needs to be addressed to avoid the risk of output error or to maximize the performance index. This can be achieved through effective tuning of the membership functions.

Conclusion:

Data clustering has emerged as the most valuable and commonly used data analysis technique and as such the present paper has made an attempt to introduce a neurofuzzy system model for data clustering. It involves the structure identification and parameter learning phases. It works by automatically generating fuzzy rules, followed by neurofuzzy network size fixation. Neurofuzzy networks blend the fuzzy set theory and the neural networks to provide a unique solution to the clustering problems. They by having the feature of self-learning, to generate rules and patterns automatically, solve a typical problem of clustering to develop different algorithms for clustering the same data for different applications. Accordingly, it has been concluded that neurofuzzy systems should be increasedly used in clustering problems.

References:

Anderberg, M. R. Cluster Analysis for Applications. Academic Press, Inc., New York, NY. 1973

C. Wong, C. Chen. A hybrid clustering and gradient descent approach for fuzzy modeling [J]. IEEE Trans.on Systems,Man and Cybernetics, Part B, 29(6) :686 - 693, 1999.

Carpenter, G. and Grossberg, S.. ART3: Hierarchical search using chemical transmitters in self-organizing pattern recognition architectures. Neural Networks 3, 129–152, 1990.

D.S. Yeung, X. Z. Wang. Using a neuro-fuzzy technique to improve the clustering based on similarity [C]// 1EEE Int. Conf. on Systems, Man, and Cybernetics, 5 : 3693 – 3698, 2000.

E. Kim, M.Park, S.Kirn, et al.'A transformed inpnt-domain approach to fuzzy modeling [Jl. IEEE Trans. on Fuzzy Systems,6(4): 596 - 604, 1998.

E.Kim, M.Park, J.Seunghwan, et al. A new approach to fuzzy modeling [J]. 1EEE Trans. on Fuzzy Systems,5(3):328- 337, 1997.

Farhat Roohi, Neuro Fuzzy Approach to Data Clustering: A Framework For Analysis, European Scientific Journal, vol.9, No.9, 183-192, March 2013

H. Hellendoom, D. Driankov: Fuzz)" Model Identification Selected Approaches [M]. Berlin: Springer, 1997.

Hertz, J., Krogh, A., and Palmer, R. G. Introduction to the Theory of Neural Computation. Santa Fe Institute Studies in the Sciences of Complexity lecture notes. Addison- Wesley Longman Publ. Co., Inc., Reading, MA. 1991.

J. Bezdek, Fuzzy mathematics in pattern classification, Ph.D. thesis, Ithaca, NY: Cornell University, 1974.

Jain, A. K. and Dubes, R. C. Algorithms for Clustering Data. Prentice-Hall advanced reference series. Prentice-Hall, Inc., Upper Saddle River, NJ. 1988.

Jain, A. K. AND Flynn, P. J. Image segmentation using clustering. In Advances in Image Understanding: A Festschrift for Azriel Rosenfeld, N. Ahuja and K. Bowyer, Eds, IEEE Press, Piscataway, NJ, 65–83. 1996.

Jain, A. K. and Mao, J. Artificial neural networks: A tutorial. IEEE Computer 29 (Mar.), 31–44. 1996.

Kohonen, T. Self-Organization and Associative Memory. 3rd ed. Springer information sciences series. Springer-Verlag, New York, NY. 1984.

M. Emami, A. A. Goldenberg, I. Turksen. Systematic design and analysis of fuzzy-logic control and application to. robotics, Part I [J]. Modeling, Robotics and Autonomous S) stems, 33 (2,3) : 65 – 88, 2000.

Rasmussen, E. Clustering algorithms. In Information Retrieval: Data Structures and Algorithms, W. B. Frakes and R. Baeza-Yates, Eds. Prentice-Hall, Inc., Upper Saddle River, NJ, 419–442. 1992.

Salton, G. Developments in automatic text retrieval. Science 253, 974–980. 1991.

X.Qiu,Y.Tang, D.Meng, et al.A new fuzzy clustering method based on distance and density [C]// IEEE Int. Conf. on Systems, Man and Cybernetics,7 : 5 – 9, 2002.