International Journal of Engineering Research-Online
A Peer Reviewed International Journal
Articles are freely available online:http://www.ijoer.in
Vol.1.Issue.1:2013

RESEARCH ARTICLE

CLUSTERING OF IMAGE DATASET USING ROUGHSET AND SOFM ALGORITHM

E. MALARVIZHI
Tamilnadu college of Engineering, Anna University, Tamilnadu, India

Dr.M.KANNAN
Professor, Tamilnadu College of Engineering, Anna University, Tamilnadu, India

Article Received:22/04/2013 | Revised on: 23/04/2013 | Accepted on: 18/05/2013

ABSTRACT
The main objective is to categories the different types of images in a huge database by means of clustering, which is the process of Data mining by SOFM (Self Organizing Feature Maps) and Rough set algorithm. Document or image clustering is the process of partitioning the unlabelled documents/Image for the purpose of organizing the data. New Rough Set neural network based on SOFM is proposed. The model perfectly solves many problems, such as the effects of training sample size and sample quality on accuracy of artificial neural network. Besides, the new network has reduced computation and time training needed, simplified the neural network structure and improved the system speed. Experimental results indicate that the system not only increases the quality and rate of diagnosis, but also reduces the measure items and diagnosis costs, which makes the result visualized and it has favorable applied prospect.

Key words— clustering, roughset, SOFM, unsupervised learning, artificial neural network

INTRODUCTION

Artificial Neural Network has the function to handle complex pattern and to associate, speculate, remember, and has fault tolerance and extensibility, especially with highly self-study, self-organization and self-adaptive capacity which has become an effective method of failure diagnosis, and has been successfully applied in many actual diagnose systems. However, experience has shown that the Neural Network has much limits of association, once beyond the limits, it will speculate in the wrong way which make the decision system produce a phenomenon of misjudge or missing. Self Organizing Feature Map (SOFM) is an artificial neural network based algorithm that organizes the data by itself without any pre description. Professor Z. Pawlak from Warsaw University, Poland, proposed Rough Set theory. The theory as a new mathematical tool is used to study the representation, learning and induction, other method of incomplete data and imprecise knowledge which overcomes the fuzzy process of learning.

An unsupervised learning method in which networks will form their data without external help by organizing the data by itself. With strong qualitative analysis is one of its outstanding merits, which does not require pre-description of a given number of a certain characteristic or attributes, it is through the relations between unresolved classes and unresolved relations to determine the approximate domain, as well as identify the inherent law of the problem. In order to make better use of the decision-making ability of Artificial Neural Networks and Rough Set theory, This article combine self-organizing feature map (SOFM) with the Rough Set theory to set a rough set neural network model in which size, sample quality such problems have been well solved, makes the training sample work on the less computation and this model simplified the neural network structure, increased the operational speed of system.

***CLUSTERING*:** Clustering is an unsupervised learning task where one seeks to identify a finite set of categories termed clusters to describe the data .Clustering is an unsupervised learning task where one seeks to identify a finite set of categories termed clusters to describe the data. SOFM is one of the simplest unsupervised learning algorithms that solve the well-known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters. The procedure follows a simple and easy way to classify a given data set through a certain number of Clusters.

The main idea is to initialize and cluster the network all the connection weights are initialized with small random values. For each input pattern, the neurons compute their respective values of a discriminate functionwhich provide the basis for competition. The particular neuron with the smallest value of the discriminate function is declared the winner. The winning neuron determines the spatial location of a topological neighborhood of excited neurons, thereby providing the basis for cooperation among neighboring neurons. The excited neurons decrease their individual values of the discriminate function in relation to the input pattern through suitable adjustment of the associated connection weights, such that the response of the winning neuron to the subsequent application of a similar input pattern is enhanced. The algorithm is significantly not sensitive to the initial randomly selected cluster centers. Clustering of the images are based on the label of the image and also the properties of the image such as intensity levels, resolution of the image. Resolution defined in the matrix and the dimension of the matrix rows and columns are considered which colour image, binary is, grayscale images are clustered based on the weights assigned by the rough set algorithm.

***ROUGH SET THEORY*:** Rough set theory is a formal mathematical tool can deal with imprecise, fuzzy issues, inconsistent and incomplete information. It can be applied to simplifying the dimensionality of datasets. Therefore, rough set has been applied in such fields as machine learning, data mining, artificial networks, etc., successfully since Professor Z. Pawlak developed it in 1982. According to the idea of rough set, the domain will be divided into positive region, boundary region and negative region. The decision rules can classify the positive and negative regions with high accuracy. It provides efficient algorithms for finding hidden patterns in data, evaluates significance of data, generates sets of decision rules from data, it is easy to understand.

Basic concepts and properties

1) Information Systems

Information Systems can be represented by a four-tuple:

$$IS = (U , A , V_a , f_a)$$

Where: is the universe, to represent a limited set of objects

$$U = \{ x_1 , x_2 , x_3 , \cdots , x_m , \}$$

A represents a limited set of attributes, including the condition attribute and decision attribute,

$$A = \{ a_1 , a_2 , a_3 , \cdots , a_n , \}$$

A is range for the attribute, for instance, the attribute represents that the object attribute is a true or a false, then value can be {true, false}; For each attribute, definition of information function is f:U V two-dimensional information table can be used to describe the information system, it means objects of U are rows of the table, and attributes of A are columns. Each value of object property is represented by each element of the table that is the value of A.

2) Reduction

Assume A as a cluster of equivalence relation, a A, If

$$\mathrm{IND}(A) = \mathrm{IND}(A - \{a\}) ,$$

then claimed a in the knowledge A can be reduced.

3) Independence of the relationship if all aA in A cannot be reduced, then we call that A is the relation of independent equivalence, otherwise it is related.

4) Core We call all the relations of A which cannot be reduced core, the set of its core set A, denoted by CORE A. If A is independent and $B \subseteq A$ so B is also Independent;

$$\mathrm{CORE}(A) = \cap \mathrm{RED}(A) , \ \mathrm{RED}(A)$$

is all reduction cluster for A.

5) Relative core of properties and reduction P and Q are two property subsets of U, when

$$\mathrm{POS}_{\mathrm{IND}(P)}(\mathrm{IND}(Q)) = \mathrm{POS}_{\mathrm{IND}(P - \{A\})}(\mathrm{IND}(Q)) ,$$

then P is reducible on Q, otherwise, a irreducible P on Q. When all a in P is irreducible on Q, called P is independent about Q. The relative reduction is fined as: if S is an independent subset of P

Q, and $\mathrm{POS}_S(Q) = \mathrm{POS}_P(Q)$, then set S belongs to P is reduction P on Q. We call all the irreducible attributes Q in subset of P are Q cores of P. Relative core and relative

6) Reduction h the following relationship:

$\mathrm{CORE}_Q(P) = \cap \mathrm{RED}_Q(P)$

B. Decision table (information table)'s reduction steps

According to the roughest theory this paper has achieved a reduction of N attributes. We adopted the recursive call the deletion of a property K of the decision table is denoted by $SD_k^{IND(A-\{a_K\})}$ of SD_k and IND(A) of SD_0 (The Initial Decision Table) is not the same ,then it indicates that the property can't be removed, otherwise, the property can be removed, calculate the remaining N-1 attributes. Figure 1 shows that main process

***SOFM ALGORITHM*:** The Self Organizing Feature Map (SOFM) is an excellent tool in exploratory phase of data mining. It projects input space on prototypes of a low-dimensional regular grid that can be effectively utilized to visualize and explore properties of the data. Learning is based on the concept that the behaviour of a node should impact only those nodes and arcs near it. Each input is connected to all output neurons. There is a weight vector attached to every neuron with the same dimensionality as the input vectors. SOFM compared with the traditional clustering method, is an unsupervised clustering method, it can form the cluster centre, which map a curve surface or plane, while maintaining the same topological structure. Figure 2 Kohonen network structure. It consists of input layer and competitive layer. Neurons form constitutes a two-dimension between the two the competitive layer, and nal planar arrays. The whole connected o layers are implemented, and sometimes layer between the neurons connected through lateral inhibition. There are two connection weights in the network, one is the connection weights between neurons, and the size of the interaction between neurons is controlled by its size, the other connection weights of neurons response to the external input. The function of neurons in the network between competition and interaction simulates the brain information processing of self-organization, self-learning and the clustering function.

A. Training of SOFM: Initially, the weights and learning rate are set. The input vectors to be clustered are presented to the network. Once the input vectors are given, based on the initial weights, the winner unit is calculated either by Euclidean distance method or sum of products method. Based on the winner unit selection, the weights are updated for that particular winner unit. An epoch is said to be completed once all the input vectors are presented to the network. By updating the learning rate, several epochs of training may be performed.

The following are the basic steps SOFM algorithm:

Step 1: We can take a small random value between the input neurons and output neurons weights. For each input vector, we executive from the second to the sixth step, until N all he input vector is trained completely. In other words, it is initialized.

Step 2: Propose a new input mode, input vector.

Step 3: Compute the Euclidean distance from the input node to each output node:

$$d_j = \sum (x_i(t) - W_{ij}(t)) \qquad j = 1, 2, \cdots, N$$

Step 4: Select the

$$d_j \ (\ j = 1, 2, \cdots, N\ )$$

minimum node j^*, $d_{j*} = \min d_j$

Step 5: Adjust j^* and the value of all the nodes in $NE_j^*(t)$ field

$$W_{ij}(t+1) = W_{ij}(t) + \eta(t)(X_i(t) - W_{ij}(t))$$
$$j \in NE_j^*(t) \qquad i = 1, 2, \cdots, K$$

And $NE_j^*(t)$ is a j*'s Euclidean field and single reduction function of time and $0 \le \eta(t) \le 1$.

Step 6: Go step 2.

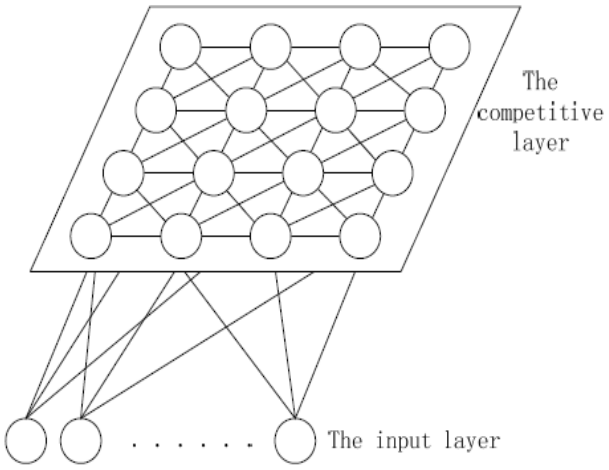

Figure 1 Kohonen architecture.

***COMBINED PROCESS OF ROUGHSET AND SOFM AN NEURAL NETWORK* :** Set of learning samples: The sample used for learning, In collected learning samples, some features is defective, some feature is repeated, so we have to classify the data, remove some redundant data, and add some

necessary feature intervals, and each interval characteristic value shows minimum conditional attribute set and the corresponding learning by reduction. The sample set retains decision model .First a learning sample set is formed according to the known field knowledge, and adopt suitable discrete method for discrete continuous characteristics, to form the organizational decision table, and use the Rough Set theory to reduce decision table, and then obtain the learning samples by eliminating the concentrated and repeated learning characteristics of the samples, to enhance the network's learning ability. For classification of decision characteristics, important and indispensable feature is the core of characteristics of the sample sets after analysing by Rough Set theory, removing the core elements or changing the value of the core elements will cause quality changes of rough set classification, so, a major characteristic of rough set neural network based on SOFM is the direct interconnection of core elements and their corresponding output layer neurons, that changes in nuclear elements reflect the neural network output directly.

CONCLUSION

The system combines the advantages of Neural Network theory and Rough Set theory. Finding core for condition attributes by rough set theory, influence of noise data in sample is well eliminated, and accuracy in system has been improved. Using Rough Set theory, the sample and the conditional properties have been simplified, a putting in the neural network, the neural network's structure has been simplified, and the speed of the system has been improved. Using this system for fault diagnosis, as the conditional properties are greatly reduced, the cost of the system is reduced, accelerates the speed of diagnosis, and enhanced the real-time performance. SOFM network can overcome some short coming of the common BP network, SOFM network has high learning efficiency, simple algorithm, lateral association functions, strong pattern clustering function, easy to diagnose, visual, and many other advantages. For enhancing these data bases the proposed algorithm defines thousands of images and it classifies those images by considering only limited properties of the image. For future clustering images can be done on the basis of all the properties of the images SOFM algorithm classifies RGB ,binary, jpeg, mpeg images for enhancing more types of images can be collected for the categorizing of those images.

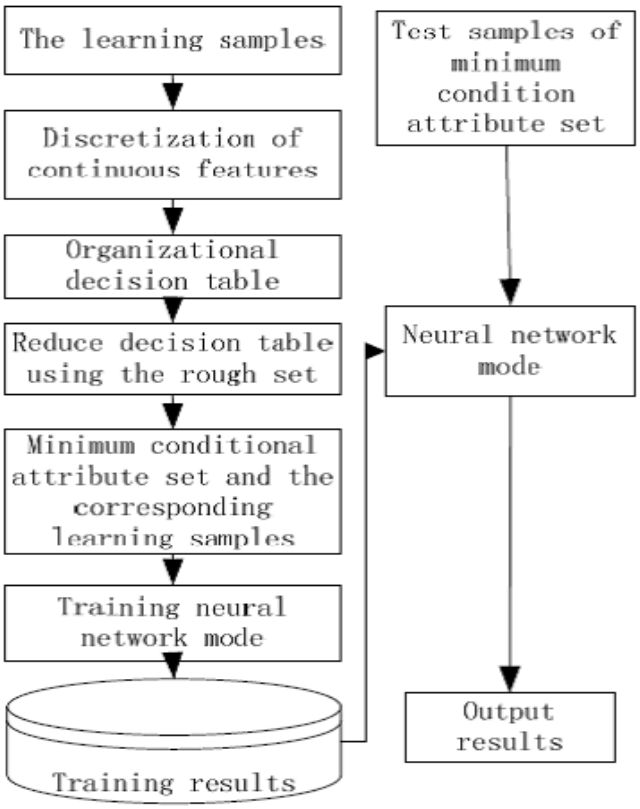

Figure 2 Architecture of roughest neural network systems using SOFM

REFERENCES

[1]. Qing Liu Rough Set and Rough Deduction [M]. Beijing: Science Press, 2001(In Chinese)

[2]. Zhitao,BaodongXu,Dingwei Wang etc. Areductionalgorithmbased on genetic algorithm of roughest knowledge [J]. System Engineering and application, 2003 21(4):116-122

[3]. Jun Duan, RuipingGeng, XuYan Tu. A CBR quick retrieval method based on rough Sets and Neural Network [J].Computer Engineering and Application,2003 39(3):25-27 (In chinese)

[4]. KohonenT.AutomaticFormatic of Topological Maps in Sef-Organizing systems inproc. For the 2nd Scandinavian Vonf. On Image Analysis, 1981, 214-220.

[5]. Vesanto J. Alboniemi E. Clustering of the self-organizing map [J]. IEEE Transaction on Neural Networks, 2000, 11(3)586-600. [6] John A F. Self-organization in SOM with a finite input [C].

In Proceedings European Symposium on Artificial Neural Newoks. Bruges(Belgium). 2000, 25-28.

[6]. Buxi Ni, LifuNetwork[J]. Computer Engineering and Design, 27(5): 855-856. (In Chinese).

[7]. Jiawei Han and MichhelineKamber, “Data mining concepts and techniques”-a reference book

[8]. Arun K. Pujari, “Data mining techniques”-a reference book. [3] .DariuszMayszko, Sawomir T Wierzchon”Standard and Genetic k-means Clustering Techniques in Image Segmentation.” (CISIM'07)-2007.

[9]. R.Xu, D.WunschA.Jain, M. Murty, P. Flynn, "Data clustering: A review", ACM Computing Surveys, 31, 1999,pp- 264

[10]. A.Jain, M. Murty, P. Flynn, "Data clustering: A review", ACM Computing Surveys, 1999,pp- 264-323.