Feature Space Reductions Using Rough Sets for a Rough-Neuro Hybrid Approach Based Pattern Classifier

Ashwin Kothari, *Member IAENG,* Allhad Gokhale, Avinash Keskar, Shreesha Srinath, Rakesh Chalasani

***Abstract*—Use of rough set theory at preprocessing stage for the targeted dimensionality reduction is very important in case of Unsupervised ANN based pattern classifiers. Such attribute compression by removing the redundancies results in reduced set of attributes acting as the input for the unsupervised neural network, based on the Kohonen's learning rule. This would result in speeding up the learning process, better understanding about the data by recognizing the significant contributors to discernibility. Here the case of classification of printed alphabetical characters is taken, for which letters A-Z of eighteen different fonts form the data set. The features extracted are broadly classified into statistical and structural categories. The statistical features are inconsistent compared to the structural features. Exploiting inconsistencies of the data by rough sets has been observed and which results in exclusion of some important structural features. Inclusion of such features results in improved classification by the ANN.**

***Index Terms*—Discernibility, Feature extraction, Pattern classification, Reducts, Rough Neuron, Rough Sets, Unsupervised neural network,**

I. Introduction

Rough sets theory exploits the inconsistency and hidden patterns present in the data. Rough sets have been proposed for a variety of applications. In particular, the rough set theory approach seems to be important for artificial intelligence and cognitive sciences, especially for machine learning, knowledge discovery, data mining, expert systems, approximate reasoning and pattern recognition. Artificial Neural Networks in the most general form aim to develop systems that functions similar to the human brain. The nature of connections and the data exchange in the network depend on the type of application. Rough sets and Neural Networks can be combined, as they would be effective in cases of real world data, which are ambiguous, and error prone.

This paper proposes a rough-neuro based hybrid approach for classification by unsupervised ANN. The rough set theory is used for reducing the input feature space and the neural network does the classification. A case study of printed character recognition has been undertaken to establish that a Rough-Neuro Hybrid approach has reduced dimensionality and hence also has lesser computations as compared to the pure neural approach. The data set used consist of characters A-Z, in eighteen different fonts. The results for pattern classification using a Pure Neural approach have been used for benchmarking. First Section discusses the steps of image preprocessing and Feature extraction. As the values obtained are continuous, the discretization steps used are explained in the following section. The feature space reduction and rough hybrid approach, are discussed in the subsequent sections. The last section presents the results and conclusions.

II. Image Processing and Feature Extraction

The original data set is subjected to a number of preliminary processing steps to make it usable by the feature extraction algorithm. Pre-processing aims at producing data that is easy for the pattern recognition system to operate accurately. The main objectives of pre-processing [1] are: binarization, noise reduction, skeletonization, boundary extraction, stroke width compensation [2], truncation of redundant portion of image and resizing to a specific size. Image binarization consists of conversion of a gray scale image into a binary image. Noise reduction is performed using morphological operations like dilation, erosion etc. Skeletonization of the image gives an approximate single-pixel skeleton, which helps in the further stages of feature extraction and classification. The outermost boundary of the image is extracted to further obtain the boundary related attributes such as chain codes and number of loops. Stroke width compensation is performed to repair the character strokes, to fill small holes and to reduce uneven nature of the characters. The white portion surrounding the image can create noise in the feature extraction process and also increases the size of image unnecessarily. Truncation is performed to remove this white portion. In the end, the image is resized to a pre-defined size: 64pixel x 64pixel in this case. All such results are indicated in fig. 1 below. In feature extraction stage, each character is represented as a feature vector, which becomes its identity. The major goal of feature

Submitted for International Conference on Soft Computing and Applications (ICSCA'08)
Ashwin kothari is Sr. Lecturer in department of Electronics and Computer Science, V.N.I.T, Nagpur, M.S. India-440010, Member IAENG. (Phone: +91 9890164705, +91 712 2801049; fax: 91 712 2223230; e-mail: agkothari72@rediffmail.com, ashwinkothari@ece.vnit.ac.in).
Dr. Allhad Gokhale is Prof. in department of Electronics and Computer Science, V.N.I.T, Nagpur, M.S. India-440010. (e-mail: agkeskar@ece.vnit.ac.in)
Dr. Avinash Keskar is Prof. and Dean R&D in department of Electronics and Computer Science, V.N.I.T, Nagpur, M.S. India-440010. (e-mail: apgokhale@ece.vnit.ac.in).
Shreesha Srinath and Rakesh Chalsani are eighth semester B.Tech students of department of Electronics and Computer Science, V.N.I.T, Nagpur, M.S. India-440010. (e-mail: shreeshasrinath@gmail.com, vnit.rakesh@gmail.com).

extraction is to extract a set of features, which maximizes the recognition rate with the least amount of elements. The feature extraction process used consists of two types of features: statistical and structural [2, 3, 4]. The major statistical features used are: zoning, crossings, pixel density, Euler number, compactness, mean and variance. In zoning, the 64 x 64 character image is divided into 16 x 16 pixel parts and pixel density of each part is calculated individually. This helps in obtaining local characteristics rather than global characteristics and is an important attribute for pattern recognition. Crossings count the number of transitions from background to foreground pixels along vertical and horizontal lines through the character image. In fig.2 there are 6 vertical crossings (white to black and black to white) and 4 horizontal crossings in both the upper and lower part. Pixel Density is calculated over the whole 64 x 64 image. Euler number of an image is a scalar whose value is the total number of objects in the image minus the total number of holes in those objects. Euler number is also calculated for each image. Structural features are based on topological and geometrical properties of the character, such as aspect ratio, loops, strokes and their directions etc. The boundary of the image is obtained and chain code is calculated for it. Then the number of ones, twos till number of eights is calculated. The number of loops present in a character is also obtained.

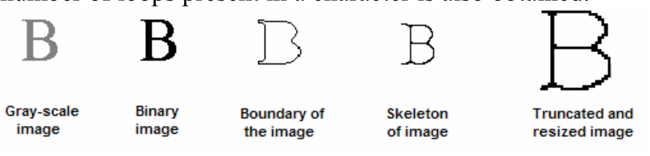

Figure 1 Output at different stages of preprocessing.

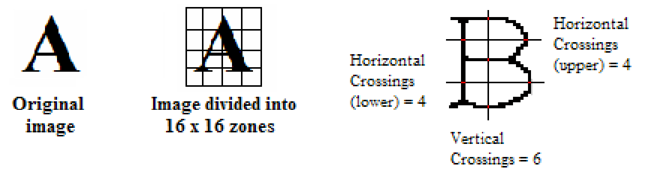

Figure 2 Image segmentation and crossings for feature extraction.

III. DISCRETIZATION

The solution to deal with continuous valued attributes is to partition numeric variables into a number of intervals and treat each such interval as a category. This process of partitioning continuous variables is usually termed as discretization. Data can be reduced and simplified through discretization. In case of continuous valued attributes, large number of objects are generated with a very few objects mapping into each of these classes. The worst case would be when each value creates equivalence with only one object mapping into it. The discretization in rough set theory has particular characteristic, which should not weaken the indiscernibility ability. A variety of discretization methods have been developed.

A. Algorithm

Consider an information system $IS = (U, A, V, f)$ where U: is the universal sets containing all objects i.e. $U = \{x_1, x_2, x_3,\ldots, x_n\}$, n is the total number of objects; $A = CU\{d\}$ where C denotes the set of condition attribute and d is the decision attribute; V denotes contains the sets of values each condition attribute can take; f is a function between the element in U and its value, the value of object x_i to the attribute a is $a(x_i)$. The process of discretization is to find the sets of cut points for each attribute and hence discretize each of them. For an attribute a, V_a the set containing the values the attribute can take. Cuts are nothing but partitions of the set V_a i.e. $V_a = [c\ d]$, where $[c\ d]$ is the interval of the continuous valued attribute. Then a partition: $c<p_1<p_2\ldots\ldots<p_m<d$,where the set $\{p_1,p_2,\ldots..p_m\}$ forms the set of cut points which divides the interval $[c\ d]$ into m intervals without intersection: $[c\ p_1)$, $[p_1\ p_2)$, $[p_2\ p_3)\ldots..(p_m\ d]$ and the continuous values of attribute a turn out to be $m+1$ discrete values: $V_1, V_2, V_3, \ldots\ldots, V_{m+1}$ by the following equations[5]:

$$\{V_1,\ \text{if } a(x) \le p_1\} \tag{1a}$$

$$V(x) = \{V_i,\ \text{if } p_{i-1}<a(x) \le p_i,\ i=1, 2, 3\ldots, m\} \tag{1b}$$

$$\{V_{m+1},\ \text{if } a(x)>p_m\} \tag{1c}$$

B. Steps to obtain cuts.

To search the cuts points we make use of the discernibility matrix. The discernibility matrix of a given decision table is defined as:

$$(M_{\{d\}}(i,j))_{nxn}$$

Where

$$M_{\{d\}}(i,j) = \{\{a_k \mid a_k(x_i) \neq a_k(x_j),\ \text{for all } a_k \text{ in } C\}\ \text{when } d(x_i) \neq d(x_j)\} \quad \{0,\ d(x_i)=d(x_j)\} \tag{2}$$

C. Steps for Algorithm

Step 1. Construct the discernibility matrix for the given table.
Step 2. For $i = 1$ to t,
Step 3. For all $1\le j\le k\le n$, if the attribute a_i is a member of the discernibility matrix entry, construct the cuts interval $[a_i(x_j)\ a_i(x_k]$.
Step 4. For every set of intersecting intervals, construct the cut point as:
$(p_j)_i = \{max(lower\ bounds) + min(upper\ bounds)\}/2$.
Step 5. Discretize attribute a_i according to the cuts obtained.
Step 6. Next i,
Step 7. End.

The tables 1&2 show an undiscretized table and the discretized table using the above-discussed algorithm. It is observed that the continuous interval valued attributes are discretized without affecting their ability to discern. It can be further seen that this helps in attribute reduction also.

Table 1Undiscretized data table

Attribute→	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	0.056274	4	2	2	2	1	0	69	69	0	0	102	102	0	70	116	116
2	0.1073	4	3	2	2	2	106	136	88	110	80	136	84	128	80	96	24
3	0.066528	-1	2	2	2	0	70	103	64	109	213	0	0	17	217	0	0
4	0.104919	-1	3	3	3	1	201	122	77	0	176	0	59	150	176	0	0
5	0.120117	1	4	2	2	0	247	112	112	112	246	96	96	78	241	16	16
6	0.111206	5	3	3	1	0	110	242	44	202	80	242	84	62	80	240	36
7	0.10791	4	4	2	2	0	19	172	141	87	157	41	55	62	176	88	83
8	0.136475	8	1	2	2	0	160	132	132	164	144	121	121	144	144	121	121
9	0.25	8	1	2	2	0	256	256	256	256	256	256	256	256	256	256	256
10	0.057068	2	1	1	1	0	0	0	0	176	0	0	0	176	0	0	0
11	0.1297	6	1	2	2	0	160	144	58	29	144	157	143	0	144	177	250
12	0.097595	5	2	1	2	0	88	232	16	0	64	224	0	0	64	224	0
13	0.137634	5	1	4	4	0	117	193	32	256	30	245	72	256	16	182	113
14	0.114502	4	1	3	4	0	236	34	0	160	166	161	22	160	160	22	163
15	0.09198	4	2	2	3	1	48	152	141	126	150	1	0	144	138	0	1
16	0.101807	3	3	3	2	1	229	112	120	137	208	0	22	201	226	96	90
17	0.092346	4	3	2	4	1	63	134	132	91	129	0	0	119	128	21	31
18	0.108459	5	2	2	3	1	200	128	133	100	179	80	101	125	165	57	168
19	0.08783	2	3	1	2	0	103	98	102	101	128	109	59	21	43	21	64
20	0.062134	0	2	1	1	0	112	157	157	112	0	80	80	0	0	80	80
21	0.077271	0	1	2	2	0	144	0	0	144	144	0	0	144	134	0	0

Table 2 Discretized data table.

Attribute→	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	1	15	3	3	3	3	1	21	29	1	1	33	45	1	21	42	47
2	70	15	5	3	3	5	30	55	34	42	21	48	34	53	22	34	9
3	11	5	3	3	3	1	10	34	26	41	59	1	1	12	57	1	1
4	67	5	5	5	5	3	73	44	32	1	47	1	21	59	45	1	1
5	81	9	7	3	3	1	85	37	42	45	69	30	40	38	67	8	6
6	76	17	5	5	1	1	33	81	18	81	21	72	34	32	22	76	16
7	72	15	7	3	3	1	5	69	66	24	39	13	20	32	45	31	31
8	94	22	1	3	3	1	58	51	58	71	36	43	53	56	38	46	48
9	99	22	1	3	3	1	86	82	86	87	71	76	79	76	68	77	81
10	2	11	1	1	1	1	1	1	1	77	1	1	1	64	1	1	1
11	91	19	1	3	3	1	58	60	23	5	36	57	59	1	38	68	80
12	55	17	3	1	3	1	20	79	10	1	18	69	1	1	18	73	1
13	95	17	1	7	7	1	40	72	15	87	12	74	28	76	8	69	46
14	79	15	1	5	7	1	81	10	1	69	41	59	7	61	41	12	64
15	46	15	3	3	5	3	7	62	66	51	38	3	1	56	37	1	3
16	63	13	5	5	3	3	80	37	50	56	57	1	7	69	62	34	36
17	48	15	5	3	7	3	9	53	58	28	31	1	1	51	31	10	12
18	73	17	3	3	5	3	72	48	60	35	49	24	43	52	43	19	67
19	38	11	5	1	3	1	27	30	39	36	29	37	21	16	13	10	25
20	7	7	3	1	1	1	36	63	72	45	1	24	31	1	1	24	29
21	21	7	1	3	3	1	50	1	1	61	36	1	1	56	35	1	1
22	6	11	1	3	3	1	41	1	1	50	23	14	14	41	5	43	48
23	14	15	1	7	7	1	12	22	30	18	20	23	30	37	18	29	34
24	12	13	1	3	3	1	19	17	22	32	1	45	54	1	5	49	53
25	5	9	1	3	1	1	50	9	13	62	10	51	62	13	1	30	35

IV. REDUCTION OF ATTRIBUTES

Use of rough sets theory in the preprocessing stage results into dimensionality reduction and optimized classification with removal of redundant attributes. Also, neural network is the most generalized tool for pattern recognition and has capability of working in noisy conditions also. Here a new Rough-Neuro Hybrid Approach in the pre-processing stage of pattern recognition is used. In this process, a set of equivalence classes, which are indiscernible using the set of given attributes, are identified. Only those attributes are kept which preserve the indiscernibility relation and the redundant ones are removed, as they do not affect the classification. A reduction is thus resulting in a reduced set of attributes, which classifies the data set with the same efficiency as that of the original attribute set.

A. Steps for finding the reduced set of attributes.

Consider an information system $IS = (U, A, V, f)$ where *U:* is the universal sets containing all objects i.e. $U = \{x_1, x_2, x_3,\ldots., x_n\}$, *n* is the total number of objects; $A = CU\{d\}$ where *C* denotes the set of condition attribute and *d* is the decision attribute; *V* denotes contains the sets of values each condition attribute can take; f is a function between the element in *U* and its value, the value of object x_i to the attribute *a* is $a(x_i)$. The total number of condition attributes is *m* i.e. $|C| = m$. The number of decision classes is *t* i.e. $|V_d| = t$. Discernibilty matrix $(t\times t)$, whose entries contain the relative significance of each attribute. In this method the relative significance of each attribute in discerning between two compared classes *x* & *y* is as given by:

$$P(x)_{x,y}|a_i \qquad (3)$$

where $P(x)_{x,y}|a_i$ is the probability of an object belonging to class *x* given the only information as attribute a_i $(i=1,2,3,\ldots,m)$ when discerning the objects of *x* from *y*. If the value turns out to be 1, then the particular attribute is the most significant and if the value turns out to be 0, then the particular attribute is the least significant. The probability described is subjective and there can be other viewpoint. The rough sets deal with uncertainty of data sets or information granules.

B. *Steps for algorithm*

The algorithm used is as follows:

Step 1. Obtain the information system whose feature space is required to be reduced.

Step 2. The discrenibility matrix (t×t) is to be constructed and entry for each attribute is given by the above method.

Step 3. The relative sum for each attribute is obtained i.e. the contribution of each attribute over the table is summed up.

Step 4. The most significant contributors based on the relative sum are selected according to the requirement or set threshold.

V. ROUGH-NEURO BASED HYBRID APPROACH

The results achieved by the hybrid approach are compared with the pure neural approach. The type of ANN used in both the cases is unsupervised with two layers. When pure neural approach is used, the input layer consisting of thirty-one nodes for thirty-one input features, fed for classification is used. Whereas the output layer contains twenty-six nodes for twenty-six output classes identified in which the input samples are to be classified. The data table used for training is formulated steps explained earlier. Many such tables with variations in data patterns were used for training and testing. The training algorithm used is (of competitive learning type) Winner take all [6-8]. Rough set approach used for preprocessing or attributes reduction and the downsized set of attributes can be fed to the neural classifier as shown below in fig. 3. Reducts derived out of input vectors using rough set postulations ease the process of making predictions and decision making which in turn gives improved classification with reduced dimensionality of feature space. For estimation of reducts discernibility matrix is first calculated, weighted contributions obtained and the significant contributors are selected as per the fixed threshold. Thus rough set mainly exploits discernibility and hidden inconsistency in the data. The results shown here are for the same table used earlier for the first approach.

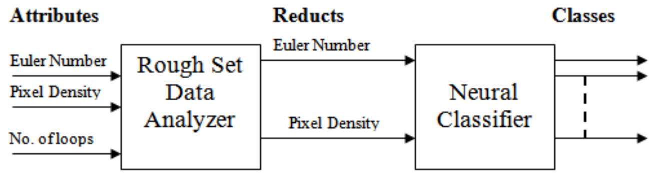

Figure 3 Block diagram showing the rough neuro classifier system

VI. RESULTS

The data set generated is given to a pure neural network described as earlier with learning rate α=0.4 and the network converges in almost 1000 iterations. The attribute set was then reduced by the proposed method and fed into the neural network with the set learning rate and iterations.The graphs below are obatined for four type of reductions of the attributes space using the above discussed method.

The horizontal axis shows the number of resulted reductions and the vertical axis shows the corresponding efficiencies. From the graph below, we conclude that the attribute reductions result in the optimum effiecincy. Thus the total 31 atributes are reduced to 21 and the efficiency obtained is still similar to that of pure neural approach..

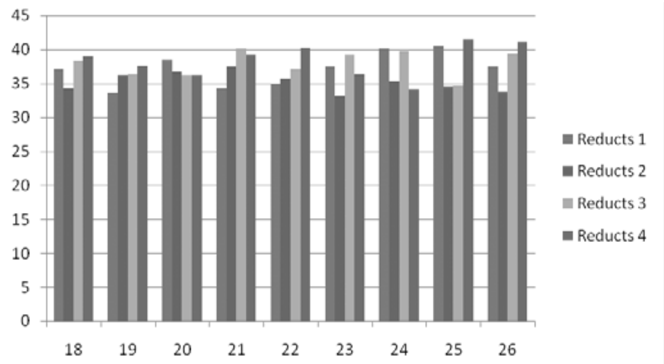

Figure 4 Classification efficiency for different methods for reducts for pure neural network.

Now, it is observed that at the point of dimentionality reduction certain structural attributes which contribte to the discernability of the system are neglected. Forcefully introducing them into the reducts yeilds a better efficency. This is shown in table 3. Note that the structural features are added to the reduced features which are 21 in number. Here the neural classifier used is of lower accuracy and efficiency and the main focus is towards the benchmarking of hybrid approach with the pure neural approach. Same results can be achieved for high accuracy and efficiency neural classifiers also such as rough SOM based or of other types.

Table 3 The type of data with corresponding classification efficiency of network.

The kind of data given to the network.	Entire Attribute space to pure neural network	Reduced attributes from the reducts algorithm	Reduced Atrributes with loops	Reduced attributes with loops and crossings.
Efficiency	43.27 %	39.46 %	41.69 %	42.03 %

VII. CONCLUSIONS

The data set generated contains both the consistent and inconsistent values.According to the rough sets theory consistent attribute values contribute less towards classification but on observation the important structural features such as the loops and crossings when included improves the efficiency of classification. Thus using rough set theory and analysing the importance of features from classification point of view efficient reduction algorithms can be developed. The need of such unconventional methods are justified by the improvement in quality of the approximation by the network. The hybrid approach discusses a new method of obtaining reductions of the data set without evaluating the reducts and cores as presented by the theory. The algorithm works even in situations when the number of reductions required are known and fixed due to the limitations of a system.

REFERENCES

[1] Rafael C. Gonzalez, Richard E. Woods, Steven L. Eddiins : "Digital Image Processing using MATLAB", First Impression, 2006, PEARSON Education, pages 348-497

[2] Jianming Hu, Donggang Yu and Hong Yan: "Algorithm for stroke width compensation of handwritten characters", Electronics Letters Online No 19961501

[3] Hongsheng Su, Qunzhan Li: "Fuzzy Neural Classifier for Fault Diagnosis of Transformer Based on Rough Sets Theory": IEEE, CS, 2223 to 2227

[4] Giorgos Vamvakas : "Optical Character Recognition for Handwritten Characters": National Center for Scientific Research, Demokritos Athens - Greece , Institute of Informatics and Telecommunications and Computational Intelligence Laboratory (CIL)

[5] Jiang-Hong Man: "An improved fuzzy discretization way for decision tables with continues attributes": Proceedings of the Sixth International Conference on Machine Learning and Cybernetics, Hong Kong, 19-22 August 2007

[6] S.N.Sivanandam, S. Sumathi, S. N. Deepa: "Introduction to Neural Networks using Matlab 6.0" first edition, 2006, Tata MCGraw Hill, pages 531-536

[7] J.S.R. Jang C.T Sun E.Mizutani : "Neuro-Fuzzy And Soft Computing A Computational Approach to Learning and Machine Intelligence", First edition reprint 2005, Pearson education, pages 327-331.

[8] Xian-Ming Huang, Ji-Kai Yi, Yan-Hong Zhang: "A method of constructing fuzzy neural network based on rough set theory", International Conference on Machine Learning and Cybernetics, 2003, Publication Date: 2-5 Nov. 2003, Volume: 3, On page(s): 1723- 1728 Vol.3.