ADVANCES in NATURAL and APPLIED SCIENCES

ISSN: 1995-0772
EISSN: 1998-1090
2016 September 10(13): pages 193-200

Published BY **AENSI Publication**
http://www.aensiweb.com/ANAS
Open Access Journal

Non Linear Dimensionality and Attribute Reduction of IRIS Code Based on Transform Domain

[1]G. Uma Mageswari and [2]Dr. M. Indra Devi

[1]Assistant Professor, Computer Science and Engineering, Bharath Niketan Engineering College, Theni, India
[2]Professor & HOD Computer Science and Engineering Kamaraj College of Eng &Tech, Virudhunagar, India

Received 7 June 2016; Accepted 12 September 2016; Available 20 September 2016

Address For Correspondence:
G.Uma Mageswari Assistant Professor / CSE, Bharath Niketan Engineering College, Theni, Tamil Nadu, India.
E-mail: g.umamaheswarime@gmail.com

Copyright © 2016 by authors and American-Eurasian Network for Scientific Information (AENSI Publication).
This work is licensed under the Creative Commons Attribution International License (CC BY).
http://creativecommons.org/licenses/by/4.0/

Open Access

ABSTRACT

In this paper, the new dimension of features in feature vector is reduced with entropy scheme and T-statistics approach. We are propose an iris recognition method based on the new technology using Genetic Algorithm (GA). We are using the data is actually not a sequential order. Using these approaches are very easy to apply the data set. Lot of differentiation are having each generations, but the IRIS data must also unique. The normalized output of entropy based approach and T- statistics approach are passed as candidate for feature pool. The IRIS data normally having huge amount of data, now we are using the techniques for feature extraction[1].Most of the traditional fesature selection or extraction schemes require sufficient number of sample data as per subject to select the most representative feature sequence . So the Unique identification must have the details about IRIS. We have developed an iris recognition method based on genetic algorithm for the optimal feature extraction, with the cost of eye scanning technology. The hybrid feature subsets are selected with the Genetic Algorithm (GA) based feature selection algorithm. The proposed Genetic Algorithm [12](GA) based feature selection approach has been experimented with cassia iris database and it reduces the feature dimension without compromising the recognition rate. Many critical genetic algorithms are followed by the iris data, but we are applied some simple algorithms to apply the genetic algorithm to present the above methods. Many iris data are the best alternative procedures proposed for this data set. Once again we are saying the Genetic Algorithm seem to constitute good choice for applying the attribute reduction technology.

KEYWORDS: Feature vector, IRIS, Statistics, Genetic Algorithm

INTRODUCTION

Processing images sourced from various imaging devices usually involves dealing with large numbers of features. The number of features in an image processing task is usually augmented by extracting features. There are two major approaches for dimensionality reduction. The first approach is through feature construction where the features in the reduced input space are functions of the original features. Examples of this approach are principle component analysis (PCA) and evolutionary-based feature [3] construction. The second approach is through feature selection (FS) where a subset of original features is selected for learning.Applying the two types of feature selection first one is the entropy feature selection method the second one is the T-Statistics feature selection method. If we are gathering any one of the methods, then proposed to the feature optimization.And the Genetic Algorithm was applied. If we are having the improved feature data the next step to calculate the entropy based methods. The iris data contains a large number of textural features and a comparatively small number of samples per class, and this makes accurate and reliable classification [11] [15] challenging. In this work, we incorporate the power of FS in searching the space of

To Cite This Article: G. Uma Mageswari and Dr. M. Indra Devi., Non Linear Dimensionality and Attribute Reduction of IRIS Code Based on Transform Domain. **Advances in Natural and Applied Sciences**. 10(13); Pages: 193-200

functions and features[3] and to find a subset of features that maximizes the complexity in an iris recognition system.

Related Work:

The selection of the most representative feature subset selection of the original features set with a relative high dimension is another important issue in the field of iris recognition [2]. The iris data usually contains a huge number of textural features and a comparatively small number of samples per subject, and thus, it becomes difficult to classify the iris patterns accurately and reliably. The Genetic Algorithm(GA) is used as an attractive approach to solve this kind of problems in they are generally quite effective in rapid global search of large, non-linear and poorly understood spaces. The work proposed in this chapter useful information obtained from the various feature selection methods to choose the most high feature subset and also to improve the matching accuracy of the data set. Show techniques in Figure1.

The work proposed in this chapter utilizes the useful information obtained from the different feature selection methods to choose the most prominent feature subset[13] [14] and also to improve the matching accuracy on the set of texture features extracted [13] with orthogonal polynomial model coefficients. The Genetic Algorithm[12] is utilized to select the significant features by combining the valuable outcomes from multiple feature selection criteria. In this proposed scheme, entropy-based method and T-statistics methods are used to provide the candidate features for the selection of the optimal feature subset with Genetic Algorithm.

3. Proposed Hybrid Feature Selection Technique:

3.1 Identification of information contained with entropy:

In entropy-based method, entropy is lower for orderly configurations and higher for disorderly configurations. Therefore, when an irrelevant feature[8] is eliminated, the entropy is reduced more than that for a relevant feature. This algorithm ranks the features in descending order of the entropies after removing each feature one at a time from the set of feature vector *fv* . The entropy measure *H (x)*of a data set is computed as follows:

$$H(x) = \sum_{i=1}^{n} p(xi) \log p(xi) \quad (1)$$

Where *x* represents the elements in the dataset and *n* represents the number of elements in the dataset and *pxi* represents the probability of element *x* in the *ith* position. In this proposed work, the entropy content *H (x*) of all features[8] is obtained by equation (1).

Algorithm:

Input: Feature vector *fv*

Output: Discriminate Feature set (*DFS*)

Steps: Begin:

Step 1: Initially *DFS*= { }.

Step 2: Compute the entropy *H (x)*for *fv* using equation (1).

Step 3: Find out the entropy *H* '(*x)*by removing one feature at a time from *fv* . Step 4: Determine the entropy loss for each feature *Ei =H(x)-H'(x).*

Step 5: Repeat from step2 to step 4 for all the feature elements in the feature vector.

Step 6: Rank the features in descending order based on the entropy loss.

Step 7: Select the top 80 ranked features and store it in DFS.

End:

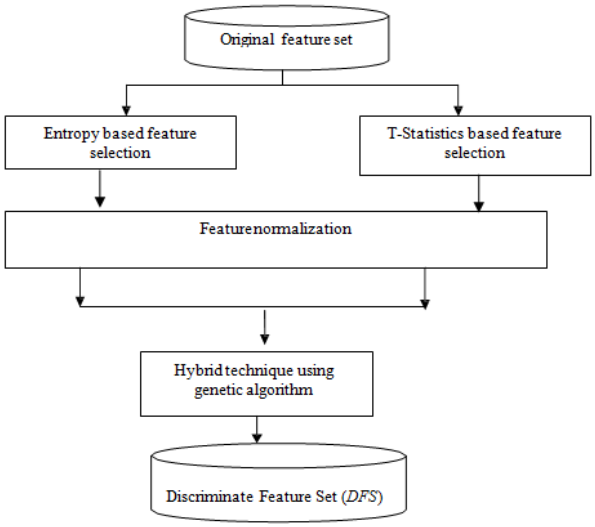

Fig. 1: Proposed feature selection technique

3.2 T-statistics method:

T-statistics is a statistical [1] measure, used to test the hypothesis that two or more examples are statistically independent in their behavior. It evaluates feature relevance by measuring the statistical [1] value with respect to the class information. From the original feature set *fv* the random features fv1 are selected with total number of features n1 and n2 respectively. The mean fv

__ *fv*1 are calculated for each feature set and the combined standard deviations *S* is calculated as

$$S=\sqrt{\frac{\sum(fv-fv)2+\sum(fv1-fv1)2}{(n1+n2-2)}} \quad (2)$$

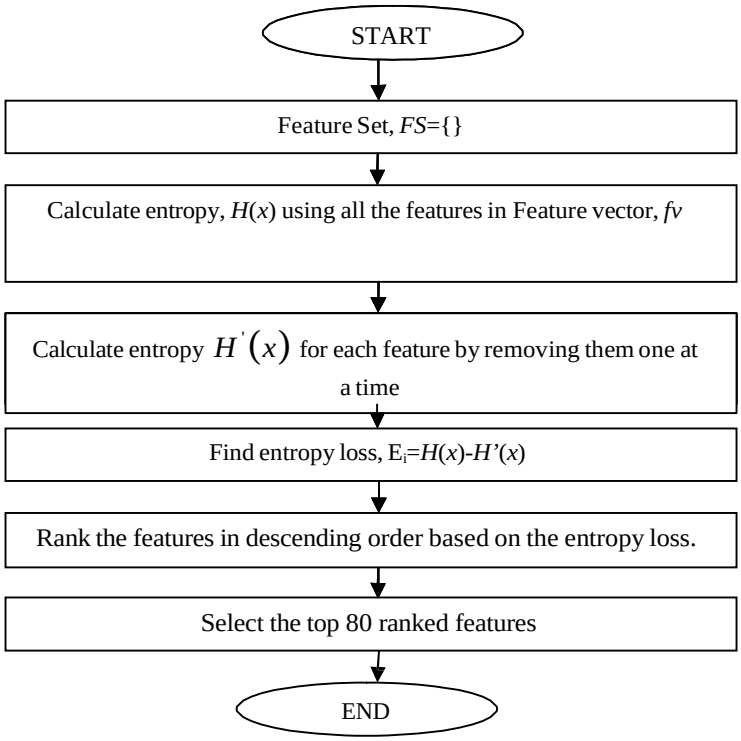

Fig. 2: Flowchart of identification of information content to entropy

Where (n1 + n2 – 2)is the degrees of freedom (dof). The *t*-test value is calculated from the mean, standard deviation and *dof* as follows

$$t = \frac{\overline{fv} - \overline{fv1}}{s} \sqrt{\frac{n1n2}{n1 + n2}} \quad (3)$$

The sample features and its *t*-test evaluation are shown below.

$$FVm = \frac{(pos)*(fvi)}{tnof} * S \quad (4)$$

where *pos* is the current position of the feature value, *fvi* is *i th* the feature value of selected

3.3 Feature Optimization:

Original Features : {2,3,4,5,7,33,34,36,37,39,55,63,69,99,103,111,167,183,231}
Random Sample Features :{5,35,55,103,231}
n1 Value :20
n2 Value : 592.8815983085337
The S value :8
Confidence Level :95%
Degrees of freedom :23
The Final t value : 0.9862018060426073
t-test table value 005 : 2.807

3.4 Hybrid techniques using Genetic Algorithm:

In this subsection, Genetic Algorithm [10] is utilized to find the prominent features based on the normalized features [7] from two feature selection [6][7] algorithms viz. Identification of Information content with entropy and T-statistics method.

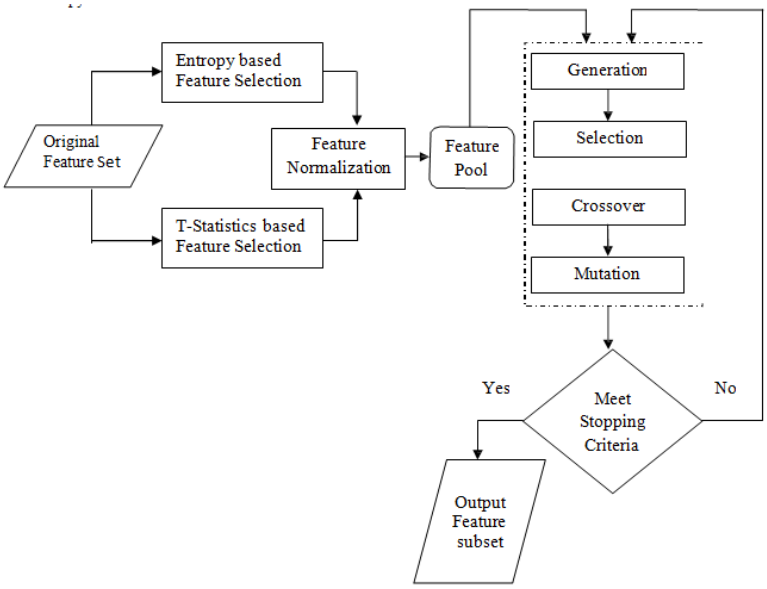

Fig. 3: Proposed Feature Selection Technique with Genetic Algorithm (Hybrid approach)

From the iris original feature set, the top ranked features are obtained from the above two feature selection [5]algorithms and forms a collection of candidate feature[11] called feature pool. From the feature pool, the Genetic Algorithm[10] is utilizwd to find an optimized feature subset by applying the genetic[9] operations such as selection, mutation and crossover. The architecture of the proposed hybrid technique with GA[12] is given in the *Figure 3*.

The Genetic Algorithm [12]searches the pool of features (denoted as population) and evaluates each feature based on the following fitness function.

$$FitVal = \frac{(pos)*(fvi)*w}{(tnof)*(Noc)} \quad (5)$$

Where *pos* is the current position of the feature value; f_{vi} is i^{th} the future value of selected five turnoff is the total number of selected features [4]; *Noc* is the number of occurrences and *w* is a constant weighting parameter. The threshold *Ta* is computed adaptively based on fixed location of the population as follows

$$Ta = \sum_{i=0}^{n} \frac{FitVal}{tonf} \quad (6)$$

where *FitVal* is the fitness value of each feature and *tnof* is the total number of selected features.

In this work, the Roulette wheel selection method is used to randomly select the individuals from the population for latter breeding. The probability that an individual will be selected is proportional to its own fitness and is inversely proportional to the fitness of the other competing hypothesis in the current population and the average fitness value of selected feature is computed using the fitness value. If the average fitness value is approximately equal to the threshold *Ta* , the sample is placed in the feature pool. Otherwise, the crossover operator is used to create the new chromosome from the selected features for latter breeding and the average fitness value of cross over feature is computed using the fitness value. If the average fitness value is approximately equal to the threshold *Ta* , the sample is placed in the feature pool. Otherwise, the mutation operator is used to create the new chromosome from the crossover features for latter breeding. In this work, the single point crossover, and two-point mutator are used. The flow chart of the proposed feature selection technique with Genetic Algorithm (GA) [12] is given in the steps

The steps involved in the proposed GA based feature selection process are as follows:

Steps: Begin:

Step 1: Initialize Population with size 160

(Combine Transformed Entropy output and T-Statistics output) Step 2: Calculate Fitness Value for every feature.

Step 3: Calculate Adaptive threshold *to* a value based on fixed location of the population

Step 4: Apply Roulette Wheel selection and select 80 features.

Step 5: Calculate FitVal for each feature and calculate Average Fitness value.

Step 6: Check average fitness value with threshold *Ta* . If the calculated fitness value is very near to the threshold, keep the sample and go to step 13; otherwise goto Step

Step 7: Apply Crossover Operation for each feature.

Step 8: Calculate the fitness value for each feature and calculate Average Fitness value.

Step 9: Check average fitness value with threshold *Ta* . If the calculated fitness value is very near to the threshold, keep the sample and go to step 13; otherwise go to Step 10.

Step 10: Apply Mutation operation for each feature.

Step 11: Calculate the fitness value for each feature and calculate Average Fitness value.

Step 12: Check average fitness value with threshold *Ta* . If the calculated fitness value is very near to the threshold keep the sample.

Step 13: If iteration is over, go to step14; otherwise go to step4.

Step 14: Sort all the samples based on fitness value and select the first sample and its fitness value.

End:

4. Experiments And Results:

The proposed feature selection method for iris recognition has been experimented with AUT- MIT iris database. Sample test image of size (752-480) with pixel values in the range (0-255) is presented in *Figure 5*. Initially, the sample test image is subjected to normalization [6] to a rectangular block with a fixed size of (648) 450. This output is presented in *Figure 6.* From the normalized image, the texture feature vector *fv is* computed with orthogonal polynomial transformation coefficients. The result of texture representation is shown in *Figure*

Fig. 4: Sample original test image considered for feature selection

Fig. 5: Rectangular normalization result of the test image shown in Figure 4

Representation of Texture:

4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 44 36 36 36

outcomes of entropy scheme, and T-statistics based feature selection algorithms as demonstrated in *subsection 3.3*.The weighting parameter used for feature normalization technique is 20 in .

The proposed GA based feature selection approach has been experimenting with the normalized output of entropy scheme, and T-statistics [1] based feature selection [4] algorithms. The weighting parameter used for fitness function is 100 for *w*. The cross over and mutation probability for the proposed technique are set as0.71 and 0.005 respectively, and the number of generation is 80. The results of hybrid approach using Genetic Algorithm corresponding to the image shown in *Figure 4* is presented in *Figure 9*.

Feature selection result with Genetic Algorithm

1600 10 75 41 12 18 8 44 45 40 54 80 76 63 60 22 64 20 28 66 88 28 68 108 63 27 5 38 92 64

19 67 100 24 20 56 80 20 96 92 76 23 ……..

Fig. 9: Feature selection result with Genetic Algorithm for the original image shown in Figure 5

5. Performance analysis:

Since the number of samples from most iris research is limited, cross- validation procedure is commonly used to evaluate the performance of a classifier. In *k*-fold cross validation, the data are divided into *k* subsets of approximately equal size. We train the classifier *k*-times, each time leaving out one of the subsets from training, but using only the omitted subset to compute the classification [10] accuracy.

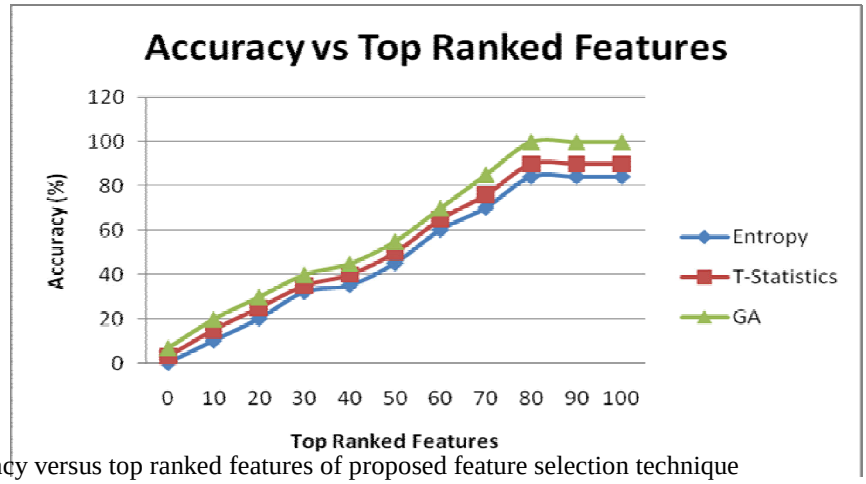

Fig. 11: Accuracy versus top ranked features of proposed feature selection technique

Conclusion:

In this paper, we present a Genetic Algorithm based feature [1] [2] subset selection for iris data set. Due to the use of different feature selection criteria, various feature selection methods often provide very different outcomes. A lot of the database is followed by many experiments, but now we are following the AUT-MIT Database.Each and every feature have the fitness value. So we additionally apply the algorithm for easy to identify their fitness value and average fitness value by the AUT-MIT data set. Applying the crossover finding for threshold value. The T-Statistics methods are correct to the hypothysed value. After the process we will get the best result and accuracy for these approaches. The proposed GA incorporates two feature selection criteria to find the subset of informative low level features that can improve the analysis of iris data. The above methods are delivering very accurate results for the classification of the iris data set. The experimental results show that the proposed method is capable of finding feature subsets with a better classification [10][11] performance and/or smaller size than each single individual feature selection algorithm does.

REFERENCES

1. Atul Bansal, Ravindar Agarwal, R.K. Sharma, 2016. Statistical Feature Extraction Based Iris Recognition System, Springer Link, 41: 507-518.
2. Mahmoud Elgamal, Nasser AI-Biqami, 2013. An Efficient Feature Extraction Method for Iris Recognition Based on Wavelet Transformation. International Journal of Computer and Information Technoloy (ISSN:2279-0764) 2: 03.

3. Dolly Choudhary , Shamik Tiwari, Ajay Kumar Singh, 2012. A survey : Feature Extraction Methods for Iris Recognition, International Journal of Electronic Communication and Computer Technology (IJECCT) 2: 6.
4. Amel Saeed Tuama., 2012. Iris Image Segmentation and Recognition. International Journal of Computer Science Engineering Technology, 3: 3.
5. Pooja Garg, Anshu Parashar, 2012. Feature Selection Method for Iris Recognition Authentication System. Global Journal of Computer Science and Technology Graphics & Vision. 12: 10.
6. Jong –Gook Ko, Youn – Hee Gil, Jang –Hee Yoo, and Kyo IL Chung, 2007. A Novel and Efficient Feature Extraction Method for Iris recognition ETR Journal., 29: 3.
7. Yong Zhang, Yan Wo, 2005. A fusion iris feature extraction method based on fisher linear discriminate. IEEE International Conference on Machine learning and cybernetics, 1: 5-9, ISSN 2160-133X
8. Seung-In Noh, Kwanghyuk Bae, Yeunggyu Park,Jaihie Kim, AVBPA 2003. A Novel to Extract Features for Iris Recognition System. Springer – Verlag Berlin Heidelberg 862-868 (LNCS 2688).
9. Roberto Baragona, Claudio Calzini, Francesco Battaglia., 2001. Genetic Algorithms and Clustering: an Application to Fisher's Iris Data, Springer Verlag Berli Heidelberg.
10. Shahamat, H., A.A. Pouyan, 2015 Feature Selection using genetic algorithm for classification of schizophrenia using fMRI data. Journal of AI and Data Mining 3 :1.
11. Mathumitha Ramamurthy and Dr. Ilango Krishnamurthi, 2016. Decision Tree Based Classification Type Question / Answer E- Assessment System. Advances in Natural and Applied Sciences, 10: 22-25.
12. Nandhini, K., S. Pushpalatha, 2016. QoS Based Service Selection using a Genetic algorithm and Association Rule Mining in MANET. Advances in Natural and Applied Sciences, 10(4): 35-43.
13. Ruba Arockia Archana, S., Dr.M.S. Thanabal, 2016. Feature Selection for Heart Disease using Enhanced Cuckoo Search , Advances in Natural and Applied Sciences, 10(4): 50-56.
14. Maruthu Pandi, J., K. Dr. Vimala Devi, V. Anitha, 2016. Efficient Feature Extraction for Text Mining . Advances in Natural and Applied Sciences, 10(4): 64-73.
15. Anusha, D.N., R.R. Bhavani, Classification of Varicose Ulcer Tissue Images, Advances in Natural and Applied Sciences 10(4) :227-232.