A Study of Feature Subset Selection Methods for Dimension Reduction

Saqib Hayat[1], Abdul Basit Siddiqui[2] and Sajid Ali Khan[2]

[1]*Shaheed Zulfikar Ali Bhutto Institute of Science and Technology (SZABIST), H-8/4 Islamabad, Pakistan*
[2]*Department of Software Engineering, Foundation University, Rawalpindi, Pakistan*
itsaqi@gmail.com,sajidalibn@gmail.com, absiddiqi@hotmail.com

Abstract

The interest and focus for quite some time has been on Feature Selection and lot of work has been made in this field. With databases getting larger in volume so machine learning techniques are required which results in demand for feature selection. Feature selection is commonly used method for performing data mining in the field of data preprocessing that is scaled on large amount of data sets. In this paper, several kinds of feature selection methods are used which may result in different subsets of features with evaluation criterion.

Keywords: Feature selection; dimension reduction; feature extraction; feature optimization

1. Introduction

In this new era one can get very useful kind of input from a great amount of data in the form of information that could be really informative or interesting. Data can be gathered around from a suitable environment and that can be classify accurately in class or other, fields that do involve in such criteria is medical sciences and engineering. All of data gathered from such fields and applications like industrial or scientific or even commercial are becoming very tough for a man kind to understand and make something out of it as to them it doesn't mean something that makes no sense. As this data is getting very difficult to understand so scientists or researchers are moving towards computers to gather a data or generate a data which would make lot of sense to them, even business are using computers for databases maintenance like to information about fir their customers, products which they are selling is also maintain by database for stock and inventory. Now a day's even buyer's uses computers in to buy a product online on the internet and don't have to go the shop to collection it. This the real use of computers in modern arena. Bioinformatics a field of medical but that needs computer scientists to involve in equipments for their survival and this is a new field of science embrace to computers and computer technology, field is generate by computer scientist and biologists and both have to work together to achieve resolution for more complex problems to design and perform high diagnosis system Gathering informative data are mostly done through fields such knowledge discovery and data mining, large amount of data in sense of medical information is digitally stored in such way that other can get benefit from this data stored and get diagnosed and treatment when ever this data is sued for such purpose. Prediction models of high performance do attractive practitioners; lives can save from fatal diseases by use of optimal classifier in proper channel [1].

Pattern recognition or classification also fall into the field of science involves in given patterns as an input and these pattern are classified into different kinds of classes. This

classification process of different classes is being used in widely areas of applications such as computer vision, bioinformatics, digital image processing (DIP), computer-assisted decision, analysis of medical data, machine learning, automatic processing of mail or bank receipts, management science, quality control and financial engineering. Classification of patterns is further divided into three sub classes such as preprocessing of data, representation of data for the generation of feature or for extraction and classification or recognition of data.

The main objective of pattern classifications to further divide or classify given input patterns or data into different kind of classes, like in the case of gender classification only two kind of classes exist one is male and the other is female and the given image of face is associated to one of these two classes that is male or female. Classification of classes can be improved or in other words are dependent on the quality of system that used the input data and classifies it. Regarding machine learning process, gender classification field is interesting with two way classes but is some time very complicated classification area or system problem. Considering this problem that is inter communicated between human and computer, the created of optimal classifiers is introduced to overcome such problem and to enhance the performance of face recognition. The work of optimal classifiers is in main focus of this thesis.

To analyze the problem of classification a suitable and statistical techniques are used, also lot of other techniques that are based on intelligent decision also used in wide range like is case of expert systems; fuzzy logic and sets, decision tree based, evolutionary Computing and artificial neural networks (ANN) and thus these techniques that are more intelligent are comparatively more effective in applications than that of traditional statistical [2]. High performance is more offered in case decisions based intelligent techniques and meet their given requirements more easily as are focused or tuned up in achieving it results according to the application. Therefore more accurate decision can be obtained from such intelligent techniques compare to expert in mankind.

2. Approaches for Suitable Features Design

The main focus of these approaches is on to select or to extract a feature in such a way that the informative part is preserver the most and eliminates an inappropriate or surplus data or components. The stage of extraction of features for classification is to transforms the data. The performance of classifiers can be improved, representing a faces as a features subset has been a difficulty in assuming that what are the features that need to be selected and what are the relationships between these features suitable for classification. When a classifiers are trained on set of well known images, so to reduce the dimensions of an given images or data to allow results in accurate with training data set so for this reason methods used are such as selection of feature and extraction of feature.

2.1. Pattern Recognition

In pattern recognition and statistics extensively areas have used methods of feature selection studies by society of Machine learning and data mining. Feature selection is considered to be an issue in algorithm of machine learning with respect to design. The objective to use a proper method of feature selection and extraction within given input data of reduced dimensions in order to get optimal performance.

2.2. Removal of Features

Sometimes removal of features from a given input has significant benefits, collecting some feature are very costly so in this scenario elimination of features can make the classification

more valuable and efficient to end user. Like in given example for medical diagnosis system, tests from thyroid disease may consist of features and also may not comfort the weak patient in addition to cost as well, so a risk factor may be attached to these features to some instance. Some of thyroid disease features can be removed without accuracy reduction of diagnosis system and as a result patient risk can be minimized.

2.3. Feature Extraction

Problems regarding feature may not be known in many application. Candidate can present better to the domain to solve problems and wide amount of candidate features are selected for this purpose, which result in few of feature extracted be irrelevant or redundant. Feature selection is of two kinds that is a filters and a wrappers framework. In filters, feature set performance are measures in such a way that classification algorithm are not included. Algorithm for feature selection filter without referring classifier on training data uses performance metric. In regards to wrapper, classification algorithm is incorporated.

3. Literature Review

Our main Topic covers the area of feature selection which has significant importance in areas like signal processing, data mining, image processing, text categorization, pattern recognition, remote sensor image recognition and many more. The ideal purpose of feature selection is to perform a heuristic or random search and find an available feature of subset by eliminating needless features. In this particular paper [3] author described the lessons of selection of feature method base lying on ant colony optimization (ACO) also known as ACOFS is implemented. The algorithm, artificial ants go over on top of a digraph by way of simply 2n arc. Within this algorithm author take taking place classifier concert as well as heuristic in sequence resting on the quantity of the preferred features, to select the most optimal and accurate element subset solution into provisions of the feature deposit volume plus classifier performance. Now the most important question is how to find optimal feature subset, for feature selecting the majority of algorithms that are based on ACO do normally use a complete graph by which all means that the trail created by ants to be the element of nodes. First feature to be selected is independent with the last feature added to the partial solution and is a feature selection problem complete graph with edges in the ACO algorithm is needless. At this moment through apply ACO algorithm on behalf of element selection supplementary proficiently, the representation graph used should be more redefined. Base scheduled our ant's resolution elected feature breaking up; classified training data sets have been used to find quality solution within expressions of categorization precision which be evaluate. Correctly classified in test accuracy are measured by the number of examples. In addition, quality function in quantity of facial appearance here the places are furthermore well thought-out. In these experiments, author used system P 4, Win XP, P 1.7G, via VC++ 6. 0, as well as visualize each and every one these consequences on top of Matlab 6.0. Round of 80 images of data set in 4 classes, 19 features of data set for primary along with secondary organize starting point instant are included, now applied ACFS algorithm on image set to pick relevant features, approach based on GA is compared and for feature selection ACO algorithm is modified. Standard bring to mind of ACOFS be 100% here 610-fold CV analysis lying on trying statistics whereas 87.5% in favor of GAs along with meant for ACO is 98.61%, as a result ACOFS are much better than ACO and GAs algorithms.

In this paper [4], author has proposed a method named as jointly optimizes a model of subspace in excess of parameters and picking the features of an image. Recreation rough calculation is proposed due this problem NP-hard. Gradient-descent is its base plus quadratic

encoding, two optimization strategies evaluate. Subspace selection setback as a thin eigenvector estimation has been reformulated by author and therefore solved this with the help of using a backward selection algorithm. In general example, we can say that the ideal purpose to choose a breaking up of these features that best contest a model which in sense of an image (*e.g.*, shape, appearance or dimensions *etc.*,). Author has previously reviewed work on two main topics that are related to this paper *i.e.*, the identical of an image and selection of features. Now match of an image, minimization of the distance to subspace mock-up of a subset of features chosen by author to extend previous approaches. Now taking feature selection into consideration, author describes that by selecting better encoded of subset of image features with a mock-up can be pose as a feature collection difficulty. To overcome this optimization problem describe by author to opt for a subset of facial appearance that make light of the expanse to a mock-up and for that reason how to put together these shape and outward show models. To detect image features author uses Harris corner detector, Harris detector manually corresponds to the label landmark of a look, they've computed information on the Multi PIE. Now two different models been built by author that is appearance and shape. Harris detector has been used for facial features detection and similarly three modes of deviation of the figure mock-up. Viola and Jones face detector has been used by author for prevent from spreading in addition to scale-normalize the face in difficult intention. The main objective and intention of this paper is to build up an algorithm to facilitate optimally decides on the subset of landmarks from an image feature with the aim of minimizes the expanse to a subspace mock-up. Two algorithms proposed by author immature gradient-descent within addition in the direction of a quadratic encoding algorithm. Regarding naive gradient-descent, the resultant be obtain as a result of select the point so as to keep up a correspondence to the superior maximum value of all one chain of P everywhere this P be parameterized at the same time as the Hadamard invention of two matrix. Within this experimentation, author compared the results of all three proposed algorithms with respect to performance and problems of facial feature detection Images of 1500 are composed for test set and database used is MultiPIE. Author added a noise in actual data set. Ran three proposed algorithms with appearance and shape models, by selected algorithm has been observed to facilitate present be supplementary regularity among have inferior energy in addition to in good health outcome. Obtained better performance by quadratic programming and greedy approach is followed it as well. The rate of error in general has increased as exact landmark position detection where does not Harris detector doesn't properly detect.

In this paper [5], author provides history of face detection and how it started from 90's and up to today's modern world. Gender classification is the main goal in this paper and is important in applications such as electronic marketing. Digital Camera shows the gender of a person in front of its display range and its respected age. Methods involving this procedure is not a simple task as faces depend on illumination, facial expressions, gestures, pose and other highly variability. Face appearance directly affects its detection and classification. Authors in this paper reveal a report of analysis on Gender classification research from the experts from 90's to till 2011. All researchers used different methods like classification of genders. Gender categorization on top of the FERET database on the whole outcome finest be 99.07%. Feature selection criterion uses Mutual information (MI) as its robustness in noisy environments, redundancy within random variables and good representation of relevance and data transformations. Choice of feature base on top of MI moreover fusion of three facial appearances for gender categorization reported in this paper. Features extracted fusion showed from six different methods that has improved significantly classification performance and compared the domino effect of these three unlike joint in row procedures: minimum redundancy along with maximal relevance (mRMR) (Ding and Peng 2003), normalized

mutual information feature. The best subset of relevant features given in the form of input data which also only consist of useful information and is used to distinguish one class from the other for selection mechanism. In many real time applications, reduced computational time is important for essential classification results. Feature subset selection is the main objectives that characterizes the best targeted results using different forms of methods, maximal relevance and minimum redundancy, normalized mutual information feature selection (NMIFS), provisional shared in sequence feature choice. Author used two databases for their experiments and results *i.e.* From FERET database, 94.3% was the best result in gender classification with 24*24 face size and with 36*36 result is 95.6% and finally the best result with dimension of 48*48 is 96.78%.

Ojala, the researcher proposed a method Local Binary Pattern (LBP) feature extraction that successfully has been used in a number of applications whereas Sun *et al.*, use LBP in support of iris feature extraction. Allocate patch transversely everyone in excess of an image so with the purpose of image is enclosed by using standard way of LBP-based feature extraction. All patches have the same uniform size so that overlap is not possible. In this paper [6], author created a process base happening genetic along with evolutionary computing (GEC) with the purpose of evolve LBP extractors with the aim of include irregularly scattered; overlap patches so as to possibly will not envelop the intact image. Simulated evolution base technique to solve a general problem is GEC. GEC develops inhabitants of folks, breeds these individuals in the direction of construct add up to of offspring. The offspring are built-in into the inhabitants as a result of replace a quantity of weaker persons. This procedure of growth is repetitive in anticipation of a user-specified stop circumstance have be reach *i.e.*, opt for parents; create children, furthermore incorporate the children interested in the inhabitants. For LBP method, usually gray-scale image is used as described above by author. Apply LBP headed for each one pixel of to facilitate exacting patch outcome here a histogram which represent the feature distinctiveness of a territory. GEC use during this article is a Steady State Genetic Algorithm (SSGA) moreover is used in the direction of evolves an inhabitant of applicant feature extractors (FEs). First of all population of candidate FEs is generated randomly, evaluate in addition to assign robustness. Whilst the preliminary inhabitants comprise be produced, two parents chosen by the use of dual competition collection in addition to form one children through uniform crossover along with Gaussian mutation. The dataset used here author trial was collected of 105 subject in use commencing the FRGC dataset, each one subject matter have a entirety of three images in use with the purpose of various to some extent. In research to facilitate used dataset was collected of 105 subject obtain as of the FRGC dataset, total of three images to each subject got that varied slightly. Images had dimension in which width of 100 pixels as well as with the purpose of height of 127 pixels, typical Local Binary Pattern Method (SLBPM) in the company of the Genetic-Based LBP process was compared and set of 24 patches of uniform, non-overlapping, and entire image covered were used for SLBPM. Accuracy rate of 99.04% for SLBPM used all 100% of the patches. Patches of average is 36.90% whereas 99.84% is its accuracy average.

Face detection is used in such fields as criminal identification, airports, employee identification *etc.*, for authentication purpose. Face detection also plays an important role in biometrics. Face detection uses various methods in biometrics but some of these methods won't be efficient much. Features extracted beginning the appearance that is face, teach the dataset among face this extracted facial appearance and from the given input comparison is perform with these features. The input data can be any of these image, video *etc.*, Ojala, the researcher proposed a method Local Binary Pattern (LBP) feature extraction that successfully has been used in a number of applications whereas Sun et al in case of iris used LBP for feature extraction. Allocate patches transversely every single one in excess of an image as a

result to image is enclosed by using standard way of LBP-based feature extraction. All patches have the same uniform size so that overlap is not possible. LBPs are considered only the uniform, gives uniform LBP at a pixel if the transitions between "0" and "1" are related binary tag order have no more than two along with all pairs such as "10001111" along with "00011000" are uniform LBPs nevertheless "01001111" is not a the same LBP at the same time as it has four transitions. At the moment on the way to overcome this issue author has proposed to utilize prevailing local binary patterns (DLBPs) is believe the nearly all normally occur pattern within a face image. In this paper [7], author proposed a prearranged algorithm which used to find out the utmost prevailing LBP pattern within an image. LBP is determined based resting on middle pixel value moreover the region pixel values. For classification point of view, Support Vector Machine (SVM) is an admired method to be considered which performs map of statistics interested in a superior dimensional aspect space whereas separate data used linear algebra and geometry with the aim of is simply distinguishable by means of nonlinear set of laws into the put in space. Separate face and non-face images used SVM precisely. Area particular choice of a kernel function is allowed by SVM. Mostly used kernel functions are as like polynomial, linear in addition to RBF kernels. Experiments are experienced method on top of the MIT+CMU front face trial deposit authors own database, more than 2500 of faces are used. Face of non face images were used to train detector and images were scaled to 32*32 in dimensions of facial ones, performed feature selection procedure and observer 200 of sub problems by extracting 10% each time Now defined and performed a training set, validation and test sets, DLBP trained on training data, DLBP tuned going on the validation set used for regularization limitation and after that as a final point evaluated the concert of categorization scheduled on trial set ROC curve constructed over test set. Detection of accuracy is increased with the use of SVM classifier and the reduced the false detection rate.

Face detection is used in such fields as criminal identification, airports, employee identification *etc.*, for authentication purpose. Active Shape Models (ASM) is a face model; several points within face are preferred from an input image other than a few point considered to be very helpful in favor of characterizing the face at the same time as others comprise some sort of noise in them. Some improvement has been made in ASM techniques but still the corresponding errors are generally far above the ground at some face location. As a result of weighting the features according on the way to a process base the performance of ASM was improved on adjusted mutual information. Author in this paper [8] present a wrapper meant for choosing the majority pertinent features for face identification. Author used a Genetic Algorithm GA to discover the space of sufficient solution, guide through the accurateness obtain within a face categorization job. Given example of image in this paper, authors describes to facilitate the shape are forced through these point distribution models (PDM) toward differ simply according toward an instruction set of example. Points set present the shape of an object and algorithm match the mock-up to a fresh image, ASM is at the moment taught taking place a set of teaching face images. Genetic algorithms (GA) are meta-heuristic optimization methods; GA consists of three dissimilar operators: selection, variant and substitution. Unlike choice of facial appearance extracted since an enter image as a result of means of ASM, the anticipated wrapper consists scheduled a traditional GA in the company of binary chromosome. Therefore now fitness function uses a classifier algorithm, thus with the intention of the achievement categorization rate is assign because fitness value for every evaluate entity. Classifier base on top of the k-nearest neighbors (KNN) has also been used within classify toward calculate its sorting presentation scheduled on the trial set. For better generalized capability to be obtained in further analysis, author used Smoothed Bootstrap Resampling (SBR) process toward broaden the teaching figures set. Set of images taken from

Essex Face Database be used within these experiment. In experiments, Essex Face Database is used with set of images. In first experiment, GA-136+MEAN is referred and chromosome size 136 is used probabilities of crossover set to 0.8 and mutation to 0.025. Performance of 97.2% is achieved on test data set with KNN classifier. In next trial GA-136+MEAN+SBR be referred in addition to used 20 examples toward appraise the presentation of categorization throughout optimization process, 68 features of GA converged and obtained 97.4% success rate on test data. In third experiment, GA-68+MEAN+SBR is referred and obtained accuracy classification of 98.0% on test data, best solution of after 63 generation was found.

Author in this paper [9] presents a thought of the characters of Binary Particle Swarm Optimization also recognized like Accelerated Binary Particle Swarm Optimization (ABPSO). Face Recognition technology has been dealing with optimization problem of Feature Selection. Over last 10 years or so Feature Selection has been the in eye of researchers which is being vastly used in many fields such as medical, electronics etc. Author in this paper has highlighted 3 main techniques although there are lot more techniques available regarding Image Pre-processing. Three techniques mostly are such as Edge detection using Laplacian of Gaussian (LoG), Histogram equalization (HE) as well as Bi-Cubic interpolation. The very first technique Image interpolation which produces high resolution/quality images from a low resolution. Bi-Cubic interpolation is far better than bilinear algorithm as it increases the sharpness of an images and conversion of resolution. The second technique is Histogram equalization which highlights the brightness area so that it is more visible to human and is a nonlinear process. Obtaining efficient and detection resolution edges, Cropped Yale B database has been applied by Log. With Discrete Cosine Transform (DCT) higher rates of recognition with computational lower complexity are achieved and have great advantages in energy compaction. Potential solutions in PSO are considered to be particles which tend to fly in the difficulty space follow through the existing most favorable solution. Author in this paper proposed Binary PSO of an improved version also known as ABPSO on the way to act upon intended for multimodal problems. The scatter index in each generation is estimated where fitness value is returned every particle is evaluated en route in favor of function S of scatter index. Regarding experiment for ORL Database, this has 400 images to 40 separate individuals. For training set experiment randomly 4 images for every individual are preferred in addition to meant for testing remaining 6 images are used. Feature Selection based on ABPSO algorithm on different sizes of DCT coefficients is tested with feature vectors. Furthermore these experiments were carried out and DCT coefficient matrix of 1/4th of the upper left is considered which covers most of the low frequency components. The obtained results of 50% reduction in the number are found to yield higher recognition rate of selected features.

Designing systems of classification of patterns, set of features chosen appropriate is critical in framework of supervised learning. In this paper [10] author describes that most of the modern day researchers are focus in deploying feature selection in application like gender classi1cation, face detection, vehicle detection and many more but have ignored the problem of feature selection. The main idea for selection mechanism to find features which suits the particular problem and solves by selecting a subset of size with smallest classification error. Two components describes by author in this paper for feature subset selection: Search Strategy and Evaluation Method. The first method search strategy is further divided into three sub parts: optimal, heuristic and randomized. Search approaches like relief algorithm and other extension are randomized search, process of sampling or steps involved are probabilistic. Heuristic feature's two most commonly known selections be Sequential forward selection (SFS) as well as sequential backward selection (SBS). SFS at the beginning takes an empty feature as a start set which later selects one the best standalone feature and adds it up to

the feature set whereas SBS let goes it's features at each step which lacks least decreases the performance. The most accurate approach in achieving optimal feature selection is the exhaustive search. The second method evaluation method is again further divided into two sub parts: Filter and Wrapper. Wrapper approaches are less efficient than those of Filter approaches. Author describes method overview in section of paper and said that there are main three steps involve in making of pattern classification by means of supervised learning. Initially pertain a few preprocessing on the way to the given patterns resembling toward return to normal them to their respective size, variations of light, reduction of noise etc. Secondly, apply feature extraction for the representation of patterns by set of features. Train a classifier is the last step which involves the learning process and assigns these given inputs to correct category. Feature extraction is also very important and author in this paper used PCA for a feature extraction, images of Eigenspace representations uses PCA. Author takes about 10 images of vehicle detection and 10 images for face detection and applies the principal components in which eigenvectors corresponds to distribution of the training images of the largest eigenvalues. Images are represented as combinations of eigenvectors allowed by projection coefficients. From the results of above data sets of vehicle detection and face detection, some of eigenvectors encoded more lighting information than others, some encoded local information in given data sets and similar is the case with face detection. One of Eigenvector in both cases encoded mostly noise. Author in this paper used GAs for feature selection to search the eigenvectors space. Subset of different eigenvectors has been represented as each chromosome. Better performance can be achieved by the use of fewer features which is the main goal of subset of feature selection. Initial population is randomly generated; selection mechanism was obtained by cross generation or crossover. Crossover has three types: point one, point two plus uniform crossover. Probability or possibility of crossover and mutation was 0.66 and 0.04 respectively in all of author's experiment used. Learning machines are implemented by Support vector machines SVMs used by different kernels whereas Gaussian kernel communicate in the direction of a neural network of radial basis function (RBF), therefore Gaussian kernel be used within this paper and outperformed other kernels with respect to author's given application. In vehicle detection, author considered the difficulty of rear-view from its gray-scale images. Data set images used in this paper for experiment purpose were taken in fall and spring of 2001 in Michigan, USA. Randomly slated teaching dataset interested in three periods on behalf of teaching through maintaining 80% of vehicle subimages along with 80% of non-vehicle whereas the remaining data of 20% was used feature selection for validation. From the experimental result SVMs were tested with top eigenvectors, while performing several experiments the average inaccuracy rates obtain were 18.21%, 10.89%, 10.24%, and 10.80%, correspondingly. Now using eigenvector selection, the average error rate obtained were 6.49% of SVM in GAs case, and average error rate were 9.07% in the case of SFBS. In face detection, teaching set consists of 616 faces and 616 nonfaces subimages physically extracted since a gender dataset in addition to the CMU of detection of face dataset. Randomly slatted instruction dataset interested in three periods of time for teaching through maintaining 84% of face subimages as well as non-face whereas the remaining data of 16% was used feature selection for validation. From the experimental result SVMs were tested with top eigenvectors in vehicle detection, while performing several experiments average rates of error obtained were 12.31%, 11.57%, 13.81% along with 14.93%, respectively. Now using eigenvector selection, the average error rate obtained were 8.21% of SVM in GAs case, and average error rate were 10.45% in the case of SFBS.

Author in this paper [11] describes the feature selection process and that they proposed an efficient and simple solution regarding feature selection, basic principles of PCA are being

used from the data sharing. This proposed selection is much easier and faster in speed then other methods which also do use PCA directly. PCA used for Feature Extraction is summarized in mathematical equation which is described in this paper with formulas and equations. Images are always measured in x-axis and y-axis in vector for matrix. Using training sets of face images, an average is obtained. The difference is calculated between average face and each face. Corresponding eigenvalues matrix is estimated with using method Singular Value Decomposition (SVD) and eigenvectors smaller in numbers are kept. Subtracting it means in eigenspace is new face image reconstructed and original face of new representation is considered. The main focusing point of this paper is Feature Selection in which author proposed components of principle based on PCA in which corresponding axes selects the space of projection in original space. The target here in this paper is to discover theory axis xj adjoining toward ei in original space; the result of jth is selected; now the principle components of data division are e1 and e2 sort toward their subsequent eigen values. Beginning with e1 and x1 adjoining toward e1 is originate through apply a planned technique of feature-selection, thus w1 is the first feature and is selected. Regarding experimental results for training data, author used 7,000 face and also 7,000 non-face samples using proposed feature-selection method for detection of face. In Pixel-Based Features experiment, SVM classifiers were trained using LibSVM on selected feature subsets of RBF kernel. Now author compared the SVM classifiers with their proposed method and feature extraction based on PCA for subset features selection while in this process top-100 and top-200 eigenvectors were used plus result of author's proposed method was better than SVM classifier, 86 patterns per second (PPS) for proposed method and 80 patterns per second (PPS) on other method. Author's method is better in facial features of face and particular clearer such mouth, nose eyes *etc*. Author also used features of Haar wavelet based AdaBoost for face detections as it evaluates quickly, rich in features. 200 features of AdaBoost are selected then first-50 features added in process of training are selected and other first-50 features from author's proposed method are used, the performance of authors proposed method is better in terms of speed. Conclusion of this paper is that author proposed an efficient method closest to principle components computed by PCA and has been efficient in all experiments and results in which it runs faster used for face detection.

In this paper [12] Gender Classification method is studied. Author in this paper proposed a new method of gender classification with few defined steps, the given input data is FERET and SUMS face databases. First of all author applied histogram equalization to test the images, which minimized the illumination an acts as a result. Second, Viola and Jones technique used to detect the face section. Then features based on geometric have been extracted by LBP and features based on appearance have been extracted by DCT. Extracted features geometric and appearance use BA and PSO to be optimized, then to achieve high accuracy rate those optimized features are passed to optimized classifier based on ensemble. Author described the usage of Viola and Jones technique for face detection as it calculates fast, starts extracting faces from top most left and ends in bottom right corner in a matrix format of an image. Furthermore three modules used in face detection, first to make image very fast, second using learning algorithm adaboost for feature selection and last one is of elimination of background regions of image using AdaBoost classifiers which also speeds up the detection process. They selected 16 coefficients from an image, selected each block coefficient of DCT, number of coefficients increases so do size of feature vector. Author initially used active shape model in five components of facial *i.e.*, eyebrows, eyes, mouth, nose and chin for locating purpose, cropped these features and then from cropped components they applied LBP to extract these features. Particle Swarm Optimization (PSO) and Bee algorithm (BA) optimizes local and global features but here author used/implemented binary

PSO (BPSO) for evaluation of optimized features. Representation of each particle for binary value, author used 100 particles and each particle is associated as a feature vector. During each iteration, genders are eliminated/discarded which carried less discriminative information by using BPSO the rest are optimized features required. With respect to feature vectors repetition of step is required. Author used BA to optimized features and performed few steps in using this algorithm. First of all they selected different feature vectors, secondly Each set of feature computes accuracy fitness, now if given criteria doesn't not meet then fitness is performed again and new features are selected, then only those feature victors are picked which fulfill given criteria for each patch and process is terminated if a criteria is met accordingly. Classifier ensemble is derived from output of different classifiers, author used three different trained classifiers and evaluated on given optimized features which are as back propagation NNs (BPNNs), KNN and SVM and optimized through GA. High accuracy classifiers are more efficient than those of less used for dataset and are weight is measured in a way that the classifier contribution is directly proportional to its accuracy in ending decision. GA has been used to optimize ensemble classifiers, weight of classifier normalized by GA. Applied mutation and crossover to obtain weight of ensemble classifiers, GA is terminated when fitness solution is obtained from a generation. Author used MatLab in their experiments, FERET and SUMS were used as databases, 200 female and 200 male images were selected from these databases. Training-to-testing 1:9 and 9:1 ratio used, using Viola and Jones algorithm to detect face facials, to normalize HE was performed., face images further divided into blocks of 8 using DCT in zigzag order then passed selected coefficients to BA and PSO for optimized features. Now performed GA to obtain optimize the weights of the ensemble classier, average coverage of GA is 53. The results of BA on optimized feature set to 90% accuracy rate after experiments but with KNN classifier results got 79.8% and then author performed this same give training and test data set on proposed GA, 95.7% accuracy rate received.

Gender Classification technique is studied in this paper. Author in this paper [13] proposed a new method of gender classification. Three modules such as feature selection/extraction, classification and face detection of gender classification method are described and discussed. Face detection technique used in this paper is Viola and Jones to extract face from the image while passed through a detector; contrast of an image is stretch with the help of histogram equalization. Image of dimensions 32x32 is then resized to 8x8 of each block, sorted in zigzag and passed to KNN classifier. Author described the usage of Viola and Jones technique for face detection as it calculates fast, starts extracting faces from top most left and ends in bottom right corner in a matrix format of an image. Author further more described three technique modules used in face detection, first one makes computational feature of an image very fast, second one for feature selection using learning algorithm adaboost and third one for background elimination of regions of an image using AdaBoost classifiers which also speeds up the detection process. Regarding Feature Extraction, author used discrete cosine transform (DCT) which is efficient in image compression applications, face recognition application and also for dimension reduction. Those coefficients are fetched first which have high variance while DCT coefficients were sorted in zigzag order. Scan performed in zigzag order which is from left most towards right most in addition to top most towards bottom taking place DCT coefficient matrix because coefficients with high importance are found in top most left angle of corner of the building block. Then preferred 16 coefficients from image then first of each block coefficient of DCT is selected, if number of coefficients increased so does the size of feature vector. 32, 48, 64,128 and 256 size was created in this same manner like 16 feature vectors. Regarding classifier, KNN has been assigned its K closest neighbors where the parameter K is number of neighbors of a class and to find distance Euclidean is

used, others can also be used for measuring of distance. Author performed three types of experiments, first training set data v/s set size test, second DCT coefficients effect of normalization and third selected number of features. 50 training data to testing ratio is suitable which gives 99.3% of classification accuracy with KNN classifier. DCT and LDA with Viola and Jones technique increased the results whereas performance of LDA was not much obtained with DCT.

Author in this paper [14] proposed a method for feature selection named class-dependent because selection along with extraction are the most important parts. The proposed method works after the feature has been extracted, method is obtained from the principal components. Author used ORL face database in their experiments as experiments have high dimension, also used SVM to evaluate each feature subset as every time 3 features are going be added into feature subset of previous ones then highest classification accuracy feature subset are picked as superior one of existing class. Author converted few two-class problems from multi-class problem based on "one-against-all" strategy and performed this strategy on face image and transformed into 2-class classification issues where apiece integrated two classes, first because of original class along with second having all other classes. Class separability measure (CSM) is adopted for each 2-class problem, evaluated features to rank each class. Classification task becomes easier if the distance between different classes is greater. Let suppose for problem 1, in classifying class 1 the importance of features measured by its ranking from the other classes and as result of that the aspect significance grade directory is exclusively assign to a class 1. Procedure can be same on behalf of class 2nd, class 3rd and then class C. A classifier SVM, used to determine subset features of the majority contributive single in favor of apiece class. They used mask of feature for expression of each feature state, '0' and '1' are two elements of feature mask in which absence of particular feature is represented by '0' and presence by '1'. SVM with RBF kernel are adopted for the classification but class-dependent facial appearance straightforwardly cannot input in the direction of original SVM so class-dependent SVM classifier is used. Regarding experiment, author choose data from Cambridge ORL face database for all of their experiments which consists of 40 subjects i.e. classes or faces along with apiece has to some extent 10 unlike images of face. There are 400 face images in total and each of them are within a gray scale of 112 * 92 pixels in dimensions, reshaped matrix of image into column vector of original image. Subtract mean and divide by variance of each image then in case of energy all images were distributed equally.

Author in this paper [15] proposed new method fuzzy entropy procedures meant for taking care of categorization issues of feature subset selection. Author initially gives a brief detail on existing measures of entropies like Shannon's entropy which more commonly used, Zadeh brief on fuzzy entropy, Luca and Termini fuzzy entropy measure etc. After giving a general brief on existing entropy measure author now discusses his proposed a measure of fuzzy entropy of its set and calculating all equations, measures and entropies from given samples, results from these indicate with the intention of Shannon's entropies of both I1 as well as I2 interval are identical and can't notable the entropies of both intervals but in author's proposed measures of fuzzy entropy indicates that interval I2 is relatively additional unclear than to facilitate of interval I1. An algorithm is proposed to build a function of the relationship of every aspect of fuzzy set anywhere fuzzy set of membership function represents each linguistic. Features of two categories, one nominal and other numeric and each contain their association function of fuzzy sets respectively. Discredited numeric feature into finite fuzzy sets, classification result are affected by number of fuzzy sets and is important procedure. Unsupervised learning techniques are used is good method to discrete numeric feature. Author applied k-means clustering algorithm. Feature of fuzzy entropy decreases when

increase the number of clusters, however overfitting problem and reduction of accuracy rate when a new instance classify may occur regarding too many cluster. To overcome this problem, threshold value Tc is used to stop the increase of number of clusters and by subtracting feature fuzzy entropy the decrease rate is obtained and calculated by feature values clustering into k clusters. Algorithm of membership functions of numeric fuzzy feature sets follow few steps, first set cluster to 2 of number k then step 2 uses k-mean clustering algorithm, step 3 to built membership functions of fuzzy set, step 4 calculate feature fuzzy entropy and at the end decrease the rate of feature fuzzy entropy f is superior than threshold Tc value followed by accede to k = k+1 along with go to Step 2 Other than that let k = k−1 and terminate. Author used Lotus Notes of IBM with its Version 4.6 on system Pentium 4, four types of classifiers such as SMO, LMT, C4.5 and Naive Bayes in addition to made two different experiments UCI data sets of 4 different types used in their first experiment such as breast cancer data set, Pima diabetes, MPG plus the Iris data set data set whereas second experiments used 8 different types UCI data sets for the purpose of comparing classification of average accuracy rate of those chosen features. Regarding experiment 1, author applied 4 UCI data sets and consists of two main steps, first corresponds to each fuzzy set of membership function of each feature whereas second steps based on fuzzy entropy measure to select feature subsets. After then used four types of classifiers as mentioned above to evaluate chosen feature subsets presentation through unlike technique and various types of classifiers in WEKA and applied on 10-fold cross-validation as well as split data set of apiece toward 10 subsets equally sized and 10 times executed, 10 subsets selected each time as testing data and classifiers are trained by 9 subsets that ware remaining to obtain the accuracy rate of classification to that of each features subset selected. Now by executing this process 10 times and end result in case of average accuracy rate of classification is higher than the others. Now author discussed about experiment 2 in which data sets used are the Correlated, M of N-3-7-10, Cleve, Crx, pima Diabetes, Monk-1, Monk-2 furthermore Monk-3 data set and applied all steps like in above experiment 1 and got the end results in which proposed average accuracy rate of classification is higher than ones of which preferred through Dong-and-Kothari's technique.

Author in this paper [16] used method an iris recognition method for most favorable features subset selection base on top of genetic algorithms (GA). Types of regions those are not relevant like pupil, sclera, eyelids and some other have been enclosed Iris. Author uses pupil and collarette borders of Iris not its entire regions, regions which are noisy also do need to be eliminated to get better of the performance. Rubber sheet model has been used to stabilize the isolated collarette region, in view of the fact that iris image with the purpose of is normalize have small contrast for which histogram equalization procedure is performed in the direction of conquer this small contrast issue. Pattern that is normalized is convolved by way of 1-D Log-Gabor filters. Author in this paper adopted GA to come across the most favorable subset of revealing features by combining several feature selection criteria; author then used four algorithms of feature selection rather than having each and every one feature set extracted from its inventive iris features sequence. On the way to extract features and breed ranking of those features, each algorithm is applied, select top number of features and provide these feature to feature pool. Author describes four feature selection algorithms in minority expressions. Based on entropy process, lesser entropy intended for orderly and superior for unmanageable configurations so entropy is reduced when the irrelevant feature is eliminated and features are rank in downward order of entropies after eliminating one at a time. Therefore this approach is used for unsupervised data as class in sequence is not required. Now in support of selecting individuals genes which are appropriate for cancer categorization difficulty SVM-RFE has been used, to find the extract features progression the top ranked iris

features and to exclude those that are worst feature one at a time and approach is set to backward selection. Evaluation of nonparametric feature subset choice is theoretical based on direct method in k-nearest neighbor rule (K-NNR) characteristic selection and denoted by 'leave-one-out (LOO)'. Process follows in few steps, first apply K-NN as classifiers, second for recognition rate estimation LOO test is used, third first selected feature has the highest LOO of all, then to select one from all unselect to obtain maximum gratitude rate, replicate this procedure in anticipation of a suitable recognition rate is obtained at level. SVM classifier is used in this paper as induction algorithm to split the cases of artificial accept along with refuse, as probabilistically used Roulette wheel selection, selected relative to its have possession of robustness in addition to inversely to robustness of the others in the in progress inhabitants challenging hypothesis. Single point crossover, possibility of each one entity in addition to to mutate in used. In every stage of mutation, m bits of numbers randomly selected to be flipped. In the direction of precisely classify the iris pattern in this paper Multi-class Gaussian mixture model (GMM) is used. The main goal here is to guess the parameters of GMM of a given training sample which in then matches the best training feature vectors distribution. The purpose is to achieve the mock-up parameters of the GMM to facilitate make best use of the probability provide in the agreed teaching data. In this case of iris recognition, purpose is the obtain person model that maximum a probability of posteriori in given observation sequence, GMM's represent groups of subjects. Two iris data sets named as ICE (Iris Challenge Evaluation) and WVU (West Virginia University) for author experiment purpose. Left iris images of 1528 corresponding to that of 120 subjects in total used in this experiments as author only considered left iris images of ICE whereas WVU iris dataset has 1852 in total from 380 kind of different people of iris images. Success rate has been evaluated on both of these given experimental datasets and obtained 98.80% for ICE and 97.95% for WVU data sets in which it is observed that recognition accuracy is obtained to some instance and drop in matching error from 3.61% to 2.48% for ICE data set, If values of pixel is increased in WVU then highest matching accuracy is obtained which is 96.79%. GA for the ICE dataset gets the accuracy of 97.80% highest at 90 level of generation with features subset reduced of 520 and 95.90% is obtained for WVU data set with features subset reduced to 680. On ICE data set, 97.90% is obtained when number of components of Gaussian is 32 whereas highest recognition accuracy for WVU data set is 96.30% of 27 mixture components.

4. Critical Evaluation

I

Ref	Techniques Used	Strength	Weakness
[3]	Ant colony Optimization	High accuracy rate by utilizing the smaller feature sets	It does optimize feature selection but it also has limitations as behavior is stagnative, huge computational time and premature convergence
[4]	Convex relaxation approximation, backward greedy algorithm	Minimize the distance to a subspace	Dimensional primal problem appears by using convex relaxation
[5]	Mutual information, feature fusion	Robustness in noisy environments and data transformations	It is computationally too intensive and complex

Ref	*Techniques Used*	*Strength*	*Weakness*
[6]	Local Binary Pattern (LBP)	Increases accuracy facial recognition	Speed of recognition reduces on large-scale face database
[7]	Dominant local binary pattern (DLBP)	Robustness in noisy image	Long or large pixel range not effected coverage of neighborhood circular
[8]	Evolutionary computation techniques	Improves the performance of classification used	Unsuccessful offspring produces when mating chromosome between different niches
[9]	Accelerated Binary Particle Swarm Optimization (ABPSO)	Improves recognition rate and reduces feature subset	Takes lot of time to optimize
[10]	Genetic algorithms (GAs)	Improves facial detection rates	Genetic algorithm has no absolute assurance to find a global optimum.
[11]	Principle component analysis (PCA)	High prediction accuracy rate by utilizing feature sets	PCA cannot deal with data that have unknown pixels
[12]	Particle swarm optimization (PSO)	Minimization of computation time and classification error rate	Takes lot of time to optimize.
[13]	K-nearest neighbor classifier (KNN)	Implementation is simple and has Robustness to the search space	Its testing is expensive, very sensitive attributes with noisy/irrelevant and to unbalanced datasets
[14]	Class-dependent feature selection	Improves accuracy and reduction of data dimensionality	Class-dependent feature selection method is likely to be more computationally expensive
[15]	Fuzzy entropy measures Handling classification problems	Higher average classification accuracy rates	Its complex and time consumption
[16]	Genetic algorithms (GA)	Improves facial detection rates	Genetic algorithm have no absolute assurance to find a global optimum

5. Conclusion

In this paper, the main focus is on the reduction of dimension on the basis of feature selection of facial appearance (local and global). Techniques and methods proposed on this paper to optimized using PSO and BA the features that are selected from an image and classifiers like ensembles to get high accuracy rate. Although there is still work left in dimensionality reduction techniques and this can be the next step for future work in field of feature selections.

References

[1]   K. Rajpoot and N. Rajpoot, "SVM Optimization for Hyperspectral Colon Tissue Cell Classification", LNCS3217. Springer-Verlag, **(2004)**, pp. 829-837.

[2]   S. Goonatilake and P. Treleaven, "Intelligent Systems for Finance and Business", John Wiley and Sons, **(1995)**.

[3]   L. Chen, B. Chen and Y. Chen, "Image feature selection based on ant colony optimization", Advances in Artificial Intelligence, pp. 580-589, **(2011)**.

[4]   G. Roig, X. Boix and F. De la Torre, "Optimal feature selection for subspace image matching", Computer Vision Workshops (ICCV Workshops), **(2009)**, pp. 200-205.

[5]   C. Perez, J. Tapia, P. Estevez and C. Held, "Gender classification from face images using mutual information and feature fusion", International Journal of Optomechatronics, vol. 6, **(2012)**, pp. 92-119.

[6]   J. Shelton, G. Dozier, K. Bryant, J. Adams, K. Popplewell, T. Abegaz and K. Ricanek, "Genetic based LBP feature extraction and selection for facial recognition", Proceedings of the 49th Annual Southeast Regional Conference, **(2011)**, pp. 197-200.

[7]   A. Bamini and T. Kavitha, "Dominant Local Binary Pattern Based Face Feature Selection and Detection", International Journal of Engineering and Technology, vol. 2, **(2010)**, pp. 77-80.

[8]   L. Vignolo, D. Milone, C. Behaine and J. Scharcanski, "An evolutionary wrapper for feature selection in face recognition applications", Systems, Man, and Cybernetics (SMC), 2012 IEEE International Conference, **(2012)**, pp. 1286-1290.

[9]   M. U. Aneesh, A. A. Masand and K. Manikantan, "Optimal Feature Selection based on Image Pre-processing using Accelerated Binary Particle Swarm Optimization for Enhanced Face Recognition", Procedia Engineering, **v**ol. 30, **(2012)**, pp. 750-758.

[10] Z. Sun, G. Bebis and R. Miller, "Object detection using feature subset selection", Pattern recognition, vol. 37, **(2004)**, pp. 2165-2176

[11] D. D. Le and S. I. Satoh, "An efficient feature selection method for object detection", Pattern Recognition and Data Mining, **(2005)**, pp. 461-468.

[12] S. A. Khan, M. Nazir and N. Riaz, "Optimized features selection for gender classification using optimization algorithms", Turk J Elec Eng & Comp Sci., vol. 21, **(2013)**, pp. 1479-1494.

[13] M. Nazir, M. Ishtiaq, A. Batool, M. A. Jaffar and A. M. Mirza, "Feature selection for efficient gender classification", Proceedings of the WSEAS international conference, Wisconsin, **(2010)**, pp. 70-75.

[14] Z. Nina and L. Wang, "Class-Dependent Feature Selection for Face Recognition", Advances in Neuro-Information Processing, **(2009)**, pp. 551-558.

[15] J. D. Shie and S. M. Chen, "Feature subset selection based on fuzzy entropy measures for handling classification problems", Applied Intelligence, vol. 28, **(2008)**, pp. 69-82.

[16] K. Roy and P. Bhattacharya, "Optimal features subset selection using genetic algorithms for iris recognition", In Image Analysis and Recognition, **(2008)**, pp. 894-904.

Authors

Saqib Hayat is a postgraduate at Department of Computer Science, Shaheed Zulfikar Ali Bhutto Institute of Science and Technology, Islamabad, Pakistan. His research interests include feature selection, extraction and dimensional reduction.

Abdul Basit Siddiqui obtained Ph.D degree in Computer Science from the FAST, Islamabad, Pakistan. Currently he is working in Department of software engineering, Foundation University, Rawalpind, Pakistan as Assistant Professor. His research interests are in the fields of Digital Image processing.

Sajid Ali Khan is a Lecturer in Department of Software Engineering at Foundation University, Rawalpindi, Pakistan. His research area is Pattern Recognition, Face Analysis and Digital Image Processing. Sajid Ali has also been supervising MS students in Independent Study research. He received his BS(CS) degree from PMAS University of Arid Agriculture, Rawalpindi, Pakistan. He completed his MS (CS) degree from SZABIST University Islamabad, Pakistan.