MULTI-LEVEL DIMENSIONALITY REDUCTION METHODS USING FEATURE SELECTION AND FEATURE EXTRACTION

Veerabhadrappa[1,2] Lalitha Rangarajan[2]

[1] Department of Computer Science, University College, Mangalore – 575 001, India
Email: virabadrappa@gmail.com
[2] Department of Studies in Computer Science, University of Mysore, Mysore- 570 006, India
Email: lali85arun@yahoo.co.in

ABSTRACT: *This paper presents a novel feature selection method called Feature Quality (FQ) measure based on the quality measure of individual features. We also propose novel combinations of two level and multi level dimensionality reduction methods which are based on the feature selection like mutual correlation, FQ measure and feature extraction methods like PCA(Principal Component Analysis)/LPP(Locality Preserving Projection). These multi level dimensionality reduction methods integrate feature selection and feature extraction methods to improve the classification performance. In the proposed combined approach, in level 1 of dimensionality reduction, feature are selected based on mutual correlation and in level 2 features are selected from the reduced set obtained from level 1 based on the quality measure of the individual features and vice versa. In another proposed combined approach, feature extraction methods like PCA and LPP are applied on the reduced feature set obtained in level 1 or /and level 2. To evaluate the performance of the proposed methods several experiments are conducted on standard datasets and the results obtained show superiority of the proposed methods over single level dimensionality reduction techniques.*

KEYWORDS: *Mutual Correlation, dimensionality reduction, feature quality measure, feature selection/extraction.*

1. INTRODUCTION

In recent years, improvement in data acquisition capacity, lower cost of data storage and development of database and data warehousing technology have led to the emergence of high dimensional dataset. The increase of data size in terms of number of instances and number of features becomes a great challenge for the feature selection algorithms. Many of these features are irrelevant and redundant which increase the search space size resulting in difficulty to process the data further. This curse of dimensionality is a major obstacle in machine learning and data mining applications. Large amount of storage space and computational time is required in handling of high dimensional data and hence we need to reduce the dimension. Dimensionality reduction is an active research area in the field of pattern recognition, machine learning, data mining and statistics. The purpose of dimensionality reduction is to improve the classification performance through the removal of redundant or irrelevant features. Dimensionality reduction can be achieved in two different ways namely feature selection and feature transformation.

Feature extraction/transformation is a process through which a new set of features is created. The feature transformation may be a linear or nonlinear combination of original features. Feature selection is a process, through which no new set of features will be generated, but only a subset of original features is selected and feature space is reduced. The choice between feature selection and feature extraction depends on the application domain and the specific data set. Variety of feature selection methods have been developed in the literature, which can be classified into three main categories: filter, wrapper and hybrid approaches. Filter methods apply an independent test without involving any learning algorithm, while wrapper methods require a predetermined learning algorithm for feature subset evaluation. Filter and wrapper methods have their drawbacks and are complementary to each other. The filter approaches have low computational cost with low reliability in classification while wrapper methods tend to have superior classification accuracy but require great computational effort. Filter approaches select features using characteristics of individual features. Wrapper approaches use a specific machine learning algorithm/ classifiers such as decision tree or SVM (Support Vector Machine) and utilize the corresponding classification performance to select features. Advantages of the filter-based techniques are that they can easily scale up to high-dimensional datasets and that they are computationally fast and independent of the learning algorithm. For application of large datasets, filter based approaches have proven to be more practical than wrapper approach because of their speed. A common disadvantage, however, is that the interaction with the classifier and the dependence among features are ignored, which leads to varied classification performance when the selected features are applied to different classification algorithms. On the other hand, the advantage of wrapper approaches is that they have a high probability of producing classes with better classification performances than the filter approaches as they take into account the feature dependencies and their collective contribution to model generation. A common drawback, however, is that the wrapper approaches have a higher risk of over-fitting and can be very computationally intensive when processing a large number of features. Hybrid approach is a recent technique which exploit the advantages of both filter and wrapper approach. A hybrid approach employs both an independent test and performance evaluation function of the feature subset.

Instead of using either feature selection or feature extraction technique to reduce the dimension, the combinations of feature selection and feature extraction methods can be applied as multi level approach, i.e., apply one approach like feature selection on the original feature set to reduce its dimension and then apply another approach (again either feature selection or feature extraction) on this reduced feature set to further reduce its dimension. These cascading approaches can be categorized as follows:

a) Feature selection algorithm followed by another feature selection algorithm.
b) Feature selection algorithm followed by feature extraction algorithm.
c) Feature extraction algorithm followed by another feature extraction algorithm.

In this paper, we propose the following dimensionality reduction methods:

I. Single level dimensionality reduction method
    a. A novel feature selection method called Feature Quality (FQ) measure.
II. Two level dimensionality reduction methods
    a. Feature selection using mutual correlation followed by applying another feature selection method on the reduced feature set based on feature quality measure of individual features.

b. Feature selection using feature quality measure of individual features followed by applying another feature selection method on the reduced feature set based on mutual correlation.
c. Feature selection using feature quality measure of individual features followed by applying feature extraction method like PCA/LPP on the reduced feature set.
d. Feature selection using mutual correlation followed by applying feature extraction method like PCA/LPP on the reduced feature set.

III. Multi level dimensionality reduction methods

a. Feature selection using mutual correlation and feature quality measure or feature quality measure and mutual correlation followed by applying feature extraction method like PCA/LPP on the reduced feature set.

The filter approaches based on mutual correlation and feature quality measure are very simple and have low computational complexity are applied to select a feature subset and then the feature extraction method PCA/LPP is applied to reduce the dimension further. PCA has proven record of high success in reducing dimensions. So in this method we have extracted features using PCA and again retaining only first few principal components. PCA is a global transformation and LPP is useful where local information is at most important. In our second method, features are extracted using LPP. It may be noted that feature extraction methods PCA/LPP are used on reduced set of features selected in the first step and hence the time required for transformation is far less than using PCA/LPP on the original set of features. Hence in this paper, we propose to combine the filter based feature selection method based on mutual correlation/feature quality measure with feature transformation methods PCA/LPP which is expected to perform better since is has the advantage of both feature selection and extraction.

This paper is organized into 5 sections. A brief review of the related work is given in section 2. The existing feature selection method based mutual correlation is presented in section 3. The proposed methods are presented in section 4. Experimental results are presented in section 5 followed by conclusion in section 6.

2. RELATED WORK

Several feature selection techniques have been proposed in the literature and survey of feature selection algorithms may be found in Molina et al. [1] and Guyon and Elisseeff [2]. Many researchers are involved in the study of goodness of a feature subset in determining an optimal one [1-6, 8, 9, 14, 27]. The wrapper model uses the predictive accuracy of a predetermined learning algorithm to determine the goodness of the selected subset. A serious drawback about this method is the higher computational cost [4]. The filter model selects features that are independent of any learning algorithm and it relies on various measures of the general characteristics of training data such as distance, information dependency and consistency [3]. In these methods, the relevance of each feature is evaluated individually and a score is given to each of them. The features are ranked by their scores and the ones with a score greater than a threshold are selected. According to the availability of class labels, there are feature selection methods for supervised learning [5, 6] as well as for unsupervised learning [7, 8]. Existing feature selection methods mainly exploit two approaches: individual feature evaluation and subset evaluation [2]. Individual evaluation methods rank features according to their importance in differentiating instances of different classes and can only remove irrelevant features as redundant features with similar rankings. Methods of subset evaluation search for a minimum subset of features that satisfies some goodness of measure and can remove irrelevant features as well as redundant ones [9].

Frequently used filter methods include *t*-test [10], chi-square test [11], mutual information [12], Pearson correlation coefficients [13] and Relief [29]. The RELIEF, one of the most used filter methods was introduced by Kira and Rendell [29] for a two-class problem, and extended later to the multiclass problem by Kononenko [30]. In the RELIEF, the relevance weight of each feature is estimated according to its ability to distinguish instances belonging to different classes. Thus, a good feature must assume similar values for instances in the same class and different values for instances in other classes. The relevance weights are set to be zero for each feature and then are estimated iteratively. In order to do that, an instance is chosen randomly from the training dataset. Then, the RELIEF searches for two closest neighbors to such instance, one in the same class, called the *Nearest Hit* and the other in the opposite class called the *Nearest Miss*. The relevance weight of each feature is modified according to the distance of the instance to its *Nearest Hit* and *Nearest Miss*. The relevance weights continue to be updated by repeating the above process using a random sample of *n* instances drawn from the training dataset. Filter methods are fast but lack of robustness against interactions among features and feature redundancy. In addition, it is not clear how to determine the cut-off point for rankings to select only truly important features and exclude noise.

In the wrapper approach, features selection is "wrapped" in a learning algorithm. The learning algorithm is applied to subsets of features and tested on a hold-out set and prediction accuracy is used to determine the feature set quality. Generally, wrapper methods are more effective than filter methods. Since exhaustive search is not computationally feasible, wrapper methods must employ a search algorithm to search for an optimal subset of features. SBS (Sequential Backward Selection) starts with the set of all features and progressively eliminates the least promising ones. SBS stops if the performance of learning algorithms drops below a given threshold due to removal of any remaining features. SBS relies heavily on the monotonic assumption [14], i.e., prediction accuracy never decreases as the number of features increases. In reality, the predictive ability of a learning algorithm may decrease as the feature subspace dimensionality increases after a maximum point due to a decreasing number of samples for each feature combination. When faced with high-dimensional data, SBS often finds difficulties in identifying the separate effect of each explanatory variable on the target variable. Because of this, good predictors can be removed early on in the algorithm (in SBS, once a feature is removed, it is removed permanently), where as SFS (Sequential Forward Selection) starts with an empty set of features and iteratively selects one feature at a time starting with the most promising feature until no improvement in classification accuracy can be achieved. In SFS, once a feature is added, it is never removed. SBS is robust to feature interaction problems but sensitive to multicollinearity. On the other hand, SFS is robust to multicollinearity problems but sensitive to feature interaction. The problem with SFS and SBS is their single track search. Hence, Pudil et al. [15] suggest floating search methods (SFFS, SFBS) that performs greedy search with provision for backtracking. However recent empirical studies demonstrate that sequential floating forward selection (SFFS) is not superior to SFS [16] and sequential floating backward selection (SFBS) is not feasible for feature sets of more than about 100 features [17]. Stochastic algorithms have been developed for solving large scale combinatorial problems such as Ant Colony Optimization (ACO), Genetic Algorithm (GA), Particle Swarm Optimization (PSO) and Simulated Annealing (SA) are at the forefront of research in feature subset selection [14, 18–20]. These algorithms efficiently capture feature redundancy and interaction and do not require the restrictive monotonic assumption. However, these algorithms are computationally expensive. Recently, several authors proposed hybrid approaches taking advantages of both filter and wrapper methods. Examples of hybrid algorithms include *t*-statistics and a GA [21], a correlation based feature selection algorithm and a genetic algorithm [22], principal component analysis and an ACO algorithm [23], chi-square approach and a multi-objective optimization algorithm [24], mutual information and a GA [25,26]. The idea behind the hybrid method is that filter methods are first applied to select a feature pool and then the wrapper method is applied to find the optimal subset of features from the selected feature pool. This makes feature selection faster since the filter method rapidly reduces the effective number of features under consideration. Advocates of hybrid methods argue that the risk of eliminating good predictors by filter methods is minimized if the filter cut-off point for a ranked list of features is set low. However, hybrids of filter and wrapper methods may suffer in terms of accuracy because a relevant feature in isolation may appear no more discriminating than an irrelevant one in the presence of feature interactions [27].

3. EXISTING METHOD

3.1 Feature Selection based on Mutual Correlation

Correlation is a well known similarity measure between two random variables. If two random variables are linearly dependent, then their correlation coefficient is close to ±1. If the variables are uncorrelated the correlation coefficient is 0. The correlation coefficient is invariant to scaling and translation. Hence two features with different variances may have same value of this measure. The p-dimensional feature vectors of N number of instances is given by

$$X_i = \left[{}^{i}x_1, \dots {}^{i}x_p \right] \quad \text{i= 1,...,N}$$

The mutual correlation [28] for a feature pair x_i and x_j is defined as

$$r_{x_i,x_j} = \frac{\sum_k {}^{k}x_i \, {}^{k}x_j - N\bar{x}_i \bar{x}_j}{\sqrt{\left(\sum_k {}^{k}x_i^2 - N\bar{x}_i^2\right)\left(\sum_k {}^{k}x_j^2 - N\bar{x}_j^2\right)}} \tag{1}$$

where k = 1,...N

If two features x_i and x_j are independent then they are also uncorrelated, i.e. $r_{x_i,x_j}=0$. Let us evaluate all mutual correlations for all feature pairs and compute the average absolute mutual correlation of a feature over δ features

$$r_{j,\delta} = \frac{1}{\delta} \sum_{i=1, i \neq j}^{\delta} \left| r_{x_i,x_j} \right| \tag{2}$$

The feature which has the largest average mutual correlation

$$\alpha = \arg\max_j \; r_{j,\delta} \tag{3}$$

will be removed during each iteration of the feature selection algorithm. When feature x_α is removed from the feature set, it is also discarded from the remaining average correlation, i.e.

$$r_{j,\delta-1} = \frac{\delta r_{j,\delta} - \left| r_{x_\alpha, x_j} \right|}{\delta - 1} \tag{4}$$

Algorithm 1: Feature selection based on mutual correlation

Input: Original feature set X of size N x p

Output: reduced feature set of size N x D (D<<p)

Method:

1. Initialize $\delta = p$
2. Discard features x_α for α determined by equation (3).
3. Decrement $\delta = \delta - 1$, if $\delta < D$ return the resulting D dimensional feature set and stop otherwise.
4. Recalculate the average correlations by using equation (4).
5. Go to step 2.

This algorithm produces the D-dimensional feature subset from the original feature set based mutual correlation concept.

4. PROPOSED METHODS

4.1 Single level Dimensionality Reduction methods

4.1.1 Feature selection based on feature quality measure (FQ measure)

Motivated by the work of Luis Daza and Edger Acuna[31] who introduced a measure to evaluate the quality of an instance, we propose to the compute feature quality measure of individual features. For the ijth feature (jth feature of the ith sample) of the dataset of size N x p, where N is number of samples and p is the number of features, the feature quality measure is given by $Q_{ij} = (r_{ij} - d_{ij})/\max(d_{ij}, r_{ij})$, i=1,….N, j = 1,…p, where d_{ij} is the distance of the ijth feature to the centroid of its class, and r_{ij} is the minimum distance of the ijth feature to the centroid of the classes where it does not belong to. It is evident that $-1 \le Q \le 1$. A feature with a Q value near to 1 has a good quality. Negative Q indicates noise in the feature.

Algorithm 2: Feature selection based on feature quality measure

Input: Original Dataset X of size N x p, class label

Output: Reduced feature set of size N x k (k <<p)

Method:

1. Compute the centroid of each class.
2. Compute the d_{ij} as the distance between centroid of each class and the ijth feature of that class.
3. Compute the r_{ij} as the minimum distance between the ijth feature and centroid of the class where ijth feature does not belong to.
4. Compute $Q_{ij} = \dfrac{r_{ij} - d_{ij}}{\max(r_{ij}, d_{ij})}$ as measure of feature quality of individual features.
   where i = 1,…,N , j=1,…p.
5. Find the average feature quality measure as

$$\overline{Q_j} = mean(Q_{ij})$$ where i= 1,…N, j = 1,…p.

6. Find the weight of each feature using

$$W_j = e^{\overline{Q_j} - 1}$$ where j = 1,…,p.

7. The p features are ranked in descending order of their weight.
8. Select the first k features as the reduced feature set.

4.2 Two level Dimensionality Reduction methods

4.2.1 Feature selection based on mutual correlation followed by Feature selection based on feature quality measure

The feature selection methods explained in section 3.1 and 4.1 are combined together to get the optimal reduced feature set. In the first level, Algorithm 1 is applied on p original features to obtain the reduced D features using the concept of mutual correlation. In the next level algorithm 2 is applied on the reduced feature set using the concept of feature quality measure to get the further reduced feature set.

Algorithm 3: Dimensionality reduction using feature selection based on mutual correlation + feature selection based on feature quality measure

Input: Original Dataset X of size N x p, class label
Output: Reduced feature set of size N x D (D <<p)
Method:
Level I : Apply Algorithm 1 on X to obtain the reduced set of size N x f.
Level II: Apply Algorithm 2 on the reduced feature set of size N x f obtained from level I to get the further reduced feature set of size N x D.

4.2.2 Feature selection based on feature quality measure followed by feature selection based on mutual correlation

In this case at the first level, Algorithm 2 is applied on p original features to obtain the reduced D features using the concept of feature quality measure. In the next level algorithm 1 is applied on the reduced using the concept of mutual correlation to get the further reduced feature set.

Algorithm 4: Dimensionality reduction using feature selection based on feature quality measure + feature selection based on Mutual correlation

Input: Original Dataset X of size N x p, class label
Output: Reduced feature set of size N x D (D <<p)
Method:
Level I : Apply Algorithm 2 on X to obtain the reduced set of size N x f.
Level II: Apply Algorithm 1 on the reduced feature set of size N x f obtained from level I to get the further reduced feature set of size N x D.

.

4.2.3 Feature selection based on mutual correlation followed by PCA or LPP

Here feature selection based on mutual correlation which is very simple and has low computational complexity is applied to select a feature subset followed by feature extraction method using PCA or LPP to reduce the dimension further. At level 1, Algorithm 1 is applied on p original features to obtain the reduced f features using the concept of mutual correlation. And in the level 2, the PCA or LPP is applied to get the further reduced feature set of size N x D.

Algorithm 5: Dimensionality reduction using mutual correlation followed by PCA or LPP.

Input: Original Dataset X of size N x p, class label

Output: New transformed feature set Y of dimension D (D<<f).

Method:

Level I: Apply Algorithm 1 on X to obtain the reduced set of size N x f.

Level II: Apply PCA or LPP on the reduced feature set of size N x f obtained from Level I.

i.e Apply PCA on $X=[x_1,\ldots,x_f]$ to obtain $Y(y_1,\ldots,y_D)=T(x_1,\ldots,x_f)$.

OR

Apply LPP on $X=[x_1,\ldots,x_f]$ to obtain $Y(y_1,\ldots,y_D)=T(x_1,\ldots,x_f)$

4.2.4 Feature selection based on feature quality measure followed by PCA or LPP.

Here feature selection based on feature quality measure which is also very simple and has low computational complexity is applied to select a feature subset followed by feature extraction method using PCA or LPP to reduce the dimension further. At level 1, Algorithm 2 is applied on p original features to obtain the reduced f features using the concept of feature quality measure. And in the level 2, the PCA or LPP is applied to get the further reduced feature set of size N x D.

Algorithm 6: Dimensionality reduction using feature quality measure followed by PCA or LPP.

Input: Original Dataset X of size N x p, class label

Output: New transformed feature set Y of dimension D (D<<f).

Method:

Level I: Apply Algorithm 2 on X to obtain the reduced set of size N x f.

Level II: Apply PCA or LPP on the reduced feature set of size N x f obtained from Level I.

i.e Apply PCA on $X=[x_1,\ldots,x_f]$ to obtain $Y(y_1,\ldots,y_D)=T(x_1,\ldots,x_f)$.

OR

Apply LPP on $X=[x_1,\ldots,x_f]$ to obtain $Y(y_1,\ldots,y_D)=T(x_1,\ldots,x_f)$

4.3 Multi level Dimensionality reduction methods

In this case, we propose to combine two feature selection algorithms namely feature selection based mutual correlation (Algorithm 1) and feature quality measure (Algorithm 2) followed by feature extraction methods PCA or LPP which results in 3 levels of dimensionality reduction.

4.3.1 Feature selection based on mutual correlation + feature selection based on feature quality measure followed by PCA or LPP

Here feature selection based on mutual correlation is applied to select a feature subset at first level (Algorithm 1); at the second level another feature selection method based on feature quality measure is applied on the reduced subset (Algorithm 2). At third level, apply feature extraction methods PCA or LPP on this reduced feature subset to reduce the dimension further.

Algorithm 7: Dimensionality reduction using mutual correlation + feature quality measure followed by PCA or LPP.

Input: Original Dataset X of size N x p, class label

Output: New transformed feature set Y of dimension D (D<<f).

Method:

Level I: Apply Algorithm 1 on X to obtain the reduced set of size N x f.
Level II: Apply Algorithm 2 on the reduced feature set obtained from Level I to get the further reduced feature set of size N x fr (fr<<f).
Level III: Apply PCA or LPP on the reduced feature set of size N x fr obtained from Level II.

i.e Apply PCA on $X=[x_1,\ldots,x_{fr}]$ to obtain $Y(y_1,\ldots,y_D)=T(x_1,\ldots,x_{fr})$.

OR

Apply LPP on $X=[x_1,\ldots,x_{fr}]$ to obtain $Y(y_1,\ldots,y_D)=T(x_1,\ldots,x_{fr})$

4.3.2 Feature selection based on feature quality measure + feature selection based on mutual correlation followed by PCA or LPP

Here feature selection based on feature quality measure is applied to select a feature subset at first level (Algorithm 2); at the second level another feature selection method based on mutual correlation is applied on the reduced subset (Algorithm 1). At third level, apply feature extraction methods PCA or LPP on this reduced feature subset to reduce the dimension further.

Algorithm 8: Dimensionality reduction using feature quality measure + mutual correlation followed by PCA or LPP.

Input: Original Dataset X of size N x p, class label
Output: New transformed feature set Y of dimension D (D<<f).
Method:
Level I: Apply Algorithm 2 on X to obtain the reduced set of size N x f.
Level II: Apply Algorithm 1 on the reduced feature set obtained from Level I to get the further reduced feature set of size N x fr (fr<<f).
Level III: Apply PCA or LPP on the reduced feature set of size N x fr obtained from Level II.

i.e Apply PCA on $X=[x_1,\ldots,x_{fr}]$ to obtain $Y(y_1,\ldots,y_D)=T(x_1,\ldots,x_{fr})$.

OR

Apply LPP on $X=[x_1,\ldots,x_{fr}]$ to obtain $Y(y_1,\ldots,y_D)=T(x_1,\ldots,x_{fr})$

The standard complete linkage clustering algorithm has been employed on these reduced feature set to obtain the reliable clusters in all the above algorithms.

5. EXPERIMENTAL RESULTS

In this section, we present the experimental results of proposed models to corroborate the success of the proposed models. The well known existing dimensionality reduction techniques such as PCA, LPP and correlation based feature selection has been considered for comparative study. The superiority of the proposed models is established through the parameters precision, recall and F measure of the obtained clusters. Results of experiments performed on the standard datasets like WDBC (Wisconsin Diagnostic Breast Cancer), WBC (Wisconsin Breast Cancer), CORN SOYABEAN and WINE datasets are shown in tables 2, 3, 4 and 5. The summary of these datasets is given in the Table 1.

To measure the accuracy of the clusters obtained, precision, recall and F measure parameters are computed. The precision, recall and F measure are defined as follows:

Precision $= \dfrac{C_a \cap C_r}{C_r}$, Recall $= \dfrac{C_a \cap C_r}{C_a}$ and

F Measure $= \dfrac{2 * precision * recall}{precision + recall}$

where C_a is the actual number of elements in the cluster and C_r is the number of elements in the clusters obtained.

Comparative analysis is carried out and the average precision, average recall and average F measure values for datasets WDBC, WBC, CORN SOYABEAN and WINE datasets are tabulated in the tables 2, 3 4 and 5 respectively. In all these tables, methods 1 to 4 are single level dimensionality reduction methods, methods 5 to 10 are bi-level dimensionality reduction methods and methods 11 to 14 are multi-level dimensionality reduction methods. From these tables, it is clear that multi-level approaches show higher value of F measure when compared to single level and bi-level dimensionality reduction methods. The F measure performance of these methods for all the 4 datasets is shown in the figure 1. From this figure, it is clear that for WBC and WINE datasets, the methods in combination with LPP perform better when compared with other combinations.

Table 1: Summary of datasets used in this paper

Dataset	Instances	Features	Classes	Size of classes
WDBC	569	30	2	212: Malignant samples 357: Benign samples
WBC	683	9	2	239: Malignant samples 444: Benign samples
CORN SOYABEAN	61	24	2	32: Corn, 29:Soyabean
WINE	178	13	3	59:Class 1, 71: Class 2, 48:Class 3

6. CONCLUSION

In this paper, a novel dimensionality reduction method based on feature quality measure is proposed. Also we introduced multi-level dimensionality reduction approaches based on mutual correlation + FQ measure with PCA/LPP and FQ Measure + mutual correlation with PCA/LPP. The proposed models achieve better performance in terms of F measure with reduced feature set. Experiments are conducted on well known datasets like WDBC, WBC, CORN SOYABEAN and WINE to demonstrate the superiority of the proposed models. As a future work, we are exploring the possibility of fusing the features obtained from these multilevel approaches to get the further reduction of dimension with better performance in terms both accuracy and time.

Table 2: Comparison of various methods for WDBC dataset

Sl. No.	Methods	Average Precision	Average Recall	Average F measure
1	PCA	87.947	75.663	81.344
2	LPP	84.575	86.747	85.627
3	Mutual Correlation	89.523	78.538	83.638
4	FQ Measure	88.738	87.721	88.300
5	FQ Measure + Mutual Correlation	89.743	88.664	89.200
6	Mutual Correlation + FQ Measure	90.637	88.664	89.640
7	Mutual Correlation + PCA	92.166	87.271	89.652
8	Mutual Correlation + LPP	92.121	87.124	**89.752**
9	FQ Measure + PCA	92.426	87.360	89.822
10	FQ Measure + LPP	92.521	87.979	**90.193**
11	Mutual Correlation + FQ Measure + PCA	93.339	89.483	91.370
12	Mutual Correlation + FQ Measure + LPP	93.452	90.102	**91.746**
13	FQ Measure + Mutual Correlation + PCA	91.614	91.796	91.705
14	FQ Measure + Mutual Correlation + LPP	93.963	90.758	**92.333**

Table 3: Comparison of various methods for WBC dataset

Sl. No.	Methods	Average Precision	Average Recall	Average F measure
1	PCA	89.454	78.114	83.400
2	LPP	88.782	82.121	85.322
3	Mutual Correlation	91.484	83.135	87.110
4	FQ Measure	92.564	85.452	88.866
5	FQ Measure + Mutual Correlation	92.745	85.871	89.176
6	Mutual Correlation + FQ Measure	92.745	85.871	89.176
7	Mutual Correlation + PCA	92.451	86.981	89.633
8	Mutual Correlation + LPP	95.820	94.191	**94.999**
9	FQ Measure + PCA	92.451	86.981	89.633
10	FQ Measure + LPP	94.937	90.151	**92.482**
11	Mutual Correlation + FQ Measure + PCA	92.745	85.871	89.176
12	Mutual Correlation + FQ Measure + LPP	96.500	96.766	**96.633**
13	FQ Measure + Mutual Correlation + PCA	92.451	86.981	89.633
14	FQ Measure + Mutual Correlation + LPP	96.354	96.959	**96.655**

Table 4: Comparison of various methods for CORN SOYABEAN dataset

Sl. No.	**Methods**	**Average Precision**	**Average Recall**	**Average F measure**
1	PCA	98.485	98.276	**98.380**
2	LPP	95.054	95.151	**95.102**
3	Mutual Correlation	98.485	98.276	**98.380**
4	FQ Measure	98.485	98.276	**98.380**
5	FQ Measure + Mutual Correlation	100.00	100.00	**100.00**
6	Mutual Correlation + FQ Measure	100.00	100.00	**100.00**
7	Mutual Correlation + PCA	100.00	100.00	**100.00**
8	Mutual Correlation + LPP	100.00	100.00	**100.00**
9	FQ Measure + PCA	100.00	100.00	**100.00**
10	FQ Measure + LPP	100.00	100.00	**100.00**
11	Mutual Correlation + FQ Measure + PCA	100.00	100.00	**100.00**
12	Mutual Correlation + FQ Measure + LPP	100.00	100.00	**100.00**
13	FQ Measure + Mutual Correlation + PCA	100.00	100.00	**100.00**
14	FQ Measure + Mutual Correlation + LPP	100.00	100.00	**100.00**

Table 5: Comparison of various methods for WINE dataset

Sl. No.	**Methods**	**Average Precision**	**Average Recall**	**Average F measure**
1	PCA	89.845	88.749	89.294
2	LPP	91.095	91.608	91.351
3	Mutual Correlation	89.945	89.191	89.566
4	FQ Measure	93.172	93.984	93.577
5	FQ Measure + Mutual Correlation	94.331	92.893	93.607
6	Mutual Correlation + FQ Measure	93.124	94.366	93.741
7	Mutual Correlation + PCA	92.458	93.447	92.950
8	Mutual Correlation + LPP	94.715	95.584	**95.148**
9	FQ Measure + PCA	94.059	94.836	94.446
10	FQ Measure + LPP	95.284	96.149	**95.714**
11	Mutual Correlation + FQ Measure + PCA	95.835	95.569	95.702
12	Mutual Correlation + FQ Measure + LPP	95.756	95.767	**95.761**
13	FQ Measure + Mutual Correlation + PCA	95.519	95.569	95.544
14	FQ Measure + Mutual Correlation + LPP	96.018	96.489	**96.253**

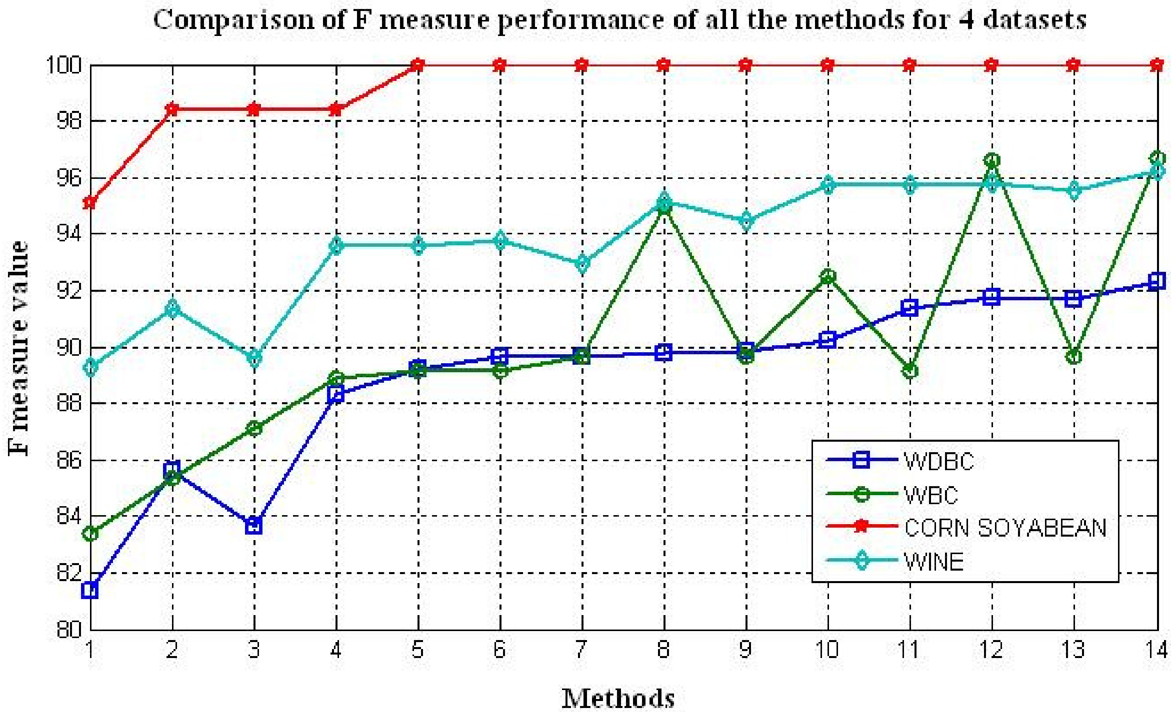

Figure 1: Comparison of F measure performance of all the methods for 4 datasets

REFERENCES

[1] Molina, L.C., Belanche, L., Nebot, A.: Attribute Selection Algorithms: A survey and experimental evaluation. Proceedings of 2nd IEEE's KDD, pp. 306-313, (2002).

[2] Guyon, I., Elisseeff, A.: An Introduction to Variable and Feature Selection. Journal of Machine Learning Research 3, pp. 1157-1182, (2003).

[3] Liu, H.A., Setiono, R.: Incremental feature selection. Applied Intelligence 9, pp. 217-230, (1998).

[4] Kohavi, R., John, G.H.: Wrappers for feature subset selection. Artificial Intelligence 97, pp. 273-324, (1997).

[5] Yu, L., Liu, H.: Feature Selection for High-Dimensional Data: A Fast Correlation-based Filter Solution. Proc. Int. Conference ICML2003, pp. 856-863, (2003).

[6] Dash, M., Liu, H.: Feature selection for classification. Intelligent Data Analysis: An International Journal 1 pp. 131-156, (1997)

[7] Park, J.S., Shazzad, K.M., Kim, D.S.: Toward modeling lightweight intrusion detection system through correlation-based hybrid feature selection. Information Security and Cryptology, Proceedings 3822, pp.279-289, ,(2005) .

[8] Dash, M., Liu, H., Motoda, H.: Consistency based feature selection. Knowledge Discovery and Data Mining, Proceedings,1805, pp. 98-109, (2000).

[9] T.Djatna, Y.Morimoto, A Novel feature selection Algorithms for Strongly Correlated Attributes using Two Dimensional Discriminant Rules, 6th IETCE Data Engg. Workshop , (2008).

[10] J.Hua, W.Tembe, E.R.Dougherty, Feature selection in the classification of high-dimension data. IEEE International Workshop on Genomic Signal Processing and Statistics, pp.1-2, (2008).

[11] X.Jin, A.Xu, R.Bie, P.Guo, Machine learning techniques and chi-square feature selection for cancer classification using SAGE gene expression profiles, Lecture Notes in Computer Science, 3916, pp. 106-115, (2006).

[12] C.Liao, S.Li, Z.Luo, Gene selection using Wilcoxon rank sum test and support vector machine for cancer, Lecture Notes in Computer Science, 4456,57–66, (2007).

[13] H.Peng, F.Long, C.Ding, Feature selection based on mutual information criteria of max-dependency, max-relevance and min redundancy, IEEE Transactions on Pattern Analysis and Machine Intelligence 27,pp. 1226-1238, (2005).

[14] J.Yang,V. Honavar, Feature subset selection using a genetic algorithm, IEEE Intelligent Systems and their Applications 13, pp. 44-49,(1998).

[15] P.Pudil , J.Novovicov, J.Kittler, Floating search methods in feature selection, Pattern Recognition Letters,15(11), pp. 1119–1125, (1994).

[16] M.Bensch, M.Schroder, M.Bogdan, W.Rosenstiel,P .Czerner,R. Montino,G. Soberger, P.Linke,R.Schmidt, Feature selection for high-dimensional industrial data ESANN 2005, pp. 27–29,(2005).

[17] H.T.Ng, W.B.Goh, K.L.Low Feature selection, perceptron learning, and a susability case study for text categorization, 20^{th} Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Philadelphia, pp.67–73, (1997).

[18] S.M.Vieira,M.C.Sousa,T.A.Runkler,Ant colony optimization applied to feature selection in fuzzy classifiers, Lecture notes in computer science, 4529, pp. 778–788, (2007).

[19] X.Wang,J.Yang,X.Teng,W.Xia,J.Richard, Feature selection based on rough sets and particle swarm optimization, Pattern Recognition Letters 28, pp. 459–471. (2007).

[20] M.Ronen,Z.Jacob,Using simulated annealing to optimize feature selection problem in marketing applications, European Journal of Operational Research,171, pp. 842–858, (2006).

[21] F.Tan, X.Fu, H.Wang, Y.Zhang, A.Bourgeois, A hybrid feature selection approach for micro array gene expression data, Lecture Notes in Computer Science,3992, pp. 678–685, (2006).

[22] K.M.Shazzad,J.S.Park,Optimization of intrusion detection through fast hybrid feature selection Proceedings of the Sixth International Conference on Parallel and Distributed Computing, IEEE Computer Society, Washington, DC, USA, pp.264–267,(2005).

[23] Z.Yan,C.Yuan ,Ant colony optimization for feature selection in face recognition, Lecture notes in Computer Science 307,pp. 221–226, (2004)

[24] K.M.Osei-Bryson, K .Giles,B. Kositanurit, Exploration of a hybrid feature selection algorithm ,Journal of the Operational Research Society 54, pp. 790–797, (2003)

[25] M.Fatourechi, G.Birch, R.K.Ward, Application of a hybrid wavelet feature selection method in the design of a self-paced brain interface system, Journal of Neuro engineering and Rehabilitation,4 (2007).

[26] J.Huang,Y.Cai,X.Xu, A wrapper for feature selection based on mutual information, 18^{th} International Conference on Pattern Recognition,vol.2, pp.618–621, (2006).

[27] Iffat A..Gheyas, Leslie S.Smith, Feature subset selection in large dimensionality domains, Pattern Recognition 43, pp. 5-13, (2010).

[28] M.Haindl, P.Somol, D.Ververidis, C.Kotropoulos, Feature Selection Based on Mutual Correlation, Proceedings of Progress in Pattern Recognition, Image Analysis and Application, 4225, pp. 569-577, (2006).

[29] Kira, K. and Rendell, L. A. A practical approach to feature selection, Proceedings of the Ninth International Conference on Machine Learning, Aberdeen, Scotland, UK, Morgan Kaufmann Publishers, San Mateo, pp. 249-256, (1992).

[30] Kononenko, I. Estimating attributes: Analysis and extension of RELIEF. In: F. Bergadano and L. De Raedt, editors, Proceedings of the European Conference on Machine Learning, Springer-Verlag. pp. 171-182, (1994).

[31] Luis Daza and Edgar Acuna, "Feature selection based on a data quality measure", Proceedings of the World Congress on Engg, Vol 11, (2008).

Authors

Veerabhadrappa obtained his B.Sc., and M.Sc., degrees in Computer Science from the University of Mysore, India, respectively, in the years 1987 and 1989. He is a faculty in the Department of Computer Science, University College, Mangalore, India and currently he is on Teacher Fellowship to carry out Ph.D in the Department of Studies in Computer Science, University of Mysore, India. He authored 8 peer-reviewed papers in journals and conferences. His area of research covers dimensionality reduction, face/object recognition and symbolic data analysis.

Dr. Lalitha Rangarajan is currently a Reader at the Department of Computer Science, University of Mysore, India. She has two master degrees one in Mathematics from Madras University, India and the other in Industrial Engineering (specialization: Operations Research) from Purdue University, USA. She has taught mathematics for five years in India soon after the completion of masters in mathematics. She is associated with the Department of Studies in Computer Science, University of Mysore, soon after completion of masters at Purdue University. She completed her doctorate in Computer Science in 2004 and since then doing research in the areas of image processing, retrieval of images, bioinformatics and pattern recognition. She has more than 30 publications in reputed conferences and journals to her credit.