Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 58 (2015) 499 – 506

Procedia Computer Science

Second International Symposium on Computer Vision and the Internet (VisionNet'15)

Multi-view Ensemble Learning using Optimal Feature Set Partitioning: An Extended Experiments and Analysis in Low Dimensional Scenario

Vipin Kumar

School of Computer and Systems Sciences, Jawaharlal Nehru University, New Delhi, INDIA

Abstract

Multi-view ensemble learning (MEL) has successfully addressed the issue related to high dimensionality of the data. It exploits the information of views of the data. To obtain views of data, an optimal feature set partitioning (OFSP) method [1] has been shown performance enhancement of MEL. Results of the experiments carried out on datasets and their statistical analysis show the effectiveness for classification problem in high dimensional scenario. In this work, classification performance of MEL using OFSP method has been analyzed in low dimensional situations. Therefore, experiments are performed on low dimension datasets using K-Nearest Neighbor (KNN), Naïve Bayesian (NB) and Support Vector Machine (SVM) classification algorithms. The experimental results and their statistical analysis show that OFSP method is also effective in low dimensional environment.

Peer-review under responsibility of organizing committee of the Second International Symposium on Computer Vision and the Internet (VisionNet'15)

Keywords: Classification, Optimal feature set partitioning, Low dimensionality, Multi-view ensemble learning.

1. Introduction

In the modern information technology, the datasets are collected from multiple sources. Therefore, a diversity of multiple pieces of information provides the insight for learning from multiple sources (views). A sub-table of data matrix with respect to the subset of attributes is considered as a view. The natural views of the dataset may be obtained on the basis of dataset acquisition and their characteristics are considered as the description of web pages on the basis of their text, images, hyperlinks, etc.; the views of the dataset may be obtained on the basis of categorical, numerical set of features and the contents in the hyperspectral imagery may be viewed as sets of bands. Moreover, many real-world datasets do not have a natural partition. Therefore, it is an important task to obtain the views of the dataset to express to enhance the classification performance of MEL [2, 3, 4, 5]. To obtain the views of the dataset, many strategies are proposed in literature such as random feature set partitioning [6, 7, 8, 9], collective performance [10, 11], reduct [12, 13, 14], rotation forest [15] etc.

Peer-review under responsibility of organizing committee of the Second International Symposium on Computer Vision and the Internet (VisionNet'15)
doi:10.1016/j.procs.2015.08.015

Optimal feature set partitioning (OFSP) method obtains a fixed number of views of the datasets. This method utilizes the classification algorithm to validate the individual view. It obtains the blocks of relevant and irrelevant features to be utilized by MEL. Therefore, analysis of a view place has reduced dimensionality in the instance space. The aim of the OFSP method is to optimize (maximize) the classification of individual views towards comparing the overall accuracy of the ensemble of classifiers. The novelty of the OFSP method has been explored in high dimensional data scenario using KNN, NB, and SVM classification algorithms. This work explores the experimental analysis of OFSP method for MEL. Experiments have been carried out on twelve low dimensional datasets, where KNN, NB, and SVM classification algorithms are used to induce classifiers for the views. The method has been compared with two state-of-art feature set partitioning methods. The results and their non-parametric statistical analysis show that the OFSP method has been outperformed in low dimension situation.

This paper is organized as follows: Section-2 describes the related work. Multi-view ensemble learning using optimal feature set partitioning is described in Section-3. Section-4 has presented a description of the datasets, experimental design and results. The results and their statistical tests are analyzed in Section-5. Finally, Section-6 has the conclusion and future research.

2. Related Work

Multi-view learning has been utilized for the supervised learning [3, 16, 17, 18, 19], clustering [20], semi-supervised learning [21], dimensionality reduction [22], ensemble learning [23], active learning [24], etc. Multi-view learning has followed two principles namely *consensus* and *complementary* [3]. The consensus principle maximizes the agreements among the classifiers, and complimentary principle utilizes the complimentary information of individual views of the dataset [3].

There are two main steps of MEL while learning; first, the view generation and second, an ensemble of learned classifiers [3]. To construct the views of the dataset, a number of methods have been proposed which are categorized such as random-based view construction, performance-based view construction, feature sets partition-based view construction and other feature set partition method [1]. In Random based view construction, a set of features is randomly split into subsets, where subsets may be disjoined. Random split for the view construction does not ensure the better performance of MEL in comparison to single view learning [21]. Tao et al. [25] reduced the discrepancy between training data size and feature vector length, which solves the related overfitting. Random subspace method (RSM) has been proposed using aggregation and bootstrapping [26]. In [8], Attribute bagging (AB) has been proposed to find the appropriate size of features for the partition and to construct the required views for the same size of features. To obtain the feature subsets automatically, View Construction using Genetic Algorithm (VC-GA) has been proposed [28]. Di and Crawford [29] constructed the views of the hyperspectral data using diversity, compatibility and accuracy measures. L. Rokach [30] employed genetic algorithm successfully for the feature set partitioning and suggested new encoding scheme.

For view evaluation, the conditional view entropy has been proposed by Christoudias [30] to filter subsets. Muslea [31] proposed view sufficiency algorithm for the prediction propose. *Heteroscedastic Bayesian Co-Training* identified the views noise [32]. A new confidence measure for intra-view and inter-view are defined in [33]. The views can be evaluated by prediction capability of the view using data mining algorithm. Kumar and Minz proposed an optimal feature set partitioning method for MEL, which utilizes the data mining algorithm for the evaluation of feature subsets [1].

3. Multi-view Ensemble Learning using Optimal Feature Set Partitioning (OFSP-MEL)

Dataset D is usually represented as a $m \times n$ data matrix, where m and n correspond to number of tuples and attribute in the data respectively. The data set D have attributes set $A = \{a_1, a_2, a_3, \dots, a_n\}$, where, $dom(a_i) = \{v_{i,1}, v_{i,2}, v_{i,3}, \dots, v_{i,|dom(a_i)|}\}$. Instance space of dataset is the cartesian product of all the input attributes set A, defined as in eq.1:

$$D = dom(a_1) \times dom(a_2) \times dom(a_3) \times \dots \times dom(a_n) \times dom(Y) \quad \text{..........................} (1)$$

where, $dom(y) = \{y_1, y_2, y_3, \dots, c_{|dom(y_i)|}\}$ are categorical labels.

The partition of dataset is the collection of subsets of the features set, which is defined as in Definition-1 [1].

Definition-1: (Feature Set Partition) [1]
"Let a data set D have nonempty features set $A = \{a_1, a_2, a_3, \dots, a_n\}$, . A partition of A is a nonempty collection of the nonempty subsets of A, $\pi_D = \{A_i \mid i \in I\}$ so that $\cup \{A_i \mid i \in I\} = A$ and $A_i \cap A_j = \emptyset$, and $i \neq j$; $j \in I; 1 \leq I \leq n$; where I is an integer and each subset A_i of π_D is a block of partition (i^{th} view of the dataset)."

The labeled view (sub-table) of the dataset can be written as in eq. 2:

$$D_i = A_i \times Y = a_{i,1} \times a_{i,2} \times a_{i,3} \times \dots \times a_{i,p} \times Y \dots\dots\dots\dots\dots\dots \quad (2)$$

where, $p < n$, the inducer f is applied to learn with each i^{th} view A_i of the dataset, where $i = 1, 2, 3, \dots, k$, which is defined as in eq. 3:

$$f : D_i \rightarrow Y \dots\dots\dots\dots\dots\dots \quad (3)$$

Therefore, a set of classifiers can be denoted as $F_{\pi_D} = \{f(D_i) | i = 1,2,3, \dots, k\} = \{f(D_1), f(D_2), f(D_3), \dots, f(D_k)\}$, which corresponds to each view of the dataset for π_D partition. The classifier of a view may be different from the other view.
If, the predicted class of the sample $s_i = (a_{i,1}, a_{i,2}, a_{i,3} \dots a_{i,p})$ of D_i is written as in eq.4:

$$\hat{y_i} = f(s_i) \dots\dots\dots\dots\dots\dots \quad (4)$$

Then, the error rate can be denoted as in eq. 5:

$$Err = \frac{1}{m} \sum_{i=1}^{m} I_f(y_i \neq \hat{y_i}) \dots\dots\dots\dots\dots\dots \quad (5)$$

where, y_i and I_f are the true class label of sample s_i and an indicator function respectively, where the false and true arguments are denoted as (0,1) values. The classification accuracy of i^{th} classifier can be denoted as in eq.6:

$$Acc_i = \frac{1}{m} \sum_{i=1}^{m} I_f(y_i \neq \hat{y_i}) = 1 - Err \dots\dots\dots\dots\dots\dots \quad (6)$$

Let, the weight of the i^{th} classifier be denoted as $\alpha_i = Acc_i$ and let, the weighted vote be denoted by $C_j(s_i)$ for class y_j for k number of classifiers, given as in eq.6:

$$C_j(s_i) = \sum_{t=1}^{k} \alpha_t \times I_f\left(f_p(s_i) = y_j\right) \dots\dots\dots\dots\dots\dots \quad (6)$$

C_j is takes care of classifiers weights for the class y_j. The prediction of the sample s_i can be written as in eq.7:

$$E^k(s_i) = \arg\max_{y_j}\{C_j(\ s_i) | \ j = 1,2,3, \dots m\} \dots\dots\dots\dots\dots\dots \quad (7)$$

where, the ensemble of the classifiers is denoted by E^k. The classification accuracy of the ensemble of the classifiers can be calculated by using eq.6.

OFSP method partition the dataset into blocks of relevant features and a block of irrelevant features, where the relevant features are measured on the basis of classification accuracy of the classifier. If a feature is not enhancing the classification accuracy of any of the classifiers that feature is considered as irrelevant. Let π_D be $(k + 1)$ blocks of partition, then k-number of blocks has relevant features and $(k + 1)^{th}$ block of partition π_D has irrelevant features. k-number of blocks of partition π_D can be identified rationally, which may be identified by system configuration for parallel programing or expert of the same domain. OFSP-MEL framework is shown in [1]. The optimal feature set partitioning is defined in Definition-2.

Definition-2: Optimal Feature Set Partition (OFSP) [1]
"Let a classifier f be a training dataset D with set of input features $A = \{A_1, A_2, A_3, \dots, A_n\}$ and target feature is y from distribution $\mathcal{P}$ over labelled instance space. A partition π_{opt} of A having k-number of blocks of partition is an optimal feature set partition, if the accuracy of the ensemble of the classifiers with respect to the block of π_{opt} is optimal i.e. $\forall\ \pi_i$ of A having k-number of blocks of partition".

Table-1: Optimal Feature Set Partitioning Algorithms (OFSP Algorithm) [1]

```
INPUTS:    D = A × Y      // A data set with | A |= n-number of attributes defined as
           k              // Number of the feature set partition;
           f              // Classifier;
OUTPUT :   π_opt          // Set of view that are optimal partitioned
Begin:
Initialization:
              for  i=1 to k
                  V_i= ∅;       // Creating k- number of  NULL views
                  Ac^i = 0;     // Initialization of classification accuracy of the classifier
              end
   do begin:
              for  attr  = 1 to n
                  for  i = 1 to k
                      V_temp =  V_i  ∪ {A_attr}; // adding attribute the i^th views.
                      Ac^i_temp = f_i  (V_temp)  // Evaluation of feature subset V_temp.
                                           //where classification accuracy is subset evaluation measure.
                      Ac^i_diff = Ac^i_temp –Ac^i;
                  end
                  if  max(Ac_diff)>0  then
                        t = arg max_i(Ac_diff);      // t^th view which has Ac^t_diff>0
                      V_t=  V_t ∪ {A_attr};
                       Ac^t =Ac^t_temp;
                  end
                 repeat
               end
              return π_opt = { V_1, V_2, V_3, … , V_k}
        end
    end
```

Table-1 shows the optimal feature set partitioning algorithm. The algorithm has three inputs namely dataset D, k-number of partition and f classifier. Initially, k-number of null views (subsets) are created. For each view V_i, an attribute A_{attr} is evaluated by classification accuracy of the corresponding classifier [1]. A_{attr} is considered relevant to the view, V_i if the previous accuracy Ac^i of the classifier for the same view is less than current Ac^i_{temp}. The view V_i would be the candidate view for A_{attr}, if $(Ac^i_{temp} - Ac^i)$ is maximum among the remaining views. Then, attribute A_{attr} is added to V_i. For the next A_{attr} from the remaining attributes, the same steps are followed in each iteration till n. OFSP method assumes that the each view satisfies learning sufficiency, because, each view would have at least one feature [1].

4. Experiments and Results

4.1. *Description of Datasets*

Twelve benchmark low dimensional datasets has been utilized for experiments, which are shown in the table-2 [34]. In this table, seven datasets are binary labeled and others are multi-class datasets.

Table 2. Description of Datasets

S.N.	Datasets	Number of Samples	Number of Attributes	Number of labels
1.	Breast Tissue	106	9	6
2.	Glass	214	9	6
3.	Ionosphere	351	33	2
4.	Leaf	340	16	30
5.	Movement Libras	360	91	15
6.	Musk2 Data	476	167	2
7.	Musk1 Data	476	167	2
8.	Parkinson	195	23	2
9.	SPECTF	80	44	2
10.	SPECT	80	23	2
11.	Sonar	208	61	2
12.	Wine	178	14	3

4.2. *Experimental Design*

The Server version of Matlab-2012b (64-bit) has been used for the implementation of the proposed method. KNN, NB and SVM based classifiers are included from each of the k-views for learning purpose [35]. Performance weighting ensemble method is applied to ensemble the view based classifiers. General classification accuracies are observed by performing 10-fold cross validation for ten times. The same setting has been done for all datasets. RFSP and VC-GA method are implemented for comparisons for OFSP method. The features set A is partitioned into k-blocks of partitioned for the RFSP method such as $\bigcup_{i=1}^{k} A_i = A$; and $A_i \cap A_j = \emptyset$; where $i \neq j; i,j = 1,2,\dots,5$. For the VC-GA method, the chromosomes length is equal to the number of attributes in the dataset, where bits are associated with features. The chromosomes are initialized randomly. Size of population and offspring are used 10 and 20 respectively. Crossover and mutation probability are used 0.66 and 0.33 respectively. Top $k = 1,2,\dots,5$ chromosomes are used for the respective number of views. For non-parametric statistical analysis the results of MEL, Friedman Align test is performed on results of classification accuracy [36, 37] and Holm procedure [37] is performed for the post-hoc test. The more non-parametric tests can be studied from [37].

4.3. *Results*

The boxplots of classification accuracies of MEL using RFSP, VC-GA and OFSP method are shown in the Figure-1 for twelve datasets. The Number 1, 2 and 3 on the x-axis show the RFSP, VC-GA and OFSP method respectively and the y-axis shows the classification accuracy of MEL in percentage. Figure-1(a), Figure-2 and Figure-3 show the boxplots of classification accuracies of MEL using KNN, NB and SVM algorithms respectively. Friedman Align and post-hoc test of the classification accuracies of MEL using KNN, NB and SVM have been shown in Table-3. In this table, ranks, p-value of post-hoc and p-value of Hommel's procedure of the RFSP, VC-GA and OFSP methods are shown for each dataset using KNN, NB and SVM algorithm.

5. Analysis

Figure-1(a) shows the boxplots of classification accuracies of MEL using RFSP, VC-GA and OFSP method for twelve datasets, where KNN classifier is used. The median of MEL classification accuracies using OFSP method is greater than RFSP and VC-GA methods for all datasets except Parkinson and SPECT datasets. Minimum classification accuracies of MEL using OFSP method are better than RFSP and VC-GA methods for all datasets except Parkinson, SPECT and Sonar datasets. MEL maximum classification accuracies using OFSP method are better than other feature set partitioning method for all datasets, with an exception of SPECT dataset.

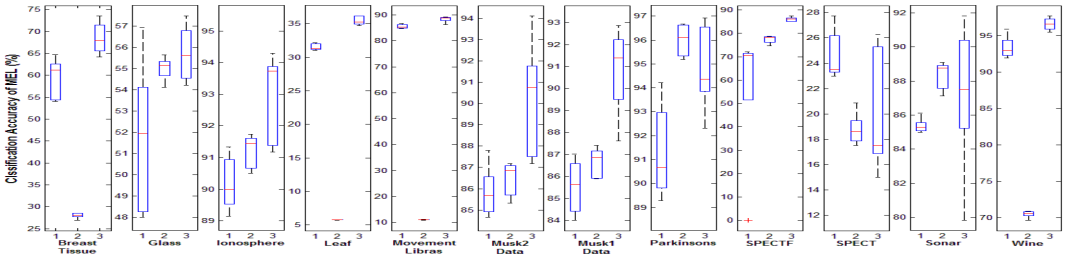
Figure-1(b) shows the boxplots of classification accuracies of MEL using RFSP, VC-GA and OFSP method for twelve datasets, where NB classifier is used for learning purpose. Median and maximum classification accuracies of MEL using OFSP method are greater than VC-GA and RFSP method for all datasets, except SPECT dataset. Except Glass, Ionosphere and SPECT datasets, the classification accuracies of MEL using OFSP method are better than other feature set partitioning methods for all datasets.

Figure-1(c) shows the boxplots of classification accuracies of MEL using RFSP, VC-GA and OFSP method for twelve datasets, where SVM classifier is used for learning purpose. The median of MEL classification accuracies using OFSP method are better than RFSP and VC-GA methods for all datasets with an exception of Except Glass, Musk1, Musk2, Parkinson and SPECT datasets,. Breast tissue, Leaf, Movement Libra, Sonar and Wine are the datasets where the minimum classification accuracies of MEL using OFSP method are better than RFSP and VC-GA methods. The maximum MEL classification accuracies using OFSP method are greater than other feature set partitioning methods for Breast tissue, Ionosphere, Parkinson, SPECTF, SPECTF, Sonar and Wine datasets.

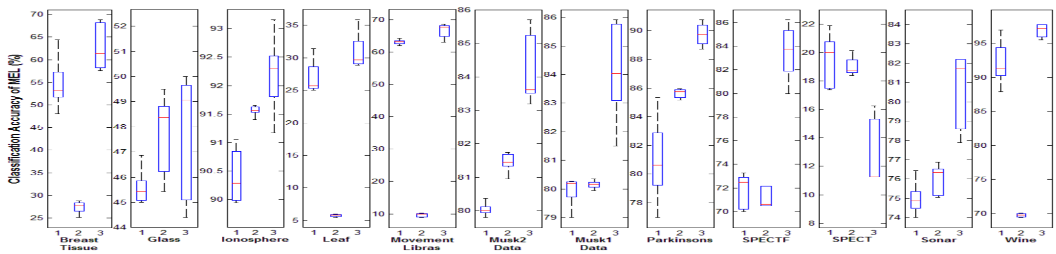
The above analysis shows that the median of MEL classification accuracies using OFSP method is better than RFSP and VC-GA methods over 10, 11 and 7 number of datasets, where KNN, NB and SVM classifier are used for learning respectively. Minimum classification accuracies of MEL using OFSP method is better than other feture set partitioning methods over 9, 9 and 5 number of datasets, where KNN, NB and SVM classifier are used for learning respectively. For 11, 11 and 6 numbers of dataset, maximum MEL classification accuracies using OFSP method is better than RFSP and VC-GA methods, where KNN, NB and SVM classifier are used for learning respectively. Therefore, it can be concluded that the classification performance of MEL using OFSP method is better than RFSP and VC-GA method.

Form Table-3; it can be observed that the ranks of OFSP method are less than RFSP and VC-GA methods for all datasets using KNN classifier, except Parkinson, SPECT and Sonar datasets. Except SPECT dataset, the ranks of OFSP method are less than other feature set partitioning methods for all datasets using NB classifier. Except Glass, Ionosphere, Movement Libra, Musk1, Musk2 and Parkinson datasets, the ranks of OFSP method are less than other feature set partitioning methods for all datasets using SVM classifier. For RFSP method, the p-values of post-hoc test are less than $\alpha = 0.05$ over 11, 11 and 10 number of datasets using KNN, NB and SVM classifiers respectively. 0, 3 and 0 are the number of datasets, where the p-values of post-hoc test for VC-GA method are less than $\alpha = 0.05$ using KNN, NB and SVM classifiers respectively. For RFSP, p-values of Hommel's procedure are less than $\alpha = 0.05$ over 10, 10 and 10 number of datasets using KNN, NB and SVM classifiers respectively. The p-values of Hommel's procedure for VC-GA method are less than $\alpha = 0.05$ over 0, 13 and 0 number of datasets using KNN, NB and SVM classifiers respectively. Form the above statistical analysis, it can be concluded that the OFSP method for MEL is better than RFSP and VC-GA methods.

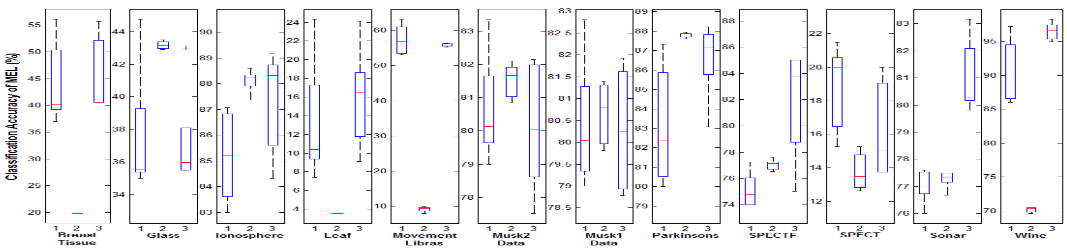
The boxplots of classification accuracies of MEL and statistical analysis of the classification results using RFSP, VC-GA and OFSP methods conclude that OFSP method is better than other feature set partitioning methods for MEL. It can also be concluded that NB classification algorithm performs better than KNN and SVM for MEL using OFSP method.



(a)



(b)



(c)

Fig. 1. (a), (b) and (c) are the boxplot of classification accuracies of MEL using RFSP (1), VC-GA (2) and OFSP (3) method on datasets, where KNN, NB and SVM algorithms are utilized respectively.

Table-3: Friedman Aligned Test of Classification Accuracy of MEL

Datasets	FSP Methods	KNN			NB			SVM		
		Ranks	p-value (Post -hoc)	Hommel's p-value	Ranks	p-value (Post -hoc)	Hommel's p-value	Ranks	p-value (Post -hoc)	Hommel's p-value
Breast Tissue	RFSP	8	0.00470	0.00081	7.8	0.00053	0.00106	6.4	0.00298	0.00596
	GA	13	0.07710	0.07710	13	0.10388	0.10388	13	0.52452	0.52452
	OFSP	3	-	-	3.2	-	-	4.6	-	-
Glass	RFSP	11.6	0.03390	0.06779	11.4	0.05624	0.11248	10.7	0.00985	0.01971
	GA	6.8	0.67137	0.67137	6.6	0.83200	0.83200	3.4	0.02156	0.02156
	OFSP	5.6	-	-	6	-	-	9.9	-	
Ionosphere	RFSP	13	0.00041	0.00081	13	0.00178	0.00178	12.8	0.00468	0.00936
	GA	8	0.07710	0.07710	7.4	0.17911	0.17911	4.8	0.57161	0.57161
	OFSP	3	-	-	3.6	-	-	6.4	-	-
Leaf	RFSP	8	0.00041	0.00081	8	0.00041	0.00081	6.6	0.00236	0.00472
	GA	13	0.07710	0.07710	13	0.07710	0.07710	13	0.43668	0.43668
	OFSP	3	-	-	3	-	-	4.4	-	-
Movement Libras	RFSP	8	0.00041	0.00081	8	0.00041	0.00081	4.8	0.00374	0.00748
	GA	13	0.07710	0.07710	13	0.07710	0.07710	13	0.62062	0.62062
	OFSP	3	-	-	3	-	-	6.2	-	-
Musk2 Data	RFSP	11.4	0.00298	0.00596	13	0.00041	0.00081	7.8	0.10388	0.20775
	GA	9.6	0.19624	0.19624	8	0.07710	0.07710	5.8	0.47950	0.47950
	OFSP	3	-	-	3	-	-	10.4	-	-
Must1 Data	RFSP	12.2	0.00114	0.00229	10.4	0.00721	0.00889	8.4	0.72367	0.77730
	GA	8.8	0.04031	0.04031	10.6	0.00889	0.00900	7.4	0.77730	0.77730
	OFSP	3	-	-	3	-	-	8.2	-	-
Parkinson	RFSP	13	0.00186	0.00373	13	0.00041	0.00081	12.8	0.00236	0.00472
	GA	4.2	0.35797	0.35797	8	0.07710	0.07710	4.2	0.32220	0.32220
	OFSP	6.8	-	-	3	-	-	7	-	-
SPECTF	RFSP	13	0.00114	0.00229	9.6	0.00298	0.00596	12	0.00236	0.00472
	GA	7.2	0.22933	0.22933	11.4	0.01962	0.01962	8.6	0.06599	0.06599
	OFSP	3.8	-	-	3	-	-	3.4	-	-
SPECT	RFSP	4.2	0.01333	0.02666	5	0.00468	0.00094	4.4	0.00801	0.01602
	GA	11.2	0.11980	0.11980	6	0.72367	0.72367	11.9	0.24332	0.24332
	OFSP	8.6	-	-	13	-	-	7.7	-	-
Sonar	RFSP	11	0.05624	0.11248	11.8	0.00186	0.00373	11.2	0.00374	0.00748
	GA	5.6	0.52452	0.52452	9.2	0.02838	0.02838	9.8	0.01621	0.01621
	OFSP	7.4	-	-	3	-	-	3	-	-
Wine	RFSP	8	0.00041	0.00081	8	0.00041	0.00081	8	0.00041	0.00081
	GA	13	0.07710	0.07710	13	0.07710	0.07710	13	0.07710	0.07710
	OFSP	3	-	-	3	-	-	3	-	-

6. Conclusion

In this work, the performance of MEL using OFSP method is studied in the low dimensional scenario. OFSP method is compared with RFSP and VC-GA methods using KNN, NB and SVM classifiers. The classification results and their non-parametric statistical analysis conclude that OFSP method is better than RFSP and VC-GA methods in low dimensional scenario. It is also recommended that NB classification algorithm may be utilized for MEL using OFSP method in low dimensional situation. Therefore, OFSP method for MEL may be utilized in low and high dimensional scenarios using NB and SVM classification algorithms for purpose of learning.

References

1. Vipin Kumar and Sonajharia Minz. Multi-view ensemble learning: An optimal feature set partitioning for high dimensional data classification. *Knowledge and Information System*, 2015.
2. R.K. Ando and T. Zhang. The two-view feature generation model for semi-supervised learning. in *ICML*, 2007.
3. Chang Xu, Dacheng Tao and Chao Xu. A survey on multi-view learning. *Learning (cs.LG),* 2013
4. S.M. Kakade and D.P. Foster. Multi-view regression via canonical correlation analysis. *In COLT*, 2007.
5. S. Yu, B. Krishnapuram, R. Rosales, H. Steck and R.B. Rao. Bayesian co-training. In *NIPS*, 2007.
6. T. K. Ho. Nearest Neighbors in Random Subspaces. *Proceeding of the Second International Workshop on Statistical Techniques in Pattern Recognition,* Sydney, Australia, August 11-13, 1998, p. 640–648,
7. S. Bay. Nearest neighbor classification from multiple feature subsets. *Intelligent Data Analysis*, 1999, 3(3), p.191-209.
8. R. Bryll, R. Gutierrez-Osunaa and F. Quek. Attribute bagging: improving the accuracy of classifier ensembles by using random feature subsets. *Pattern Recognition*, 2003, Volume 36, p. 1291-1302.
9. L. Rokach, O. Maimon and O. Arad. Improving Supervised Learning by Sample Decomposition. *International Journal of Computational Intelligence and Applications*, 2005, 5(1), p.37-54.

10. P. Cunningham and J. Carney. Diversity Versus Quality in Classification Ensembles Based on Feature Selection. In R. L. de Mntaras and E. Plaza (eds.), Proc. ECML 2000, *11th European Conf. On Machine Learning, Barcelona*, Spain, LNCS 1810, Springer, 2000, p. 109-116.
11. G. Zenobi and P. Cunningham. Using diversity in preparing ensembles of classifiers based on different feature subsets to minimize generalization error. *In Proceedings of the European Conference on Machine Learning*, 2001.
12. Q. X. Wu, D. Bell and M. McGinnity. Multi-knowledge for decision-making. *Knowledge and Information Systems*, 7(2005): 246-266
13. Q. H. Hu, D. R. Yu, M. Y. Wang. Constructing Rough Decision Forests. D. Slezak et al. (Eds.): RSFDGrC 2005, *LNAI 3642*, Springer, 2005, p. 147-156.
14. Y. Bao, N. Ishii. Combining multiple K-nearest neighbor classifiers for text classification by reducts. *In Proceedings of the 5th international conference on discovery science*, *LNCS 2534,* Springer, 2002, p. 340-347.
15. J. J. Rodriguez. Rotation Forest: A New Classifier Ensemble Method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(10), 2006, p.1619-1630.
16. M. Kudo and J. Sklansky. A Comparative Evaluation of medium and large–scale Feature Selectors for Pattern Classifiers. *In Proceeding of the 1st International Workshop on Statistical Techniques in Pattern Recognition*, Prague, Czech Republic, 1997, p. 91–96,
17. A. L. Bluma and P. Langley. Selection of Relevant Features and Examples in Machine Learning. In R. Greiner and D. Subramanian, eds., *Artificial Intelligence on Relevance, Artificial Intelligence*, volume 97, 1997, p. 245–271.
18. Vipin Kumar and Sonajharia Minz. Multi-view Ensemble Learning for Poem Data Classification using SentiWordNet. *2nd International Conference on Advanced computing , Networking, and Informatics (ICACNI-2014),* Smart Innovation, Systems and Technogies, Springer Verlag, Volume-27, 2014, p:57-66.
19. Sonajharia Minz and Vipin Kumar. Reinforced Multi-view Ensemble Learning for High Dimensional Data Classification. *International Conference on Communication and Computing (ICC-2014),* ELSEVIER, 2014.
20. M. Ben-Bassat. Pattern recognition and reduction of dimensionality. In P. R. Krishnaiah and L. N. Kanal, editors, *Handbook of statistics-II*, North Holland, 1982, p. 773–791.
21. Brefeld, C. Gartner and T. Scheffe. Multi-view Discriminative sequential learning. *Machine Learning, ECML 2005*, 2005, p. 60-71.
22. M.A. Hall. Correlation-based feature selection for discrete and numeric class machine learning. *In Proceedings of the 17th International Conference on Machine Learning*, 2000, p. 359–366.
23. H. Almuallim and T.G. Dietterich. Learning boolean concepts in the presence of many irrelevant features. *Artificial Intelligence,* 69(1-2), 1994, p.279–305.
24. P. A. Devijver and J. Kittler. Pattern Recognition – A Statistical Approach. *Prentice Hall, London*, GB, 1982.
25. D. Tao, X. Tang, X. Li, and X. Wu. Asymmetric bagging and random subspace for support vector machines-based relevance feedback in image retrieval. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(7), 2006, p.1088–1099.
26. T.K. Ho. The random subspace method for constructing decision forests. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 20(8), 1998, p.832–844.
27. S. Sun, F. Jin, and W. Tu. View construction for multi-view semi-supervised learning. *Advances in Neural Networks–ISNN 2011*, 2011, p. 595–601.
28. W. Di and M.M. Crawford. View generation for multi-view maximum disagreement based active learning for hyperspectral image classification. *Geoscience and Remote Sensing, IEEE transaction on*, (99), 2012, p.1-13.
29. L. Rokach. Genetic algorithm-based feature set partitioning for classification problem. *Pattern Recognition*, 41(5), 2008, p.1676–1700.
30. C.M. Christoudias, R. Urtasun, and T. Darrell. Multi-view learning in the presence of view disagreement. *In Proceedings of the 24th Conference on Uncertainty in Artificial Intelligence*, 2008.
31. I. Muslea, S. Minton, and C.A. Knoblock. Adaptive view validation: A first step towards automatic view detection. *In the Conference on Machine Learning-International Workshop,* , 2002, p. 443–450.
32. C.M. Christoudias, R. Urtasun, A. Kapoorz, and T. Darrell. Co-training with noisy perceptual observations. *In Computer Vision and Pattern Recognition*, 2009. CVPR, 2009, IEEE Conference on, IEEE, 2009, p.2844–2851.
33. Liu and P.C. Yuen. A boosted co-training algorithm for human action recognition. *IEEE Transactions on circuits and systems for video technolog*y, 21(9), 2011, p.1203–1213.
34. https://archive.ics.uci.edu/ml/datasets.html.
35. http://www.37steps.com/prtools/.
36. J.L. Hodges, E.L. Lehmann. Rank Method for a combination of independents experiments analysis of variance. *Annals of Mathematical Statistics*, Volume-33, 1962, p. 482-497.
37. S. Garcia, A. Fernandez, J. Luengo and F. Herrera. Advanced non-parametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: Experimental analysis of Power. *Information Science*, 18, 2010, p: 2044-2064.