Proceedings of the 5th International Conference on Computing and Informatics, ICOCI 2015
11-13 August, 2015 Istanbul, Turkey. Universiti Utara Malaysia (http://www.uum.edu.my)

Paper No. 222

ANT SYSTEM-BASED FEATURE SET PARTITIONING ALGORITHM FOR K-NN AND LDA ENSEMBLES CONSTRUCTION

Abdullah[1] and Ku Ruhana Ku-Mahamud[2]

[1]*Universitas Islam Indragiri, Indonesia, abdialam@yahoo.com*
[2]*Universiti Utara Malaysia, Malaysia, ruhana@uum.edu.my*

ABSTRACT. Combination of several classifiers has been very useful in improving the prediction accuracy and in most situations multiple classifiers perform better than single classifier. However not all combining approaches are successful at producing multiple classifiers with good classification accuracy because there is no standard resolution in constructing diverse and accurate classifier ensemble. This paper proposes ant system-based feature set partitioning algorithm in constructing k-nearest neighbor (*k*-NN) and linear discriminant analysis (LDA) ensembles. Experiments were performed on several University California, Irvine datasets to test the performance of the proposed algorithm. Experimental results showed that the proposed algorithm has successfully constructed better classifier ensemble for *k*-NN and LDA.

Keywords: *k*-nearest neighbor, linear discriminant analysis, feature set partitioning, ant system algorithm, classifier ensemble

INTRODUCTION

Several classification algorithms for pattern classification have been developed. The *k*-nearest neighbor (*k*-NN) and linear discriminant analysis (LDA) are two of the most widely used algorithm for classification tasks. However, there is not a single classifier that can be considered optimal for all pattern classification problems. Therefore the multiple classifier combination (or ensemble method) in the form of a hybrid intelligent approach is considered as a new direction in pattern classification. Combining classifier is considered as a general solution to solve classification problems (Koyuncu & Ceylan, 2013; Margoosian & Abouei, 2013). Previous studies have shown that the combination of several classifiers has been very useful in improving the prediction accuracy (Turhal et al., 2013). It has been shown that in most situations multiple classifiers perform better than single classifier. However not all combining approaches are successful at producing multiple classifiers with good classification accuracy.

Multiple classifier combinations consist of a set of classifiers called ensemble. Classifier ensemble construction aims to establish a set of accurate and diverse classifiers. It has been shown theoretically and empirically that a good ensemble is in which the individual classifier has both good accuracy and diversity (Parvin et al., 2009). However, classifier ensemble construction problems have not been fully resolved. There is no standard resolution in constructing diverse and accurate classifier ensemble (Schiele, 2002; Hernandez-Lobato & Martinez-Munoz, 2013).

The commonly used approach in constructing classifier ensemble is the training data manipulation (Yang et al., 2010). This approach works very well with unstable classifiers which can produce diverse predictions even though there is only a very small change in the training data. However *k*-nearest neighbor (*k*-NN) and linear discriminant analysis (LDA) are very stable classifiers, thus generally this approach is not suitable in constructing the classifier ensemble. Stable classifier means that small changes in the training data set will not cause large changes in the classifier output.

Another approach in building a classifier ensemble is to use input feature manipulation (Roli, 2009). Feature decomposition methods are those that manipulate the input feature set in creating diverse classifier ensemble. In this method the input features on the training set are decomposed in order to build a classifier ensemble. Feature decomposition method potentially facilitates the creation of a classifier for high dimensionality data sets without the feature selection drawback (Rokach, 2010). The feature decomposition method is also known as the feature subset based ensemble. Feature set partitioning is a special case of feature subset-based ensemble. Feature set partitioning does not just search for single useful subset but the original feature set is decomposed into several subsets and a set of classifiers trained on a disjoint feature subset. This approach is appropriate for the classification task with large number of features.

Classifier ensemble built with a different subset of features has been shown to be effective in practice. Based on this approach, one popular way to generate different feature subset is through random subspace (RS) method (Ho, 1998). This method produces feature subspaces that are selected randomly from a subset of features in the original representation of space, and then a set of classifiers is built based on the selected subspace. Random subspace method has performed satisfactorily and has proven resistant to irrelevant features. Bay (1999) presented multiple feature subset (MFS) approach that combines many *k*-NN classifiers each using random feature subset. The final decision is obtained as the majority voting result of the classifiers. The experimental results showed that the MFS improved classification performance. Ahn et al. (2007) showed that the randomly partitioned input features to several subsets thus each classifier was trained on different subsets, particularly useful for high-dimensional datasets and unbalanced data. However random selection could not find the optimal subset of features for a combination of several classifiers.

Several studies have applied Ant System (AS) algorithm for set partitioning problems were reported by Maniezzo and Milandri (2002), Randall and Lewis (2010) and Crawford et al. (2013). AS algorithm is an original and most popular variant of ant colony optimization (ACO) based algorithm that has been used and proven to solve various optimization problems (Rebeiro & Enembreck, 2013). ACO was introduced by Marco Dorigo as a metaheuristic method for the solution of hard combinatorial optimization problems (Dorigo & Blum, 2005). The Ant System is also applied to solve the set partitioning problem which is one of the most difficult (NP-Hard) and very constrained combinatorial problems due to their complexities.

The work presented here aims at optimizing the number of classifier in an ensemble by using the appropriate feature manipulation while maintaining classification performance. In this paper, a new algorithm is proposed to construct better *k*-NN and LDA ensembles. Section 2 explains the proposed ant system-based feature set partitioning (ASFSP) algorithm. Section 3 presents the experiments and comparisons were performed with several other methods by using several benchmark dataset from UCI to test the performance of the proposed algorithm. Section 4 briefly concludes the work.

PROPOSED METHOD

In this proposed ASFSP, classifier ensemble is constructed based on input feature manipulation approach. A disjoint feature set decomposition is performed based on the original training set. Feature set is partitioned into different feature subset. There is no feature in the training set that is eliminated. Furthermore each classifier in the ensemble is trained on a different projection of the original training set to induce diversity. The number of features subsets or partitions determines the number of classifiers in the ensemble. Ant system-based algorithm is developed to perform feature set partitioning. The required inputs are the feature set and class labels of original training set. The flowchart of feature decomposition algorithm is provided in Figure 1.

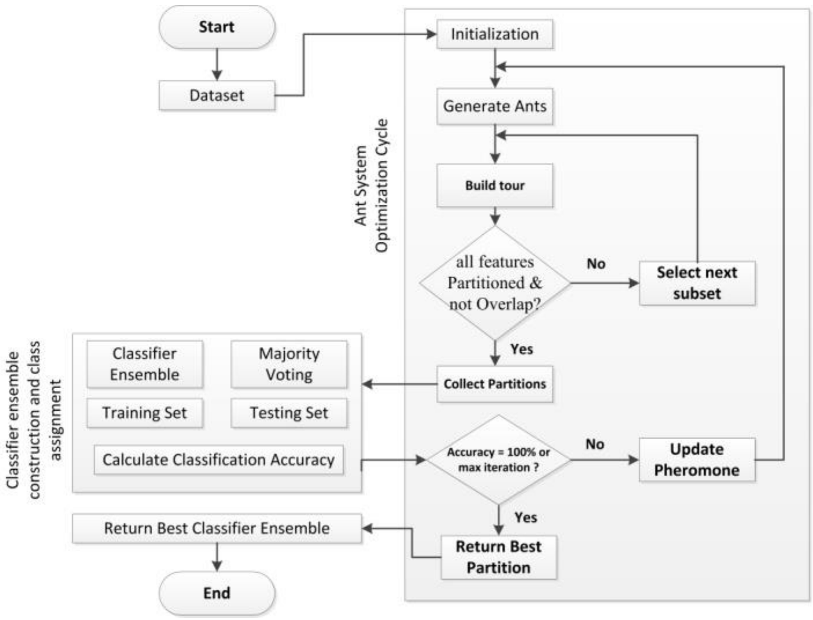

Figure 1. Flowchart of the generic ant system-based feature set partitioning algorithm

In the implementation of ASFSP, the required inputs are features in dataset. The pheromone table is initialized followed by the generation of the ants. Each ant then builds a tour in the form of a feature partition which is considered as a possible solution. The tour is evaluated if it contains all the features and no overlap features. Otherwise the next feature subset is selected until the feature partitions have been collected. This will be done repeatedly until a possible solution is built. Furthermore partitioned feature is used to construct classifier ensemble. The class assignment is performed using constructed classifier ensemble by using majority voting combiner. The best partition will be formed if classification accuracy reaches 100% or the maximum iteration limit has been reached. The pheromone is then updated and another ant is generated if any criterion is not fulfilled. The whole process is repeated until the best partition is formed.

EXPERIMENTAL RESULTS

Experiments were conducted to test the proposed algorithm to construct *k*-NN and LDA ensembles. Prediction class label of unknown pattern is obtained by using the majority voting Experiments were performed using nine (9) data sets from University California, Irvine (UCI) repository. Ten (10) experiments were performed to estimate the accuracy of the constructed classifier ensembles by using random subspace method and newly constructed classifier ensembles by using ASFSP algorithm. The 10-fold cross validation approach is used to validate the proposed algorithm. Tables 1 and 2 depict the average and standard deviation of the classification accuracy. It can be shown that a small deviation of the classification accuracy was obtained and this showed that the experiments were accurate. It can also be shown that the ASFSP algorithm give better accuracy than RS method in constructing *k*-NN ensembles.

Table 1. Classification Accuracy of *k*-NN Ensembles using RS

Experiment #	Haberman	Iris	Lenses	Liver	Ecoli	Pima	Tic-Tac-Toe	Glass	Breast Cancer
1	67.32	93.33	70.83	55.07	80.36	69.79	72.13	73.83	97.36
2	65.03	94.00	58.33	60.00	81.25	71.09	76.10	71.96	97.36
3	69.28	92.67	62.50	58.84	78.87	68.10	77.56	74.77	96.93
4	69.28	94.00	66.67	66.09	84.23	72.66	77.56	71.50	97.80
5	67.32	93.33	58.33	57.97	80.95	69.53	72.13	73.36	96.93
6	68.30	92.67	62.50	56.81	79.46	70.83	76.10	73.36	97.22
7	64.38	93.33	58.33	57.10	81.25	69.92	74.53	75.70	96.78
8	70.26	93.33	58.33	60.29	82.44	70.96	78.18	71.96	97.07
9	70.59	93.33	62.50	64.35	83.63	70.31	78.18	68.69	97.66
10	67.32	94.00	66.67	64.06	79.46	72.66	74.53	71.96	97.22
Average	**67.91**	**93.40**	**62.50**	**60.06**	**81.19**	**70.59**	**75.70**	**72.71**	**97.23**
Standard deviation	1.96	0.47	4.17	3.48	1.70	1.32	2.19	1.86	0.31

Table 2. Classification Accuracy of *k*-NN Ensembles using ASFSP

Experiment #	Haberman	Iris	Lenses	Liver	Ecoli	Pima	Tic-Tac-Toe	Glass	Breast Cancer
1	72.22	96.00	79.17	65.80	80.95	71.48	74.74	73.36	97.95
2	72.88	96.00	79.17	62.61	81.25	71.74	74.32	72.43	97.51
3	72.22	96.00	79.17	65.51	81.25	70.18	76.83	72.90	97.80
4	72.55	96.00	79.17	64.06	80.65	70.44	75.78	72.90	97.51
5	72.22	96.00	79.17	62.61	82.14	71.48	76.10	72.90	97.51
6	73.20	96.00	79.17	62.61	81.25	70.18	75.47	74.30	97.51
7	72.88	96.00	79.17	64.06	80.36	71.22	76.10	73.83	97.36
8	72.22	96.00	79.17	65.51	80.95	70.44	76.83	71.03	97.36
9	74.51	95.33	79.17	66.25	81.55	72.14	75.05	72.90	97.80
10	72.55	96.00	79.17	62.61	81.55	70.83	76.10	72.43	97.66
Average	**72.75**	**95.93**	**79.17**	**64.16**	**81.19**	**71.01**	**75.73**	**72.90**	**97.60**
Standard deviation	0.71	0.21	0.00	1.50	0.50	0.70	0.84	0.88	0.20

Tables 3 and 4 depict the average and standard deviation of the classification accuracies of constructed LDA ensembles based on RS and newly constructed LDA ensembles based on ASFSP respectively. Small deviation of the classification accuracy was obtained which indicate that the experiments were good. The ASFSP algorithm always gives better accuracy than RS method in constructing LDA ensembles for all the datasets.

Table 3. Classification Accuracy of LDA Ensembles using RS

Experiment #	Haberman	Iris	Lenses	Liver	Ecoli	Pima	Tic-Tac-Toe	Glass	Breast Cancer
1	74.84	96.00	79.17	62.03	73.23	75.00	65.66	57.48	96.19
2	72.88	95.33	79.17	62.03	73.23	74.87	65.66	57.94	96.49
3	72.88	94.67	83.33	63.77	73.03	75.00	65.66	59.35	96.19
4	74.84	95.33	87.50	62.03	73.23	74.61	66.18	60.28	96.49
5	73.20	94.67	79.17	63.48	73.23	74.87	65.66	56.54	95.90
6	73.20	97.33	79.17	62.61	74.41	75.39	64.72	59.81	96.19
7	72.88	96.00	79.17	60.58	73.23	73.44	65.66	59.81	96.49
8	74.84	96.67	79.17	63.48	73.00	75.13	64.72	60.75	95.90
9	74.84	95.33	83.33	60.58	73.23	75.13	66.18	59.81	96.05
10	73.20	96.00	75.00	63.77	72.99	75.91	66.18	60.28	96.19
Average	**73.76**	**95.73**	**80.42**	**62.44**	**73.28**	**74.94**	**65.63**	**59.21**	**96.21**
Standard deviation	0.94	0.84	3.43	1.21	0.41	0.63	0.53	1.39	0.22

Table 4. Classification Accuracy of LDA Ensembles using ASFSP

Experiment #	Haberman	Iris	Lenses	Liver	Ecoli	Pima	Tic-Tac-Toe	Glass	Breast Cancer
1	74.84	98.00	87.50	63.48	75.96	75.52	73.05	62.66	97.07
2	75.16	98.00	87.50	63.77	75.90	75.91	72.79	62.66	97.07
3	74.51	98.00	83.33	63.77	75.96	76.17	72.53	62.76	97.22
4	74.51	98.00	87.50	63.77	76.00	75.78	72.92	62.66	97.22
5	74.51	98.00	87.50	64.06	75.89	75.65	73.70	61.79	97.51
6	74.51	98.00	87.50	64.06	75.96	76.04	73.18	62.66	97.22

7	75.16	98.00	87.50	64.06	75.96	76.17	72.53	62.66	97.22
8	74.84	98.00	83.33	63.48	75.79	76.17	73.96	62.03	97.22
9	75.16	98.00	87.50	64.06	75.96	76.04	72.79	62.66	97.22
10	75.16	98.00	87.50	64.06	75.96	76.82	72.79	62.66	97.07
Average	**74.84**	**98.00**	**86.67**	**63.86**	**75.93**	**76.03**	**73.02**	**62.52**	**97.20**
Standard deviation	0.31	0.00	1.76	0.24	0.06	0.36	0.47	0.33	0.13

Table 5 and Table 6 show the summary of result in constructing *k*-NN and LDA ensembles respectively. The average of accuracy of newly constructed ensemble classifiers by ASFSP are compared with single approach and RS method.

Table 5. Comparison of Single Approach, RS and ASFSP in Constructing *k*-NN Ensembles

No	Dataset	Single Approach	RS	ASFSP
1	Haberman	66.83	67.91	**72.75**
2	Iris	95.67	93.40	**95.93**
3	Lenses	77.92	62.50	**79.17**
4	Liver	62.32	60.06	**64.16**
5	Ecoli	81.19	81.19	81.19
6	Pima	67.37	70.59	**71.01**
7	Tic-Tac-Toe	75.51	75.70	**75.73**
8	Glass	72.71	72.71	**72.90**
9	BreastCancer	95.78	97.23	**97.60**

Table 6. Comparison of Single Approach, RS and ASFSP in Constructing LDA Ensembles

No	Dataset	Single Approach	RS	ASFSP
1	Haberman	73.73	73.76	**74.84**
2	Iris	97.33	95.73	**98.00**
3	Lenses	86.25	80.42	**86.67**
4	Liver	62.35	62.44	**63.86**
5	Ecoli	72.91	73.28	**75.93**
6	Pima	75.34	74.94	**76.03**
7	Tic-Tac-Toe	65.62	65.63	**73.02**
8	Glass	58.83	59.21	**62.52**
9	BreastCancer	96.18	96.21	**97.20**

Based on the results, it can be shown that the RS method does not always give better accuracy than a single classifier. Instead ASFSP algorithm gives better results than the single classifier approach. This is because the usage of ASFSP has successfully formed the optimal feature set partition to induce diversity in constructing ensembles. Table 7 and Table 8 present the feature set partition and the number of classifier in ensemble with respect to ensemble accuracy. The number of partitions determines the number of classifiers. It can be shown that features set partitions are not formed on several dataset. This means that this proposed algorithm able to determine either the single classifier or an ensemble classifier is better for the dataset.

Table 7. Feature Set Partition and Number of *k*-NN Classifier

No	Dataset	Partition	# of Classifier	Accuracy
1	Haberman	[1 3][2]	2	72.75
2	Iris	[1 2 3 4]	1	95.93
3	Lenses	[1 2 3 4]	1	79.17
4	Liver	[1 4 6][3 5][2]	3	64.16
5	Ecoli	[1 2 3 4 5 6 7]	1	81.19
6	Pima	[1 3 4 7][5 6 8][2]	3	71.01
7	Tic-Tac-Toe	[1 2 3 4 5 6 7 8 9]	1	75.73
8	Glass	[1 2 3 4 5 6 7 8 9]	1	72.90
9	BreastCancer	[1 2 4 7 9][3 5][6][8]	4	97.60

Table 8. Feature Set Partition and Number of LDA Classifier

No	Dataset	Partition	# of Classifier	Accuracy
1	Haberman	[1][2 3]	2	74.84
2	Iris	[1 2 3 4]	1	98.00
3	Lenses	[1 2 3 4]	1	86.67
4	Liver	[1 3 4 6][2][5]	3	63.86
5	Ecoli	[1 3 5][4 6][2 7]	3	75.93
6	Pima	[1 2 3 4 5 6 7 8]	1	76.03
7	Tic-Tac-Toe	[2 4 5 6 8][1][3][7][9]	5	73.02
8	Glass	[2 3 5 7][4 8 9][1 6]	3	62.52
9	BreastCancer	[2 4 8][7 9][3][1 5 6]	4	97.20

The proposed algorithm has successfully partition the feature set to several feature subsets which may lead to a better classification performance. Improvement on accuracy is obtained on datasets where feature partitions have been performed. This is due to the classifier ensembles being constructed on datasets that form feature partition, where each individual classifier is trained on a different subset of features to induce diversity. Otherwise the relatively same accuracy with the original single classifier will be obtained on several datasets where no feature partition is performed.

CONCLUSION

A new feature set partitioning algorithm based on AS has been presented. Feature set partition is performed and ant system algorithm is used for optimization. Base classifier is trained on a different feature partition to induce diversity. The majority voting rule was used as combination rule in the experiments. The proposed algorithm was evaluated on several datasets from UCI repository. The results show that implementation of this algorithm in constructing both *k*-NN and LDA ensembles outperforms their single version and also RS method. Results indicated that the proposed algorithm can be used in constructing better *k*-NN and LDA ensembles. Future work is to implement this algorithm on other classifiers and by using other ACO-based algorithm.

REFERENCES

Ahn, H., Moon, H., Fazzari, M. J., Lim, N., Chen, J. J., & Kodell, R. L. (2007). Classification by ensembles from random partitions of high-dimensional data. *Computational Statistics and Data Analysis, 51*, 6166-6179.

Bay, S. D. (1999). Nearest neighbour classification from multiple feature subsets. *Intelligent Data Analysis, 3*, 191-209.

Crawford, B., Soto, R., Monfroy, E., Castro, C., Palma, W., & Paredes, F. (2013). A hybrid soft computing approach for subset problems. *Mathematical Problems in Engineering, 2013*, 1-12.

Dorigo, M., & Blum, C. (2005). Ant colony optimization theory: A survey. *Theoretical Computer Science, 344*, 243 – 278.

Hernandez-Lobato, D., & Martinez-Munoz, G. (2013). How large should ensembles of classifiers be? *Pattern Recognition, 46(5)*, 1323-1336.

Ho, T. K. (1998). The random subspace method for constructing decision forest. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(8)*, 832-844.

Koyuncu, H., & Ceylan, R. (2013). Artificial neural network based on rotation forest for biomedical pattern classification. *Proceedings of the 36th International Conference on Telecommunications and signal processing*, 581-585.

Maniezzo, V., & Milandri, M. (2002). An ant-based framework for very strongly constrained problems. *Proceedings of 3rd International Workshop on Ant Algorithms, LNCS, Vol. 2463*, 222-227.

Margoosian, A., & Abouei, J. (2013). Ensemble-based classifiers for cancer classification using human tumor microarray data. *Proceedings of the 21st International Conference on Electrical Engineering*, 1-6.

Parvin, H., Alizadeh, H., & Minaei-Bidgoli, B. (2009). A new method for constructing classifiers ensembles. *International Journal of Digital Content Technology and its applications, 3(2)*, 62-66.

Randall, M., & Lewis, A. (2010). Modifications and additions to ant colony optimisation to solve the set partitioning problem. *Proceedings of the 6th International Conference on e science Workshops*, 110-116.

Rebeiro, R., & Enembreck, F. (2013). A sociologically inspired heuristic for optimization algorithm: A case study on ant systems. *Expert System with Application, 40,* 1814-1826.

Rokach, L. (2010). *Pattern classification using ensemble methods.* Singapore: World Scientific Publishing.

Roli, F. (2009). Multiple classifier system. In S. Z. Li, & A. K. Jain, *Encyclopedia of Biometrics,* 981-986. New York: Springer Science+Business Media.

Schiele, B. (2002). How many classifier do I need. *International Conference of Pattern Recognition*, 176-179.

Turhal, U., Babur, S., Avci, C., & Akbas, A. (2013). Performance improvement for diagnosis of colon cancer by using ensemble classification methods. *International Conference on Technological Advances in Electrical, Electronics and Computer Engineering*, 271-275.

Yang, P., Yang, Y. H., Zhou, B., & Zomaya, Y. (2010). A review of ensemble methods in bioinformatics. *Current Bioinformatics, 5(4)*, 296-308.