

Ant-based Feature Decomposition Method in Constructing NMC and NBC ensembles

Abdullah

Department of Information System, Universitas Islam Indragiri
Tembilahan, Indragiri Hilir Riau 29213, Indonesia

Abstract—Several approaches have been proposed to construct a set of diverse classifiers within an ensemble. One of the approaches is the input features manipulation. Feature decomposition methods are those that manipulate the input feature set in creating the ensemble. However, it is difficult to determine how to partition the feature set into several feature subsets to train base classifiers which may lead to an accurate and diverse ensemble. This paper proposes ant-based feature decomposition method in constructing nearest mean classifier (NMC) ensembles and naïve bayes classifier (NBC) ensembles. Experiments were carried out on several University California, Irvine (UCI) datasets to test the performance of the proposed method. Experimental results showed that the proposed method has successfully constructed better nearest mean classifier (NMC) and naïve bayes classifier (NBC) ensembles.

Keywords Feature decomposition, classifier ensemble, ant-system algorithm

I. INTRODUCTION

The nearest mean classifier (NMC) and naïve bayes classifier (NBC) are two widely used single classifier for classification tasks. However, no single classifier that perfect for all pattern classification problems. Multiple classifier combination (or ensemble method) is considered as a general solution for classification problems [1-2]. Previous works have shown that the ensemble method has been useful in enhancing the classification accuracy [3-4]. It has been shown that in most situations combining approaches perform better than single classifier approach. However not all combining approaches are successfully in producing better classification performance.

Ensemble Classifier consists of a set of classifier ensemble and a combination rule for combining the classifier outputs. In a set of classifier, the combination is only useful if they disagree on several inputs [5]. Therefore, an ensemble should be built as diverse as possible. Roli [6] suggested several approaches to construct a classifier ensemble as follows: (1) using different base classifiers, (2) injecting randomness, (3) manipulating training data, (4) manipulating input features, and (5) manipulating output labels.

The idea of the input features manipulation approach is to train base classifiers on different

feature subsets. Feature decomposition methods are those that manipulate the input feature set in creating the ensemble. Maimon and Rokach [7] developed a general framework for feature decomposition. Figure 1 shows the general framework of feature decomposition. However, it is difficult to determine how to partition the feature set into several feature subsets to train base classifiers which may lead to an accurate and diverse ensemble.

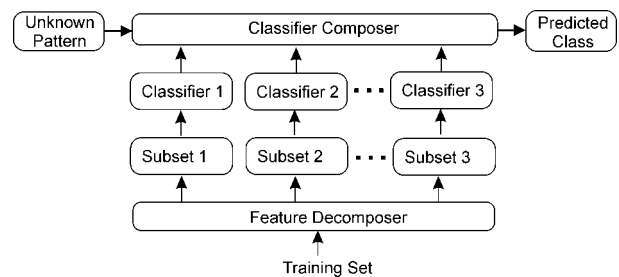


Fig. 1. General framework for feature decomposition

Feature set partitioning is a special case of feature decomposition. In feature set partitioning, the training set is decomposed into several subsets and a set of classifiers trained on a disjoint feature subset. Feature set may be partitioned by random selection, statistical approaches, and genetic algorithm [8]. Ahn et al. [9] showed that randomly partitioned input features to several subsets will enable each classifier to train on different subsets. Rokach [10] applied genetic algorithm (GA) for feature set partitioning. This technique has been tested with different datasets and results show advantages as compared to other techniques.

Ant colony optimization (ACO) algorithm has shown a better performance than genetic algorithms [11-12]. The ant system (AS) algorithm is a variant of the ACO algorithm. Ant system was the original term used to refer to a range of ACO-based algorithms, where the specific algorithm implementation was referred to as Ant Cycle. Ant Cycle algorithm is now referred to as ant system. This is the original and most famous variant of the ACO-based algorithms that has been used and is proven to solve various optimization problems [13-16]. Furthermore, AS has been successfully applied in solving the set partitioning problem [17-18].

In this paper, Ant-based feature decomposition method is proposed to construct classifier ensembles. A majority of the techniques reported in the literature focused on feature selection. However, the

assumption that the input feature set can be removed to a small subset of relevant features is not always correct. In several cases, removing features will lead to a significant loss of valuable information [7]. The proposed method will use all the features as opposed to other techniques that will remove several features which result in the loss of information.

This paper is structured as follows. The proposed Ant-based feature set decomposition method is presented in Section 2. Section 3 presents the experiments that have been conducted to test the proposed ensemble construction method, while Section 4 presents the conclusion.

II. THE ANT-BASED FEATURE DECOMPOSITION METHOD

In this study, A method in constructing classifier ensemble based on feature decomposition is proposed. The Ant-based feature decomposition method (AFDM) is used to construct NMC and NBC ensembles. A disjoint feature set decomposition is performed based on the original training set. No feature in the original training set is eliminated. The required inputs are the feature set and class labels of the original training set. The original training set is split into two parts, namely training set and validation set. Each classifier in the ensemble is trained on a different feature partition using the training set. The classification accuracy of the ensemble is obtained using the validation set.

The main steps of this method are as follows: (1) input original training set, (2) generate a graph problem based on features in the original training set, where each node will present a unique feature subset, (3) initialize the input parameter value, pheromone trail value and the number of artificial ants, (4) each ant will randomly build 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 overlapping features. Otherwise, the next feature subset is selected until feature partitions have been collected. This will be done repeatedly until a possible solution is built, (5) evaluate the classifier ensemble using validation set. The best partition will be formed if the classification accuracy reaches 100% or the maximum iteration limit has been reached, (6) if any criterion is not fulfilled, update pheromone and generate new ants. The whole process is repeated until the best partition is formed. The matlab code is used to implement this method. Figure 2 presents the pseudocode of the proposed method.

```

%Input      : Original training set
%Output     : Best feature partition, best classifier ensemble
Begin
[b,a]=loaddata('dataset.xxx');%load features in dataset
[n nod d h]=generate_problem(a) % generate graph problem
[t,iter,alpha,beta,rho,m,el]=initialization(n); %initialization
for i=1:iteration
    [app]=generate_ants(m,n) %generate ants
    [tabu]=build_tour(app,m,n,nod,h,t,alpha,beta)%built tour
    [clust]=conversion(tabu)%built tour
    [path]=substitutes(nod,clust)%collect partition
    [path error accuracy]=ensemble_accuracy(b,a,path) %evaluation
    [maxaccuracy(i),number]=max(accuracy)
    besttour(i,:)=path(number,:);
    if max(accuracy)==100
        break
    end
    [t]=ants_traceupdating1(t,clust,accuracy,rho);%update
    pheromone
end
[k,l]=max(maxaccuracy)
accuracy=k
best_partition=[{besttour{1,:}}]%return best partition
End
    
```

Fig. 2. Pseudo code for Ant-based feature decomposition

III. RESULTS AND DISCUSSION

Experiments were conducted to determine the performance of our proposed method. Ten experiments were conducted to test the AFDM in constructing homogeneous NMC and NBC ensembles. Nine data sets from UCI machine learning repository are used to perform classification experiments. The 10-fold cross validation approach is used to estimate the accuracy of constructed classifier ensembles by the AFDM. The prediction category is obtained by combining predictions using the majority voting rule. The average accuracies of constructed ensembles are compared with the average accuracies of constructed homogeneous ensembles by random subspace method (RSM) [19-21]. The number of feature subsets is set to four. The number of features for each subset is selected randomly with replacement. The number of ants is set to the number of nodes ($m=n$), $\alpha=1$, $\beta=1$ and $\rho=0.5$. The comparison between RSM and AFDM and the detailed information are as depicted in Tables 1 and 2.

Table 1. Comparison of RSM and AFDM in constructing homogeneous NMC ensembles

Dataset	Classifier ensemble construction					
	RSM			AFDM		
	Average of accuracy (%)	Feature subset	# of classifier	Average of accuracy (%)	Feature partition	# of classifier
Haberman	70.33	[1 3][1 2 3][1 2][3]	4	70.39	[1][2 3]	2
Iris	92.07	[1 2 3 4][1 2 3][1 3 4][3]	4	94.47	[1][2 3][4]	3
Lenses	66.25	[2 3 4][1 3][1 3 4][2]	4	66.67	[1 2 3 4]	1
Liver	56.43	[1 4 5][5 6][1 2 3 4 5 6][2 3 4 5]	4	64.29	[1 2 4 6][3][5]	3
Ecoli	81.67	[2 5 7][1 2 4 5 7][1 3 7][2]	4	81.82	[1 2 3 4 5 6 7]	1
Pima	67.88	[1 2 3 6 7 8][3 4 5][3 6 7][2 3 5 6]	4	73.02	[3 4 5 7][1 6][8][2]	4
Tic-Tac-Toe	64.49	[1 4 5 9][1 2 5 6][2 3 5 6][1 2 3 4 5 7 8 9]	4	73.01	[2 4 5 8][7][3 6 9][1]	4
Glass	44.44	[2 3 5 6 9][3 7 8 9][1 2 3 4 8 9][1 3 5 6 8]	4	53.22	[2 3 5 7][1 4 8 9][6]	3
Breast Cancer	96.50	[6 7 8 9][3 4 6 8][1 2 3 4 6 7 8][1 4 5 8]	4	97.23	[1 2 3 4 5 7 9][6 8]	2

Table 1 shows the comparison results of RSM and AFDM in constructing homogeneous NMC ensembles. Most of the datasets have been successfully partitioned, for instance, haberman, iris, liver, pima, tic-tac-toe, glass, and breast cancer. On

two datasets, which are lenses and ecoli, AFDM does not partition the features. This means that AFDM chooses the single classifier, instead of an ensemble classifier.

Table 2. Comparison of RSM and AFDM in constructing homogeneous NBC ensembles

Dataset	Classifier ensemble construction					
	RSM			AFDM		
	Average of accuracy (%)	Feature subset	# of classifier	Average of accuracy (%)	Feature partition	# of classifier
Haberman	74.61	[1 2 3][1 2][2][3]	4	74.81	[1 2 3]	1
Iris	94.80	[1 3 4][2 4][1 2 4][2]	4	95.46	[1 2 3 4]	1
Lenses	62.50	[2 3 4][3 4][1 2 3][1 2 3 4]	4	62.50	[1 2 4][3]	2
Liver	60.12	[2 3][2 4 5 6][1 2 3 4 6][2 5]	4	63.51	[1 2 3 4][5][6]	3
Ecoli	75.25	[2 3 4 7][2 4 5 6][4 5][1 2 3 5]	4	75.53	[1 2 3 4 5 6 7]	1
Pima	75.70	[1 2 3 5 7 8][2 4 7 8][1 2 3 4][1 2 3 7]	4	75.44	[1 2 3 4 5 6 7 8]	1
Tic-Tac-Toe	68.34	[2 3 5 7 8 9][1 3 4 5 6 7][1 3 5 8 9][1 2 3 4 6]	4	72.61	[1 2 3 4 5 6 7 8 9]	1
Glass	73.21	[1 3 6][1 3 4 6 9][3 8][2 4 5 7 8 9]	4	73.25	[1 2 3 4 5 6 7 8 9]	1
Breast Cancer	96.14	[1 3 4 7 8][5 6][2 3 4 6 8][2 6 9]	4	97.63	[4 5 8 9][1 2 7][6][3]	4

Table 2 shows the comparison results of RSM and AFDM in constructing homogeneous NBC ensembles. On all datasets, AFDM has successfully delivered better classification results. An increase is clearly seen on liver and breast cancer datasets. Features for the two datasets were successfully partitioned. The number of partitions for the liver dataset, which has been constructed by AFDM, is less than RSM. The same number of partition on breast cancer dataset constructed by AFDM and RSM has been obtained, but the partitions are of different form. The lenses dataset has also successfully been partitioned with comparable classification accuracy.

Based on the results, the two methods are different in forming partitions. In RSM, the features are randomly selected with replacement, thus, feature subsets can be overlapped and also there is a possibility that several features are not selected. However, if AFDM is used, all features will be used and no features will be used more than once. In both methods, partitions or feature subsets that are formed are used to train classifiers in the ensemble. The number of feature subsets or partitions indicates the number of classifiers in the ensemble. The number of classifiers in RSM is specified beforehand, while the number of classifiers in AFDM is automatically determined.

The usage of RSM provides lower accuracy, even though the number of classifiers is bigger. The usage of AFDM can easily determine the optimal number of classifiers. AFDM has successfully delivered better classification results with an optimal number of classifiers. The summary of results for single approach, RSM and AFDM in constructing NMC and NBC ensembles are shown in Tables 3 and 4 respectively.

Table 3. The Result in Constructing NMC Ensembles

Dataset	Single approach	RSM	AFDM
Haberman	69.97	70.33	70.39
Iris	92.07	92.07	94.47
Lenses	65.83	66.25	66.67
Liver	55.19	56.43	64.29
Ecoli	81.55	81.67	81.82
Pima	63.29	67.88	73.02
Tic-Tac-Toe	63.19	64.49	73.01
Glass	44.16	44.44	53.22
Breast Cancer	96.49	96.50	97.23

Table 4. The result in constructing NBC ensembles

Dataset	Single approach	RSM	AFDM
Haberman	74.51	74.61	74.81
Iris	95.47	94.80	95.46
Lenses	62.50	62.50	62.50
Liver	55.42	60.12	63.51
Ecoli	74.69	75.25	75.53
Pima	75.77	75.70	75.44
Tic-Tac-Toe	72.54	68.34	72.61
Glass	73.02	73.21	73.25
Breast Cancer	96.13	96.14	97.63

Based on the summary of results, it can be seen that there is an increase in accuracy on all datasets when ensembles were constructed using the proposed method. Obvious improvement accuracy is obtained on datasets that successfully form any feature set partition, because each individual classifier is trained on a different subset of features to induce diversity.

IV. CONCLUSION

A new feature decomposition method using AS algorithm for accurate and diverse classifier ensemble construction has been presented. Classifier ensembles were trained on different feature partitions to induce diversity. The utilization of AS was to produce the optimal feature set partition. The proposed method was evaluated on several benchmark datasets. The results show that the implementation of this proposed method in constructing NMC and NBC ensembles outperforms single approach and RSM. The use of AFDM able to form the optimal feature set partitions. Furthermore, the number of classifiers can be automatically determined by the number of feature set partitions that have been formed. The proposed method will decide whether a single or ensemble approach is suitable to be constructed. Future work would be to apply this method for heterogeneous classifier ensemble construction.

REFERENCES

- [1] H. Koyuncu, and R. Ceylan, "Artificial neural network based on rotation forest for biomedical pattern classification," in *Proceedings of the 36th International Conference on Telecommunications and signal processing (TSP)*, 2013, pp. 581-585.
- [2] A. Margoosian, and J. Abouei, "Ensemble-based classifiers for cancer classification using human tumor microarray data," in *Proceedings of the 21st International Conference on Electrical Engineering*, 2013, pp. 1-6.
- [3] U. Turhal, S. Babur, C. Avci, and A. Akbas, "Performance improvement for diagnosis of colon cancer by using ensemble classification methods," in *Proceedings of International Conference on Technological Advances in Electrical, Electronics and Computer Engineering (TAEECE)*, 2013, pp. 271-275.
- [4] S.K. Shukla, and A. Pandey, "Classification of Devnagari Numerals using Multiple Classifier," *International Journal of Computer Trends and Technology (IJCTT)*, vol. 12, no. 4, pp. 196-200, 2014.
- [5] L.I. Kuncheva and C.J. Whitaker, "Measures of diversity in classifier ensembles and their relationship with ensemble accuracy," *Machine Learning*, vol. 51, no. 2, pp. 181-207, 2003.
- [6] F. Roli, "Multiple classifier system," *Encyclopedia of Biometrics*, eds. S.Z. Li, and A.K. Jain. New York: Springer Science & Business Media, 2009, pp. 981-986.
- [7] O. Maimon, and L. Rokach, *Decomposition methodology for Knowledge Discovery and Datamining*, Berlin, Germany: Springer, 2005.
- [8] L. Rokach, *Pattern classification using ensemble method*, Singapore: World Scientific, 2010.
- [9] H. Ahn, H. Moon, M.J. Fazzari, N. Lim, J.J. Chen, and R.L. Kodell, "Classification by ensembles from random partitions of high-dimensional data," *Computational Statistics and Data Analysis*, vol. 51, no. 12, pp. 6166-6179, 2007.
- [10] L. Rokach, "Genetic algorithm-based feature set partitioning for classification problems," *Pattern Recognition*, vol. 41, no. 5, pp. 1676-1700, 2008.
- [11] C.T. Su, C.F. Chang, and J.P. Chiou, "Distribution network reconfiguration for loss reduction by ant colony search algorithm," *Electric Power Systems Research*, vol. 75(2-3), pp. 190-199, 2005.
- [12] C.F. Chang, "Reconfiguration and capacitor placement for loss reduction of distribution systems by ant colony search algorithm," *IEEE Transactions on Power Systems*, vol. 23, no. 4, pp. 1747-1755, 2008.
- [13] F. Shang, and Y. Wang, "An ant system optimization QoS routing algorithm for wireless sensor networks," in *Proceedings of the 3rd International Workshop on Advanced Computational Intelligence*, 2010, pp. 25-27.
- [14] A. Jevtic, D. Andina, A. James, J. Gomez, and M. Jamshidi, "Unmanned aerial vehicle route optimization using ant system algorithm," in *Proceedings of the 5th International Conference on System of Systems Engineering*, 2010, pp. 1-6.
- [15] R. Rebeiro, and F. Enembreck, "A sociologically inspired heuristic for optimization algorithm: A case study on ant systems," *Expert System with Application*, vol. 40, no. 5, pp. 1814-1826, 2013.
- [16] A.H. El Bakely, and H.A. Hefny, "Using Ant Algorithm in Green Cloud Computing to Minimize Energy," *International Journal of Computer Trends and Technology (IJCTT)*, vol. 27, No. 1, pp. 44-50, 2015
- [17] B. Crawford, C. Castro, and E. Monfroy, "A new ACO transition rule for set partitioning and covering problems," in *Proceedings of International Conference of Soft Computing and Pattern Recognition*, 2009, pp. 426-429.
- [18] B. Crawford, R. Soto, E. Monfroy, C. Castro, W. Palma, and F. Paredes, "A hybrid soft computing approach for subset problems," *Mathematical Problems in Engineering*, 2013, pp. 1-12.
- [19] T.K. Ho, "The random subspace method for constructing decision forest," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 8, pp. 832-844, 1998
- [20] G. Serpen, and S. Pathical, "Classification in high-dimensional feature spaces: Random subsample ensemble," in *Proceedings of IEEE International Conference on Machine Learning and Application*, 2009, pp. 740-745.
- [21] H. Li, G. Wen, Z. Yu, and T. Zhou, "Random subspaces evidence classifier," *Neurocomputing*, vol. 110, pp. 62-69, 2013