International Journal of Soft Computing 11 (3): 176-184, 2016
ISSN: 1816-9503

Ant System-Based Feature Set Partitioning Algorithm for Classifier Ensemble Construction

[1]Abdullah and [2]Ku Ruhana Ku-Mahamud
[1]Department of Information System, Universitas Islam Indragiri, 29213 Tembilahan, Indonesia
[2]School of Computing, Universiti Utara Malaysia, Sintok, 06010 Kedah, Malaysia

Abstract: Ensemble method is considered as a new direction in pattern classification. Accuracy and diversity in a set of classifiers are two important things to be considered in constructing classifier ensemble. Several approaches have been proposed to construct the classifier ensemble. All of these approaches attempt to generate diversity in the ensemble. However, classifier ensemble construction still remains a problem because there is no standard guideline in constructing a set of accurate and diverse classifiers. In this study, Ant system-based feature set partitioning algorithm for classifier ensemble construction is proposed. The Ant System Algorithm is used to form an optimal feature set partition of the original training set which represents the number of classifiers. Experiments were carried out to construct several homogeneous classifier ensembles using nearest mean classifier, naive Bayes classifier, k-nearest neighbor and linear discriminant analysis as base classifier and majority voting technique as combiner. Experimental results on several datasets from University of California, Irvine have shown that the proposed algorithm has successfully constructed better classifier ensembles.

Key words: Feature decomposition, classifier ensemble construction, ant system algorithm, nearst mean classifier, pattern

INTRODUCTION

The multiple classifier combination is considered as a new direction to solve classification problems. Multiple classifier combination has been widely used in several application domains (Wu and Liang, 2011; Srimani and Koti, 2012; Margoosian and Abouei, 2013). Experimental studies have shown that the combination of several classifiers has been very helpful in improving the prediction accuracy and reduces the generalization error (Kittler *et al*., 1998). The concept of multiple classifier combination was first proposed by Suen to enhance character recognition. Later, the neural network ensemble was proposed by Hansen and Salamon (1990) to transform the weak neural network into strong neural network. Multiple classifier combination aims to obtain the final classification decision by integrating the outputs of several individual classifiers. This can be written mathematically as follows: Let, $D = \{D_1, D_2, \ldots, D_L\}$ be a set of classifiers and $\Omega = \{\omega_1, \omega_2, \ldots, \omega_c\}$ be a set of c class labels. Each classifier D_i (i = 1, ..., L) receives a feature vector as an input and assigns it to one of the c class labels from Ω, i.e., $D_i: R^n \rightarrow \Omega$.

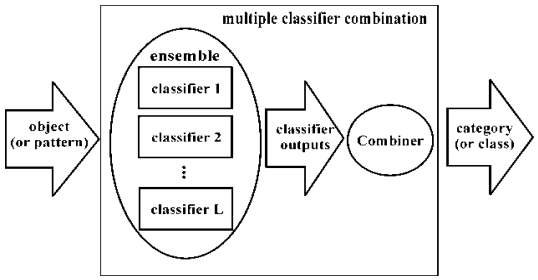

Fig. 1: Multiple classifier combination architecture

Multiple classifier combination consists of a set of classifiers (also called classifier ensemble) and a combiner for combining classifier outputs (Ponti, 2011). Figure 1 shows the architecture of a generic multiple classifier combination for any pattern classification task.

The classifier ensemble construction is to construct a set of classifiers as a base classifier of multiple classifier combination. The construction of the classifier ensemble aims to establish a set of diverse classifiers that complement each other. For this purpose, an ensemble should be built as diverse as possible. In classifier ensemble, the combination is only useful if they disagree on some inputs (Kuncheva and Whitaker, 2003).

Corresponding Author: Abdullah, Department of Information System, Universitas Islam Indragiri, 29213 Tembilahan, Indonesia

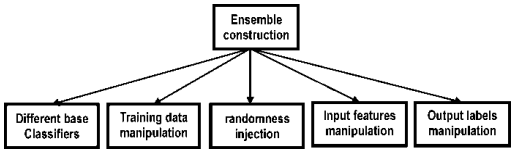

Fig. 2: Approaches for diverse classifier ensemble construction

Several approaches have been proposed to construct a set of diverse classifiers within an ensemble. Wanas and Kamel (2002) have summarized several approaches to construct an ensemble of neural network classifiers by varying the: Initial conditions, network topology, training algorithm and data training. Abreu and Canuto (2007) have summarized three ways to construct diverse classifier ensembles as follows: different parameters of classifier, such as weights and topology of neural network models, different learning algorithms, such as neural network, naïve bayes, or decision tree and different training sets, which is obtained from the original training set by resampling. Roli (2009) suggested several approaches to construct a classifier ensemble as follows: Using different base classifiers, injecting randomness, manipulating training data, manipulating input features and manipulating output labels. Figure 2 shows the five approaches as summarized by him.

All these approaches attempt to induce classifier diversity with the aim to create classifiers that make errors on different patterns and thus, they can be combined effectively. Accuracy and diversity are two important factors when combining multiple classifiers (Kang and Doermann, 2005). It has been shown empirically that a good ensemble is where the individual classifier has both accuracy and diversity (Parvin *et al*., 2009). However, there is no standard guideline for constructing an accurate and diverse classifier ensemble (Hernandez-Lobato *et al*., 2013).

One of the approaches to construct classifier ensemble is the input features manipulation. The idea of the input features manipulation approach is to train base classifiers on different feature subsets of the training set. Feature decomposition methods are those that manipulate the input feature set in creating the ensemble. Maimon and Rokach (2005) developed a general framework for feature decomposition. Figure 3 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.

Feature set partitioning is a special case of feature decomposition. It does not just search for a single useful subset. In feature set partitioning, the training set

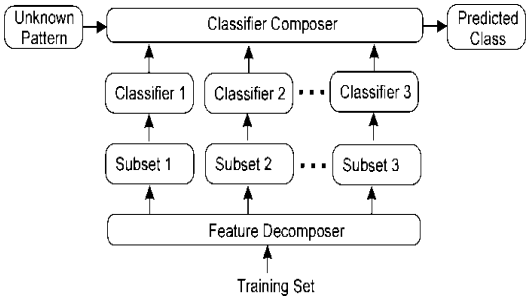

Fig. 3: General framework for feature decomposition

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 (Rokach, 2010). Ahn *et al*. (2007) showed that randomly partitioned input features to several subsets will enable each classifier to train on different subsets. Rokach (2008) applied genetic algorithm 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 (Su *et al*., 2005; Chung, 2008). Ant colony optimization was introduced by Marco Dorigo (1992) as a metaheuristic method for the solution of hard combinatorial optimization problems (Dorigo and Blum, 2005). This algorithm is inspired by the behavior of ants in finding the shortest path from the colony to the food. Ants are able to find the shortest route to the food using a chemical called pheromone. Pheromones are used by living things to recognize other individuals, groups and to assist the process of reproduction. In contrast to hormones, pheromones spread outside the body and can only affect and be recognized by other similar individuals of the same species. This process is known as pheromones relics stigmergy which is the process of modifying an environment that not only aims to remember the way back to the nest but also allows the ants to communicate with its colonies. Over time however, the pheromone trail will evaporate and will reduce the strength of its appeal. When the ants commute longer through these pathways, the pheromones will evaporate over time.

The optimal path can be obtained through the following processes: Ants move randomly to find food and at the same time, pheromone is laid on the path. Ants will bring the food back to the nest when they have found the food and thus, the food path will be formed. Other ants will follow the food path and more pheromones will be laid on the path. An ant which accidentally finds the optimal path will move faster than his colleagues, conduct more round-trips frequently and consequently leave more

pheromones on the path. A highly concentrated pheromone path will attract other ants to change paths to the most optimal path while other paths will be abandoned. Finally, all ants that travel on different paths will switch to a single most optimal path from the nest to the food.

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 (Shang and Wang, 2010; Jevtic *et al*., 2010; Ribeiro and Enembreck, 2013). Furthermore, Ant System has been successfully applied in solving the set partitioning problem (Crawford *et al*., 2009, 2014). In this study, ant system-based feature set partitioning 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 (Maimon and Rokach, 2005). The proposed technique will use all the features as opposed to other techniques that will remove several features which result in the loss of information. The use of Ant System in the feature set partitioning technique will be 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 algorithm will decide whether a single or ensemble approach is suitable to be constructed.

MATERIALS AND METHODS

In this study, The classifier ensemble is constructed based on feature decomposition in the proposed Ant System-based Feature Set Partitioning (ASFSP) algorithm. A disjoint feature set decomposition is performed based on the original training set. No feature in the original training set is eliminated. Figure 4 shows the architecture of the proposed Ant System-based feature set partitioning.

It can be seen that it is composed of two interrelated parts, namely the Ant system and classifier ensemble. The Ant System aims to produce feature partition while the classifier ensemble is to evaluate the feature partition. 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

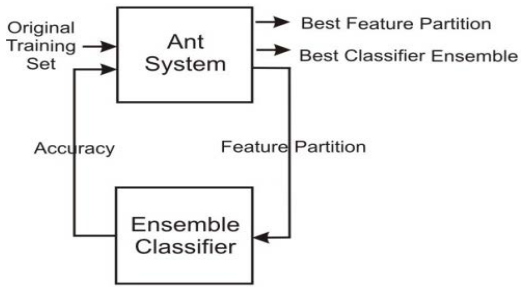

Fig. 4: The architecture of the proposed ASFSP algorithm

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 flowchart for classifier ensemble which incorporates the feature set partitioning based on Ant System is provided in Fig. 5. The main steps of this algorithm are as follows.

Input original training set: The required inputs are features in the original training set.

Generate graph problem: Generate a graph problem based on features in the original training set where each node will present a unique feature subset.

Initialization: Initialize the input parameter value, pheromone trail value and the number of artificial ants.

Feature partition generation: In the first iteration, 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.

Feature partition evaluation: Split the original training set to training set and validation set. Train a set of classifiers using the generated feature partition via the training set. Evaluate the classifier ensemble using validation set. The 10-fold cross validation method can be used to test accuracy while the majority voting technique can be used as a combiner. The best partition will be formed if the classification accuracy reaches 100% or the maximum iteration limit has been reached.

Criteria evaluation: If any criterion is not fulfilled, update pheromone and generate new ants. The whole process is repeated until the best partition is formed.

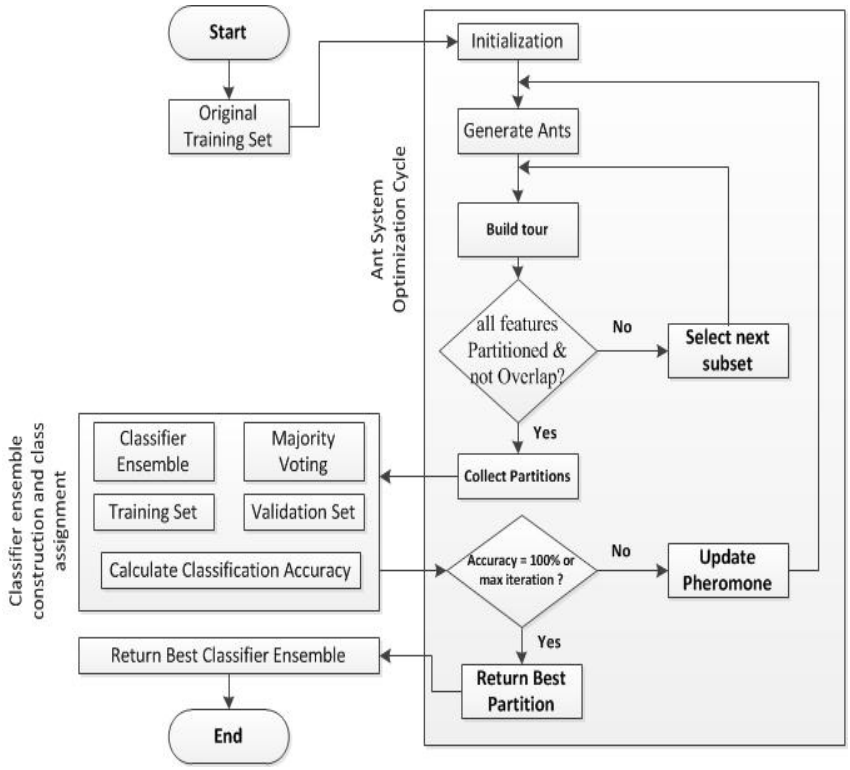

Fig. 5: Flowchart of ASFSP algorithm

The use of ant system for feature set partitioning in this algorithm will be able to find the optimal number of feature set partitions to train the classifiers. The number of classifiers can be automatically determined by the number of feature set partitions that have been formed. Furthermore, the use of this algorithm will determine if a single classifier or classifier ensemble approach is suitable to be used. presents the pseudocode of the proposed algorithm.

Algorithm 1: Generic pseudo code proposed algorithm

```
%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]=subtitutes(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{l,:}}]%return best partition
End
```

Table 1: Summary of datasets

Datasets	No. of instances	No. of classes	No. of features	Features types
Haberman	306	2	3	Integer
Iris	150	3	4	Real
Lenses	24	3	4	Categorical
Liver	345	2	6	Categorical, integer, real
E. coli	336	8	7	Real
Pima Indians Diabetes	768	2	8	Integer, real
Tic-Tac-Toe	958	2	9	Categorical
Glass	214	6	9	Real
Breast Cancer (Wisconsin)	699	2	9	Categorical

The MATLAB code is used to implement this algorithm. Nearest Mean Classifier (NMC), Naive Bayes Classifier (NBC), k-Nearest Neighbor (k-NN) and Linear Discriminant Analysis (LDA) are chosen as base classifier. The 10-fold cross validation method is used to test the classifiers. A collection of nine datasets taken from the UCI repository are used in the experiment. The datasets involved are Haberman, Iris, Lenses, Liver, *E. coli*, Pima Indians Diabetes, Tic-Tac-Toe, Glass and breast cancer (Wisconsin). The datasets consist of various numbers and types of features. A summary of the datasets is presented in Table 1.

RESULTS AND DISCUSSION

Experiments were conducted to test the proposed ensemble construction algorithm that integrates the ant system for feature set partition. Ten experiments were performed to evaluate the performance of single NMC, NBC, k-NN and LDA. Table 2-5 show the average andstandard deviation of single NMC, NBC, k-NN and LDA accuracy for the nine datasets. Based on the results, it can be seen that small deviations of the classification accuracies were obtained which showed that the experiments have been performed correctly.

The experiments were also conducted to test the ASFSP in constructing homogeneous NMC, NBC, k-NN and LDA ensembles. The prediction class label of the testing set is obtained by aggregating predictions using a combiner. The most commonly used majority voting combiner has been used in the experiments. The average accuracies of the newly constructed homogeneous ensembles by the proposed algorithm are compared with the average accuracies of constructed homogeneous ensembles by RSM. RSM is chosen because it is a widely used method in constructing ensemble classifiers (Serpen and Pathical, 2009; Li *et al*., 2013). The configuration parameter of RSM is described as follows. The number of feature subsets is set to four. The number of features for each subset is selected randomly with replacement. The configuration parameter of ASFSP is described as follows. The number of ants is set equal to the number of nodes (m = n), α =1, β =1 and p = 0.5. The comparison between RSM and ASFSP in constructing homogeneous ensembles and the detailed information of the features are as depicted in Table 6-9.

Table 2: Classification accuracy of single NMC

Experiment No.	Haberman	Iris	Lenses	Liver	*E. coli*	Pima	Tic-Tac-Toe	Glass	Breast cancer
1	69.93	92.00	70.83	55.07	81.85	63.02	62.94	44.86	96.49
2	69.93	91.33	66.67	55.94	81.85	62.89	62.73	44.86	96.49
3	70.59	92.67	66.67	54.20	81.55	63.41	64.20	44.39	96.49
4	69.61	92.67	62.50	54.49	81.25	63.15	62.73	43.46	96.49
5	70.92	92.00	70.83	53.91	80.65	63.67	64.20	44.86	96.49
6	69.93	92.00	62.50	56.23	81.55	62.89	62.63	44.39	96.49
7	69.93	92.67	66.67	55.94	81.55	63.41	63.15	43.93	96.49
8	71.24	92.00	62.50	55.36	81.85	63.28	62.84	43.93	96.49
9	67.32	91.33	62.50	55.07	82.14	63.67	62.84	44.86	96.49
10	70.26	92.00	66.67	55.65	81.25	63.54	63.67	42.06	96.49
Average	69.97	92.07	65.83	55.19	81.55	63.29	63.19	44.16	96.49
SD	1.06	0.49	3.29	0.79	0.42	0.30	0.61	0.89	0.00

Table 3: Classification accuracy of single NBC

Experiment No.	Haberman	Iris	Lenses	Liver	*E. coli*	Pima	Tic-Tac-Toe	Glass	Breast cancer
1	74.84	95.33	62.50	55.36	74.20	75.39	72.13	72.87	96.05
2	74.51	95.33	62.50	55.36	74.60	75.26	73.38	72.87	96.05
3	74.18	95.33	62.50	56.23	75.01	75.78	72.44	73.33	96.05
4	74.51	94.67	62.50	55.65	74.98	75.78	72.65	72.21	96.19
5	74.18	96.00	62.50	55.65	73.99	75.65	72.13	73.02	96.34
6	74.51	95.33	62.50	54.78	75.03	76.04	71.82	72.56	96.19
7	74.51	96.00	62.50	55.65	74.44	75.52	72.44	73.33	95.90
8	73.86	95.33	62.50	55.36	74.87	75.91	72.65	73.33	96.19
9	75.16	96.00	62.50	55.07	74.77	76.43	72.86	73.33	96.19
10	74.84	95.33	62.50	55.07	75.00	75.91	72.86	73.33	96.19
Average	74.51	95.47	62.50	55.42	74.69	75.77	72.54	73.02	96.13
SD	0.38	0.42	0.00	0.41	0.37	0.34	0.45	0.39	0.12

Table 4: Classification accuracy of single k-NN

Experiment No.	Haberman	Iris	Lenses	Liver	*E. coli*	Pima	Tic-Tac-Toe	Glass	Breast cancer
1	69.28	96.00	75.00	62.03	80.95	67,84	75.57	73.36	95.46
2	67.65	96.00	79.17	61.74	81.85	66.93	75.89	71.50	95.75
3	66.99	94.67	75.00	62.61	80.06	65.49	75.26	73.36	95.61
4	65.03	96.00	79.17	62.90	80.95	67.97	74.63	73.83	95.61
5	66.01	95.33	75.00	60.29	81.85	67.32	76.51	72.43	95.46
6	66.99	95.33	79.17	63.48	80.95	67.71	75.89	71.50	95.61
7	64.71	95.33	83.33	62.90	82.14	66.67	76.10	71.96	96.19
8	67.97	96.00	75.00	63.19	81.25	67.84	73.49	72.90	95.90
9	67.32	96.00	79.17	61.16	80.95	67.84	75.37	73.36	96.05
10	66.34	96.00	79.17	62.90	80.95	68.10	76.41	72.90	96.19
Average	66.83	95.67	77.92	62.32	81.19	67.37	75.51	72.71	95.78
SD	1.37	0.47	2.81	1.00	0.61	0.81	0.45	0.83	0.28

Table 5: Classification accuracy of single LDA

Experiment No.	Haberman	Iris	Lenses	Liver	*E. coli*	Pima	Tic-Tac-Toe	Glass	Breast cancer
1	73.86	97.33	83.33	62.90	72.33	75.78	65.55	59.35	96.05
2	73.20	97.33	87.50	62.90	72.98	74.48	65.14	60.28	96.19
3	73.53	97.33	87.50	60.87	73.00	75.65	66.28	57.94	96.19
4	74.84	97.33	83.33	63.77	72.78	75.26	65.34	58.41	96.19
5	72.88	97.33	87.50	62.61	74.01	75.78	65.24	58.41	96.19
6	73.86	97.33	87.50	63.48	74.20	74.87	65.45	58.41	96.05
7	74.18	97.33	87.50	62.32	72.12	75.52	65.55	59.35	96.19
8	74.51	97.33	87.50	62.32	72.67	75.52	65.24	57.94	96.19
9	73.53	97.33	83.33	60.00	72.98	75.39	65.87	57.01	96.34
10	72.88	97.33	87.50	62.32	72.00	75.13	66.49	61.22	96.19
Average	73.73	97.33	86.25	62.35	72.91	75.34	65.62	58.83	96.18
SD	0.66	0.21	2.01	1.14	0.73	0.42	0.46	1.24	0.08

Table 6: Comparison of RSM and ASFSP in constructing homogeneous NMC ensembles

	Classifier ensemble construction					
	RSM			ASFSP		
Dataset	Average of accuracy (%)	Feature subset	No. of classifier	Average of accuracy (%)	Feature partition	No. 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
E. coli	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 7: Comparison of RSM and ASFSP in constructing homogeneous NBC ensembles

	Classifier ensemble construction					
	RSM			ASFSP		
Dataset	Average of accuracy (%)	Feature subset	No. of classifier	Average of accuracy (%)	Feature partition	No 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
E. coli	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 8: Comparison of RSM and ASFSP in constructing homogeneous k-NN ensembles

	Classifier ensemble construction					
	RSM			ASFSP		
Dataset	Average of accuracy (%)	Feature subset	No. of classifier	Average of accuracy (%)	Feature partition	No. of classifier
Haberman	67.91	[1 2][3][1 3][2 3]	4	72.75	[1 3][2]	2
Iris	93.40	[3 4][2 4][1 4][1 2]	4	95.93	[1 2 3 4]	1
Lenses	62.50	[1][2 4][1 2 4][3 4]	4	79.17	[1 2 3 4]	1
Liver	60.06	[3][1 3 5 6][5 6][2 3 4 5]	4	64.16	[1 4 6][3 5][2]	3
E. coli	81.19	[1 2 3 4 5 6 7][1 5][4 5 6][1 6]	4	81.19	[1 2 3 4 5 6 7]	1
Pima	70.59	[1 2 3 4 5 7][3][2 3 4 5 6 7 8][2 6 8]	4	71.01	[1 3 4 7][5 6 8][2]	3
Tic-Tac-Toe	75.70	[1 2 3 5 6 7][2 6 7 9][1 2 3 4 6 7 8 9][5 6]	4	75.73	[1 2 3 4 5 6 7 8 9]	1
Glass	72.71	[4 5 6 7][1 2 3 5 7 8 9][1 6 9][1 2 4 5 6 7]	4	72.90	[1 2 3 4 5 6 7 8 9]	1
Breast Cancer	97.23	[1 2 3 6][1 3 6][5 8 9][1 3 8 9]	4	97.60	[1 2 4 7 9][3 5][6][8]	4

Table 6 shows the comparison results of RSM and ASFSP in constructing homogeneous NMC ensembles. Most of the datasets have been successfully partitioned, for instance, Haberman, Iris, Liver, Pima, Tic-Tac-Toe,

Table 9: Comparison of RSM and ASFSP in constructing homogeneous LDA ensembles

	Classifier ensemble construction					
	RSM			ASFSP		
Dataset	Average of accuracy (%)	Feature subset	No. of classifier	Average of accuracy (%)	Feature partition	No. of classifier
Haberman	73.76	[3][2 3][2][1 3]	4	74.84	[1][2 3]	2
Iris	95.73	[1 3 4][4][1 4][3]	4	98.00	[1 2 3 4]	1
Lenses	80.42	[1 4][1][1 2 3 4][1 3 4]	4	86.67	[1 2 3 4]	1
Liver	62.44	[1 2 6][3 4 5 6][1 2 3 5 6][2 3]	4	63.86	[1 3 4 6][2][5]	3
E. coli	73.28	[1 4 5 7][2 4 6][1 2 3 4][2 5 6]	4	75.93	[1 3 5][4 6][2 7]	3
Pima	74.94	[1 2 7][2 3 8][1 2 3 7][4 7 8]	4	76.03	[1 2 3 4 5 6 7 8]	1
Tic-Tac-Toe	65.63	[1 3 6 7][4 6 7 9][3 4 5 8][1 2 7 8 9]	4	73.02	[2 4 5 6 8][1][3][7][9]	5
Glass	59.21	[1 2 3 4 7 9][1 2 3 5 6 7][[7 9][3 5 9]	4	62.52	[2 3 5 7][4 8 9][1 6]	3
Breast cancer	96.21	[4 5 7 8][2 3 4 6 7][6 7 8][3 4 5 6 7]	4	97.20	[2 4 8][7 9][3][1 5 6]	4

Table 10: The result summary in constructing homogeneous NMC ensembles

Dataset	Single approach	RSM	ASFSP
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
E. coli	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 11: The result summary in constructing homogeneous NBC ensembles

Dataset	Single approach	RSM	ASFSP
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
E. coli	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

Table 12: The Result Summary in Constructing Homogeneous k-NN Ensembles

Dataset	Single approach	RSM	ASFSP
Haberman	66.83	67.91	72.75
Iris	95.67	93.40	95.93
Lenses	77.92	62.50	79.17
Liver	62.32	60.06	64.16
E. coli	81.19	81.19	81.19
Pima	67.37	70.59	71.01
Tic-Tac-Toe	75.51	75.70	75.73
Glass	72.71	72.71	72.90
Breast Cancer	95.78	97.23	97.60

glass and breast cancer. On two datasets which are lenses and *E. coli*, ASFSP does not partition the features. This means that this algorithm chooses the single classifier, instead of an ensemble classifier.

Table 7 shows the comparison results of RSM and ASFSP in constructing homogeneous NBC ensembles. On all datasets, ASFSP 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 ASFSP, is less than RSM. The same number of partition on breast cancer dataset constructed by ASFSP and RSM has been obtained but the partitions are of different form. The lenses dataset has also successfully been partitioned with comparable classification accuracy.

Table 8 shows the comparison of RSM and ASFSP in constructing homogeneous ensembles when k-NN is used as base classifier. Improvement classification accuracy frequently appears when k-NN is used as base classifier. It can be clearly seen in haberman, liver, pima and breast cancer datasets. For other datasets, the worst accuracy of ASFSP is par with the accuracy of RSM. In these results, the parameter k in k-NN is set to one (Table 9).

Table 9 comparison of RSM and ASFSP in Constructing Homogeneous LDA Ensembles. 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. There is a possibility that several features are not selected. However, if ASFSP 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 ASFSP is automatically determined.

The usage of RSM provides lower accuracy, even though the number of classifiers is bigger. The usage of ASFSP can easily determine the optimal number of classifiers. ASFSP has successfully delivered better classification results with an optimal number of classifiers. The summary of results for single approach, RSM and ASFSP in constructing classifier ensembles are shown in Table 10-13.

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 algorithm. 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.

Table 13: The result summary in constructing homogeneous LDA ensembles

Dataset	Single approach	RSM	ASFSP
Haberman	73.73	73.76	74.84
Iris	97.33	95.73	98.00
Lenses	86.25	80.42	86.67
Liver	62.35	62.44	63.86
E. coli	72.91	73.28	75.93
Pima	75.34	74.94	76.03
Tic-Tac-Toe	65.62	65.63	73.02
Glass	58.83	59.21	62.52
Breast cancer	96.18	96.21	97.20

Table 14: Result of Proposed Algorithm compared with Previous Algorithms

Dataset	(1)	(2)	(3)	(4)	(5)	(6)	ASFSP
Haberman	66.83	-	-	-	71.89	-	72.75
Iris	95.67	97.33	-	-	95.20	96.70	96.34
E. coli	81.19	-	-	-	82.79	-	81.91
Glass	72.71	-	-	-	74.23	72.50	73.54
Pima	67.37	72.68	-	71.90	-	75.70	71.22
Breast cancer	95.78	96.35	97.92	97.50	-	-	98.09

The performance of the proposed algorithm has been compared to six other algorithms. k-NN classifier is used as a base classifier for comparison. Haberman, iris, *E. coli*, glass, pima and breast cancer from the UCI repository are chosen because of the availability of results from previous studies, in which k-NN was also used as base classifier. The performance of the proposed algorithm is evaluated by comparing the results to: single classifier approach, dynamic weighted voting (Valdovinos and Sanchez, 2009), an improved k-NN classification using Genetic Algorithm (Suguna and Thanushkodi, 2010), Simultaneous metaheuristic feature selection (Tahir and Smith, 2010), Weighted k-NN ensemble method (Hamzeloo *et al*., 2012) and Direct boosting algorithm (Neo and Ventura, 2012). Table 14 presents the comparison of results of these algorithms. It can be seen that the performance of the proposed algorithm gives the best classification accuracies as compared to the other algorithms on habermann and breast cancer datasets. The performance of the proposed algorithm is at par with other algorithms for the remaining seven datasets.

CONCLUSION

A feature set partitioning algorithm based on ant system for accurate and diverse classifier ensemble construction has been presented. Classifier ensembles were trained on different feature partitions to induce diversity. The utilization of ant system was to produce the optimal feature partition. The proposed algorithm was evaluated on several benchmark datasets. The results show that the implementation of this algorithm in constructing several homogeneous ensembles outperforms single approach and RSM. The performance of the proposed algorithm has also been compared to several other algorithms. In general, the proposed algorithm gives good classification results and is comparable to previous algorithms. The use of the proposed algorithm can also provide the optimal number of ensemble members. Moreover, the use of this algorithm will determine if a single or ensemble approach is suitable to be used. The novel contribution of this algorithm can be useful as a guide to produce various combinations of classifiers.

RECOMMENDATIONS

Future work would be to apply this algorithm for heterogeneous classifier ensemble construction. Testing the ability of this algorithm to overcome the curse of dimensionality can also be considered for future work.

REFERENCES

Abreu, M.C. and A.M. Canuto, 2007. Evaluating the influence of the choice of the ensemble members in some fuzzy combination methods. Proceedings of the 2007 International Joint Conference on Neural Networks, August 12-17, 2007, IEEE, Orlando, Florida, ISBN: 978-1-4244-1379-9, pp: 448-453.

Ahn, H., H. Moon, M.J. Fazzari, N. Lim and J.J. Chen *et al*., 2007. Classification by ensembles from random partitions of high-dimensional data. Comput. Stat. Data Anal., 51: 6166-6179.

Chung-Fu, C., 2008. Reconfiguration and capacitor placement for loss reduction of distribution systems by ant colony search Algorithm. IEEE Power Syst. Transact., 23: 1747-1755.

Crawford, B., C. Castro and E. Monfroy, 2014. A new ACO transition rule for set partitioning and covering problems. Proceedings of the International Conference of Soft Computing and Pattern Recognition 2009, December 4-7, 2009, IEEE, Malacca, Malaysia, pp: 426-429.

Crawford, B., R. Soto, E. Monfroy, C. Castro and W. Palma *et al*., 2013. A hybrid soft computing approach for subset problems. Math. Prob. Eng., 2013: 1-12.

Dorigo, M. and C. Blum, 2005. Ant colony optimization theory: A survey. Theor. Comput. Sci., 344: 243-278.

Hamzeloo, S., H. Shahparast and M.Z. Jahromi, 2012. A novel weighted nearest neighbor ensemble classifier. Proceedings of the 2012 16th CSI International Symposium on Artificial Intelligence and Signal Processing (AISP), May 2-3, 2012, IEEE, Shiraz, Iran, ISBN: 978-1-4673-1478-7, pp: 413-416.

Hansen, L.K. and P. Salamon, 1990. Neural network ensembles. IEEE. Trans. Pattern Anal. Mach. Intell., 12: 993-1001.

Hernandez-Lobato, D., G. Martinez-Munoz and A. Suarez, 2013. How large should ensembles of classifiers be? Pattern Recognit., 46: 1323-1336.

Jevtic, A., D. Andina, A. Jaimes, J. Gomez and M. Jamshidi, 2010. Unmanned aerial vehicle route optimization using ant system algorithm. Proceedings of the 2010 5th International Conference on System of Systems Engineering (SoSE), June 22-24, 2010, IEEE, Loughborough, England, ISBN: 978-1-4244-8197-2, pp: 1-6.

Kang, H.J. and D. Doermann, 2005. Selection of classifiers for the construction of multiple classifier systems. Proceedings of the Eighth International Conference on Document Analysis and Recognition (ICDAR'05), August 29-September 1, 2005, IEEE, Seoul, South Korea, ISBN: 0-7695-2420-6, pp: 1194-1198.

Kittler, J., M. Hatef, R.P. Duin and J. Matas, 1998. On combining classifiers. IEEE. Trans. Pattern Anal. Mach. Intell., 20: 226-239.

Kuncheva, L.I. and C.J. Whitaker, 2003. Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy. Mach. Learn., 51: 181-207.

Li, H., G. Wen, Z. Yu and T. Zhou, 2013. Random subspace evidence classifier. Neurocomputing, 110: 62-69.

Maimon, O. and L. Rokach, 2005. Decomposition Methodology for Knowledge Discovery and Data Mining. In: Data Mining and Knowledge Discovery Handbook. Maimon, O. and L. Rokach (Eds.). Springer-Verlag, Berlin, Germany, ISBN: 978-0-387-24435-8, pp: 981.

Margoosian, A. and J. Abouei, 2013. Ensemble-based classifiers for cancer classification using human tumor microarray data. Proceedings of the 2013 21st Iranian Conference on Electrical Engineering (ICEE), May 14-16, 2013, IEEE, Mashhad, Iran, pp: 1-6.

Neo, T.K.C. and D. Ventura, 2012. A direct boosting algorithm for the k-nearest neighbor classifier via local warping of the distance metric. Pattern Recognit. Lett., 33: 92-102.

Parvin, H., H. Alizadeh and B.B. Minaei, 2009. A new method for constructing classifier ensembles. J. Digital Content Technol. Appl., 3: 62-66.

Ponti, J.M.P., 2011. Combining classifiers: From the creation of ensembles to the decision fusion. Proceedings of the 2011 24th SIBGRAPI Conference on Graphics, Patterns and Images Tutorials (SIBGRAPI-T), August 28-30, 2011, IEEE, Alagoas, Brazil, ISBN: 978-1-4577-1627-0, pp: 1-10.

Ribeiro, R. and F. Enembreck, 2013. A sociologically inspired heuristic for optimization algorithms: A case study on ant systems. Expert Syst. Appl., 40: 1814-1826.

Rokach, L., 2008. Genetic algorithm-based feature set partitioning for classification problems. Patt. Recognit., 41: 1676-1700.

Rokach, L., 2010. Pattern Classification Using Ensemble Methods. World Scientific, Singapore, ISBN: 9789814271073, Pages: 225.

Roli, F., 2009. Multiple Classifier System. In: Encyclopedia of Biometrics. Li, S.Z. and A.K. Jain (Eds.). Springer Science and Business Media, New York, USA., ISBN: 978-0-387-73002-8, pp: 981.

Serpen, G. and S. Pathical, 2009. Classification in high-dimensional feature spaces: Random subsample ensemble. Proceedings of the International Conference on Machine Learning and Applications ICMLA'09, December 13-15, 2009, IEEE, Miami Beach, Florida, ISBN: 978-0-7695-3926-3, pp: 740-745.

Shang, F. and Y. Wang, 2010. An ant system optimization QoS routing algorithm for wireless sensor networks. Proceedings of the 2010 Third International Workshop on Advanced Computational Intelligence (IWACI), August 25-27, 2010, IEEE, Suzhou, China, ISBN: 978-1-4244-6334-3, pp: 339-343.

Srimani, P.K. and M.S. Koti, 2013. Medical diagnosis using ensemble classifiers-a novel machine-learning approach. J. Adv. Comput., 1: 9-27.

Su, C.T., C.F. Chang and J.P. Chiou, 2005. Distribution network reconfiguration for loss reduction by ant colony search algorithm. Electr. Power Syst. Res., 75: 190-199.

Suguna, N. and K. Thanushkodi, 2010. An improved k-nearest neighbor classification using genetic algorithm. Int. J. Comput. Sci. Issues, 7: 18-21.

Tahir, M.A. and J. Smith, 2010. Creating diverse nearest-neighbour ensembles using simultaneous metaheuristic feature selection. Pattern Recognit. Lett., 31: 1470-1480.

Valdovinos, R.M. and J.S. Sanchez, 2009. Combining Multiple Classifiers with Dynamic Weighted Voting. In: Hybrid Artificial Intelligence Systems. Corchado, E., W. Xindong, O. Erkki, A. Herrero and B. Bruno (Eds.). Springer Berlin Heidelberg, Berlin, Germany, ISBN: 978-3-642-02318-7, pp: 510-516.

Wanas, N.M. and M.S. Kamel, 2002. Weighted combination of neural network ensembles. Proceedings of the 2002 International Joint Conference on Neural Networks IJCNN'02, May 12-17, 2002, IEEE, Honolulu, Hawaii, ISBN: 0-7803-7278-6, pp: 1748-1752.

Wu, C.H. and W.B. Liang, 2011. Emotion recognition of affective speech based on multiple classifiers using acoustic-prosodic information and semantic labels. IEEE. Trans. Affective Comput., 2: 10-21.