Filomat 32:5 (2018), 1917–1930
https://doi.org/10.2298/FIL1805917S

Published by Faculty of Sciences and Mathematics,
University of Niš, Serbia
Available at: http://www.pmf.ni.ac.rs/filomat

Rough Set based Ensemble Learning Algorithm for Agricultural Data Classification

Lei Shi[a], Qiguo Duan[b], Juanjuan Zhang[a], Lei Xi[a], Hongbo Qiao[a], Xinming Ma[a]

[a]*College of Information and Management Science, HeNan Agricultural University/ Collaborative Innovation Center of Henan Grain Crops, Zhengzhou 450002, China*
[b]*Zhengzhou Commodity Exchange, Zhengzhou 450008, China*

Abstract. Agricultural data classification attracts more and more attention in the research area of intelligent agriculture. As a kind of important machine learning methods, ensemble learning uses multiple base classifiers to deal with classification problems. The rough set theory is a powerful mathematical approach to process unclear and uncertain data. In this paper, a rough set based ensemble learning algorithm is proposed to classify the agricultural data effectively and efficiently. An experimental comparison of different algorithms is conducted on four agricultural datasets. The results of experiment indicate that the proposed algorithm improves performance obviously.

1. Introduction

Agricultural data classification has become an important technique in the community of intelligent agriculture because of the massive volume of agricultural data available. Ensemble learning is a popular algorithm which construct multiple base classifiers and then aggregate these classifiers to predict unknown instances [1]. Ensemble learning is able to obviously increase generalization performance of the unstable classifier, and thus has become a very popular issue in data mining and machine learning fields. Recently, the application of ensemble learning in agricultural data classification is receiving a lot of attention. For instance, ensemble learning technique is used to enhance SVM to detect rice parcels and yield [2]. However, existing ensemble techniques often tend to combine all the trained base classifiers into a unnecessarily large ensembles. Two problem with these ensemble learning techniques are that they require more random access memory (RAM) to store all based classifiers, and also need more time to complete classification. Rough set theory is an approach to process vague and uncertain information, and it provides efficient tools to find hidden patterns in data [3]. Since it was first presented in the 80's of the 20th century, rough set theory has extracted much attention from various research fields such as credit rating [4], text classification [5], image segmentation [6], etc. Attribute reduction is an important method in rough set theory, it does not need any preliminary information to obtain minimal sets of attributes [7].

An ensemble learning algorithm based on rough set is presented to classify the agricultural data effectively and efficiently in this paper. A set of trained individual classifiers is created by applying bootstrap

2010 *Mathematics Subject Classification*. Primary O159.
Keywords. Rough set; Agricultural data; Ensemble learning.
Received: 17 October 2017; Accepted: 7 December 2017
Communicated by Hari M. Srivastava
Email addresses: sleicn@126.com (Lei Shi), dqgcn@126.com (Qiguo Duan), zhangjuan_2003@126.com (Juanjuan Zhang), hnaustu@126.com (Lei Xi), qiaohb@126.com (Hongbo Qiao), xinmingma@126.com (Xinming Ma)

samples firstly and a decision table is constructed. Then a subset of the individual classifiers is selected and used for aggregation by using rough set theory. An experimental evaluation is conducted on the public agricultural datasets. The results of experiment show that the proposed algorithm achieves significant performance improvement compared with popular ensemble methods.

The rest of this paper is organized as follows. Section 2 reports related work. Section 3 reviews the basic concepts of ensemble learning technique and rough set theory. Section 4 detailedly introduces the proposed rough set based ensemble learning algorithm. Section 5 reports experimental evaluation and discussion. Section 6 summarizes the conclusions and presents directions of future work.

2. Related work

Recently, some classification techniques in the machine learning community are applied in agriculture field to solve the classification problem. In optimal site-specific application of herbicides, the Self-Organizing Map neural network is applied in the classification of crops and weeds for reducing input costs and environmental impact [8]. One of the major problem affecting the quality of peppers and tomatoes is cuticle crack. Artificial neural network is introduced to accurately classify cuticle crack in both peppers and tomatoes produced in greenhouses [9]. Because frost may have serious consequences on crop production, artificial neural network, Naive Bayes and k-nearest neighbor (kNN) classifiers have been applied to develop an empirical prediction system for frost protection of fruits and vegetables [10]. For obtaining a map of the crops on the scene, Support Vector Machine (SVM) is used to develop crop cover classifiers for crop classification using hyperspectral images [11]. The automatic classification of birds by their vocalization is useful in some practical applications, such as avoiding collisions between aircraft and birds. In the studies presented in [12], the bird sound signals have been represented by the descriptive parameters model, and then SVM classifier with a Gaussian kernel is applied in the classification for the bird sound and other different sounds. In early crop growth stage, stresses detection is helpful for the application of site-specific remedies. In [13], SVM method is applied to classify hyperspectral data for identification of nitrogen stresses and weed in cornfield. Because the coughs detection has important effect to check the presence of diseases, artificial neural network is used as the classification method for distinguishing cough sounds of pigs from other sounds, like background noise, metal clang and grunts to reveal the possibility of an epidemic [14]. Because the identification of soybean development stages is useful to make appropriate and timely management decisions, an artificial neural network classifier is developed for the classification of soybean flowering and physiological maturity [15]. To provide better crop management information, the classification and regression trees (CART) technique is applied in the classification of spectral data from the corn fields [16].

3. Preliminaries

3.1. Ensemble learning

As a famous supervised learning algorithm, ensemble learning combines several machine learning classifiers into one predictive model in order to decrease variance and improve predictions [1]. Previous work has shown that combination of multiple classification models generally increases predictive performance.

Recently, ensemble learning has received considerable attention and has been successfully used in many fields, such as biomedical prediction task [17], protein structural classification [18] and agricultural data classification [19]. The ensemble learning framework is shown in Fig. 1.

Firstly, subset O_k, $k = 1, \cdots, T$ of the original training set O is formed. Secondly, one base classifier h_k is trained by using subset O_k. For each new instance, the predictions of all base classifiers are aggregated in some strategy $h^* = F(h_1, h_2, \cdots, h_T)$ for generating the classification result of the ensemble. Two of the most popular ensemble learning algorithms are Bagging and Boosting, which we briefly introduce next.

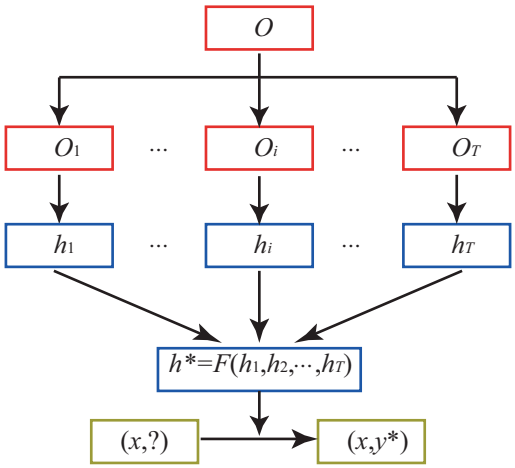

Figure 1: Framework of a classifier ensemble

3.1.1. Bagging

Bagging, one of the earliest and most intuitive ensemble algorithms, has been successfully applied in constructing ensembles of unstable classifiers [20]. Bootstrap sampling of the original data is used to generate different training data subsets. Then, base classifier is trained by using each training subset and their individual prediction results are aggregated to form a final ensemble decision.

Given an original training dataset S of size n and the integer parameter T which specify the number of iterations, then T new training sets $S_1, S_2, \cdots, S_T$ are generated by sampling from S uniformly and with replacement. A classifier C_i is trained by using S_i and then T classifiers are aggregated to construct the final classifier C. The classification label of instance x is assigned to the category classified by most number of the base classifiers, i.e., $C_1, C_2, \cdots, C_T$. The Bagging method is shown as Algorithm 1 in detail.

Algorithm 1. The Bagging ensemble algorithm
Input: dataset S={$(x_1, y_1), \cdots, (x_N, y_N)$}, number of iterations T;
Output: Bagging ensemble classifier.
(1) For $t = 1$ to T
(2) Generate a training set S_t by using bootstrap sample from S;
(3) Construct a classifier C_t on S_t;
(4) Output $C(x) = argmax_y(\sum_{t=1}^{T} I(C_t(x) = y))$.
where I is a function such that $I(true) = 1, I(false) = 0$.

3.1.2. Boosting

The Boosting is one of the most popular and successful implementations of ensemble methods [21]. As a two-step approach, Boosting uses subsets of the original data to produce a series of performing models firstly, and then combines them together using a particular cost function such as majority voting. Unlike bagging, the subset creation of Boosting is not random and depends upon the performance of the previous models. Boosting assigns a weight for each instance. When a classifier is trained from the training set, the weights for misclassified instances are increased. The instances with higher weight have more influences on the next classifier learned. Boosting is in fact a family of algorithms since there are many variants. AdaBoost is a popular boosting algorithm and is shown as Algorithm 2.

Algorithm 2. The AdaBoost ensemble algorithm
Input: dataset S, number of iterations T;
Output: AdaBoost classifier.
(1) Init data weights $d_n^{(1)} = \frac{1}{N}$ for all $n = 1, \cdots, N$;
(2) For $t = 1$ to T
(3) Train classifier with respect to the weighted instance set $\{S, d^{(t)}\}$ and obtain hypothesis: $h_t : x \to \{-1, +1\}$,

i.e., $h_t = L(S, d^{(t)})$;
(4) Compute the weighted training error ε_t of h_t: $\varepsilon_t = \sum_{n=1}^{N} d_n^{(t)} I(y_n \neq h_t(x_n))$;
(5) Set $\alpha_t = \frac{1}{2} log \frac{1-\varepsilon_t}{\varepsilon_t}$;
(6) Update the data weights $d_n^{(t+1)} = \frac{d_n^{(t)} exp\{-\alpha_t y_n h_t(x_n)\}}{Z_t}$,
where Z_t is a normalization constant, such that $\sum_{n=1}^{N} d_n^{(t+1)} = 1$;
(7) Break if $\varepsilon = 0$ or $\varepsilon_t \geq \frac{1}{2}$ and set $T = t - 1$;
(8) Output $f_T(x) = \sum_{t=1}^{T} \frac{\alpha_t}{\sum_{r=1}^{T} \alpha_r} h_t(x)$.

3.2. Rough set theory

As an efficient mathematical theory, rough set can handle imprecise or vague information [3]. Some important notions of rough set theory are introduced in this section.

Definition 1. Information system. An information system is defined as a pair as follows:

$$I = (\mathbb{U}, \mathbb{A}) \tag{1}$$

where $\mathbb{U}$ is a set of instances, and $\mathbb{A}$ is a set of attributes.

Definition 2. Decision table. In rough set theory, decision table is defined as a pair as follows:

$$I = (\mathbb{U}, \mathbb{C} \cup \mathbb{D}) \tag{2}$$

where $\mathbb{C}$ is a set of conditional attributes, and $\mathbb{D}$ is a set of decision attributes. For instance, Table 1 shows a dataset which has four conditional attributes (a, b, c, d), one decision attribute (e), and six objects $(x_1, x_2, x_3, x_4, x_5, x_6)$.

Definition 3. Indiscernibility relation. Objects may be indiscernible because available information is limited. For any $P \subseteq \mathbb{A}$, an associated equivalence relation $IND(P)$ on $\mathbb{U}$ is defined as follows:

$$IND(P) = \{(x, y) \in \mathbb{U}^2 | \forall a \in P, a(x) = a(y)\} \tag{3}$$

where $a(x)$ indicates the value on attribute a of object x. If $(x, y) \in IND(P)$, x and y are indiscernible by attributes from P. In rough set theory, the indiscernibility relation is one of the most important concepts. The equivalence classes of the P-indiscernibility relation are defined as $[x]_P$, $x \in \mathbb{U}$.

Table 1: An example dataset

$x \in \mathbb{U}$	a	b	c	d	e
x_1	1	0	0	1	1
x_2	1	0	0	0	2
x_3	0	0	1	2	2
x_4	0	1	2	1	1
x_5	2	2	1	0	1
x_6	0	1	2	2	0

$\mathbb{U}/IND(P)$ is generated by $IND(P)$, and it is computed as follows:

$$\mathbb{U}/IND(P) = \otimes\{\mathbb{U}/IND(a) | a \in P\} \tag{4}$$

where

$$A \otimes B = \{X \cap Y : \forall X \in A, \forall Y \in B, X \cap Y \neq \emptyset\} \tag{5}$$

According to the definition of $\mathbb{U}/IND(P)$ and data from illustrative example in Table 1, if $P = \{b, c\}$, then objects x_1 and x_2 are indiscernible. Similarly, objects x_4 and x_6 are indiscernible too. Therefore, $IND(P)$ generates the partition of $\mathbb{U}$ as follows:

$\mathbb{U}/IND(P) = \{\{x_1, x_2\}, \{x_3\}, \{x_4, x_6\}, \{x_5\}\}$.

Definition 4. Lower approximation and upper approximation. Let $I = (\mathbb{U}, \mathbb{A})$ be an information system, $X \subseteq \mathbb{U}$ and $P \subseteq \mathbb{A}$. The lower approximation $\underline{P}(X)$ and upper approximation $\overline{P}(X)$ of X with respect to P are defined as follows:

$$\underline{P}X = \{x \in \mathbb{U} | [x]_P \subseteq X\} \tag{6}$$

$$\overline{P}X = \{x \in \mathbb{U} | [x]_P \cap X \neq \emptyset\} \tag{7}$$

The lower approximation of a set X is the set of objects of $\mathbb{U}$, which can be certainly classified as X with respect to P. The upper approximation of a set X is the set of objects of $\mathbb{U}$, which can be possibly classified as X with respect to P. Suppose $X = \{x_3, x_4\}$ with the previous example, it is noted that $\underline{P}X = \{x_3\}$ and $\overline{P}X = \{x_3, x_4, x_6\}$.

A diagram of a rough set X within the lower approximation and upper approximation is shown in Fig. 2.

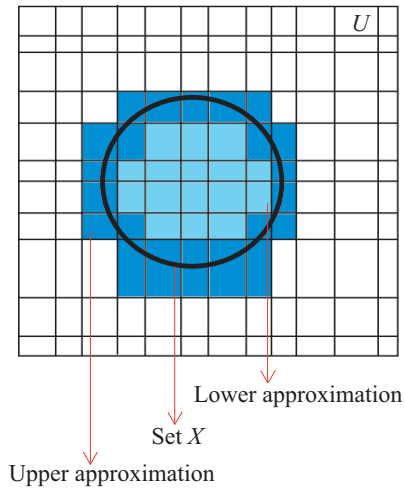

Figure 2: Lower and upper approximation

Definition 5. Positive region, negative region and boundary region. Let P and Q be subsets of $\mathbb{A}$, then the positive region, negative region and boundary region are defined as follows:

$$POS_P(Q) = \bigcup_{X \in \mathbb{U}/Q} \underline{P}X \tag{8}$$

$$NEG_P(Q) = \mathbb{U} - \bigcup_{X \in \mathbb{U}/Q} \overline{P}X \tag{9}$$

$$BND_P(Q) = \bigcup_{X \in \mathbb{U}/Q} \overline{P}X - \bigcup_{X \in \mathbb{U}/Q} \underline{P}X \tag{10}$$

The positive region, $POS_P(Q)$, is the set of all objects in the universe $\mathbb{U}$ that certainly belong to classes of $\mathbb{U}/Q$. The negative region, $NEG_P(Q)$, is the set of objects that cannot be classified to classes of $\mathbb{U}/Q$ by using attributes of P. The boundary region, $BND_P(Q)$, is the set of objects that can possibly belong to classes of $\mathbb{U}/Q$. X is definable with respect to P if $BND_P X = \emptyset$, otherwise X is rough with respect to P. For instance, let $P = \{b, c\}$ and $Q = \{e\}$, then $POS_P(Q) = \{x_3, x_5\}$, $NEG_P(Q) = \emptyset$ and $BND_P(Q) = \{x_1, x_2, x_4, x_6\}$.

If values of attributes of P uniquely determine values of attributes of Q, then the set of attributes Q depends totally on the set of attributes P, denoted $P \Rightarrow Q$. The definition of dependency is given as following.

Definition 6. Dependency degree. When Q and P be subsets of $\mathbb{A}$, we will say that Q depends on P to a degree $k(0 \leq k \leq 1)$, denoted $P \Rightarrow_k Q$, if

$$k = \gamma_P(Q) = \frac{|POS_P(Q)|}{|\mathbb{U}|} \tag{11}$$

where $|\mathbb{U}|$ is the cardinality of set $\mathbb{U}$. If $k = 1$, we say that Q depends totally on P, and if $0 < k < 1$, we say that Q depends partially on P. If $k = 0$ then we say that Q does not depend on P [22]. For instance, let $P = \{b, c\}$ and $Q = \{e\}$, then the dependency degree of attribute set Q from the attribute set P is $\gamma_P(Q) = \frac{|POS_{\{b,c\}}(\{e\})|}{|\mathbb{U}|} = \frac{1}{3}$.

Definition 7. Reduct. Let $I = (\mathbb{U}, \mathbb{A})$ be an information system, $P \subseteq \mathbb{A}$ and $a \in P$. Attribute a is redundant in P if $\mathbb{U}/P = \mathbb{U}/(P - a)$, otherwise a is indispensable in P. P is independent if every $a \in P$ is indispensable in P. P is a reduct of $\mathbb{A}$ if $\mathbb{U}/P = \mathbb{U}/\mathbb{A}$ and P is independent. Thus, a reduct of attributes is sufficient subset to represent the category structure.

Attribute reduction selects a minimal subset of attributes with the same information of classification, and it has become a much more important issue in rough set theory. QuickReduct algorithm [22] is a popular technique to discover reduct. The QuickReduct algorithm is shown as follows.

Algorithm 3. QuickReduct algorithm
Input: The set of entire conditional attributes $\mathbb{C}$, the set of entire decision attributes $\mathbb{D}$;
Output: the reduct Q of $\mathbb{C}$.
(1)$Q = \emptyset$;
(2)Do
(3) $H = Q$;
(4) $\forall x \in (\mathbb{C} - Q)$
(5) If $\gamma_{Q\cup\{x\}}(\mathbb{D}) > \gamma_H(\mathbb{D})$
(6) $H = Q \cup \{x\}$;
(7) $Q = H$;
(8)Until $\gamma_Q(\mathbb{D}) == \gamma_{\mathbb{C}}(\mathbb{D})$
(9)Return Q.

4. The proposed algorithm

A rough set based ensemble learning algorithm is introduced for effective classification of the agricultural data in this section. A schematic diagram of the proposed algorithm is illustrated in Fig. 3.

The proposed algorithm consists of six stages, i.e., partitioning stage, bootstraping stage, training stage, constructing stage, selecting stage and combining stage. In the first stage, the original agricultural dataset S is randomly partitioned into two subsets S_1 and S_2 for building the proposed algorithm. Then, a bootstrap sampling method is applied to construct different training datasets from the dataset S_1 for improving the diversity of the classifier in the second stage. The multiple base classifiers are trained by using training subsets obtained from the previous stage in the third stage. All base classifiers are applied to classify instances in the S_2 dataset and accordingly classification labels are obtained in the fourth stage. The real labels and classification labels of instances in the S_2 dataset are then used to create the decision table I. In the decision table I, the real class label of the instances are used as the value of decision attribute, and the classification label of the mth instance given by nth base classifier is indicated as the element I_{mn}. The reduct of the decision table I is obtained by using QuickReduct algorithm in the fifth stage. In the reduct, the appropriate ensemble members are selected from the multiple trained base classifiers. Finally the selective base classifiers are combined in terms of majority voting strategy to classify the new unlabeled instance in the sixth stage. The proposed method is shown in Algorithm 4.

Algorithm 4. The proposed rough set based ensemble learning algorithm
Inputs:original agricultural dataset S, trials T;

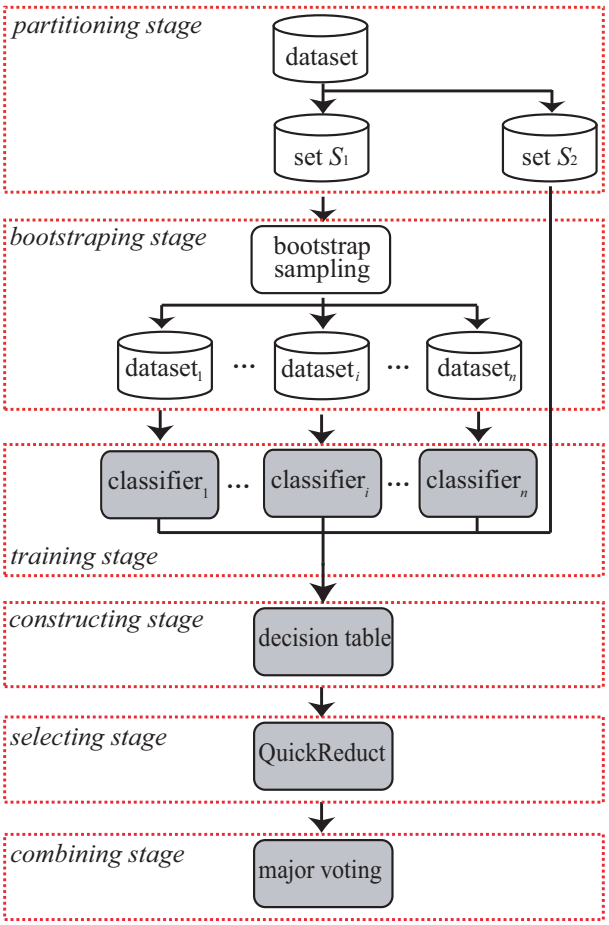

Figure 3: General architecture of the proposed algorithm

Outputs:ensemble classifier.
(1) Partition the S into two sets S_1 and S_2;
(2) Generate T training datasets from the dataset S_1;
(2.1) For $t = 1$ to T
(2.2) Build set S_t via bootstrap sample from S_1;
(2.3) End For
(3) Train T base classifiers;
(3.1) For $t = 1$ to T
(3.2) Train the base classifier C_t on S_t;
(3.3) End For
(4) Create the decision table;
(4.1) Apply T base classifiers to classify the instances in the S_2;
(4.2) Create a decision table in which the classification class labels of instances in the S_2 are used as conditional attributes and real class labels are used as decision attribute;
(5) Select the base classifiers;
(5.1) Apply QuickReduct algorithm in the decision table to obtain a reduct;
(6) Create the final ensemble classifier;
(6.1)The base classifiers in reduct are used to classify instance and then majority voting strategy is used to aggregate the corresponding predictions.

5. Experimental results and discussions

The experimental evaluation is conducted on four public agricultural datasets in this section. First, the dataset descriptions and evaluation measures are given. Next, experimental settings and results of the experiments on the four datasets are presented.

5.1. Experimental datasets

In the experiment, four agricultural datasets [23], i.e., the white clover dataset, the grub damage dataset, the eucalyptus dataset and the pasture dataset, are employed as benchmarks. The white clover dataset consists of 63 instances and 32 attributes. The grub damage dataset consists of 155 instances and 9 attributes. The eucalyptus dataset consists of 736 instances and 20 attributes. The pasture dataset consists of 36 instances and 23 attributes.

5.2. Performance metrics

Two popular metric techniques F_1 and AUC are used as performance measures to demonstrate the effectiveness of different methods. The decision of classifier is indicated as a confusion matrix in Table 2 [24].

Table 2: A confusion matrix

Class C		Result of classifier	
		belong	Not belong
Real classification	Belong	TP	FN
	Not belong	FP	TN

The four categories of confusion matrix are described as follows:
*T*rue *P*ositives (***TP***): the number of positive instances correctly classified as positive.
*F*alse *P*ositives (***FP***): the number of negative instances incorrectly classified as positive.
*T*rue *N*egatives (***TN***): the number of negative instances correctly classified as negative.
*F*alse *N*egatives (***FN***): the number of positive instances incorrectly classified as negative.
TP and TN are the observations that are correctly predicted. FP and FN occur when actual class contradicts with the predicted class. According to these four parameters, *precision*, *recall*, and F_1 measure can be defined. *Precision* is the rate of correctly classified positive instances to the total results which are classified as positive. *Recall* is the rate of correctly classified positive instances to the all observations in actual class. F_1 measure is the weighted average of *precision* and *recall*.

$$precision = \frac{TP}{TP + FP} \tag{12}$$

$$recall = \frac{TP}{TP + FN} \tag{13}$$

$$F_1\ measure = 2 \times \frac{precision \times recall}{precision + recall} \tag{14}$$

Receiver operating characteristic (ROC) curve is a useful tool that presents the diagnostic ability of classifier without regard to class distribution or error cost [25]. The ROC curve is generated according to the trade-off between the true positive rate (TPR) and the false positive rate (FPR) for various cut-off values of a parameter. The TPR and FPR are defined as follows:

$$True\ positive\ rate = \frac{TP}{TP + FN} \tag{15}$$

$$False\ positive\ rate = \frac{FP}{FP + TN} \tag{16}$$

ROC curve is not convenient for making a direct comparison of different classifiers.

The Area Under the receiver operating characteristic Curve (AUC) uses simple scores between 0 and 1 to measure the performance, and it is intuitive and more sensitive for comparison of the classifiers' performances [26]. Typical AUC are shown in Fig. 4 where the two ROC curves represent classifiers A and B, respectively.

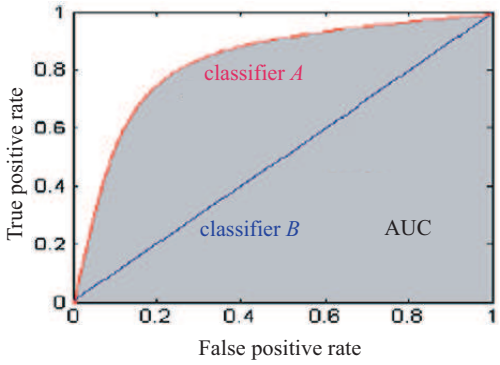

Figure 4: A schematic view of AUC

AUC score of classifier A is larger than AUC score of classifier B, and thus performance of classifier A is better than performance of classifier B. Recently, AUC is increasingly employed in the study of classification systems. In the experiment, F_1 and AUC are used to evaluate the classification performance of different methods.

5.3. Experimental settings

In the experiments, two popular ensemble algorithms Bagging and AdaBoost are implemented and the classification results of them are included as benchmark. Bagging, AdaBoost and the proposed algorithm use decision tree C4.5 algorithm [27] as base classifier. The experimental result of C4.5 algorithm is also included for comparison. Bagging and Boosting change the ensemble sizes from ten to fifty respectively to construct a various range of parameter.

The 10-fold cross validation is conducted on the datasets to avoid over-fitting. Each dataset is first partitioned into ten equal-sized parts. Any nine of the ten parts is selected to perform training. The remaining tenth acts as an independent holdout test set for evaluation. For ensuring the effectiveness of the comparison among different algorithms in experiments, the algorithms are repeated three times with different random seeds of each training and test process in 10-fold cross validation. Thus, thirty experiment results for each technique can be obtained and the mean of the thirty performances is used as final results for the C4.5, Bagging and Boosting methods respectively. Moreover, the process of partitioning the training set into S_1 and S_2 is done for the proposed algorithm. For each dataset, α% of the training dataset instances is selected randomly to construct the set S_1 in the experiment. The rest of the training set instances is used as the set S_2. According to Algorithm 4, the set S_1 and S_2 are employed to training the proposed rough set based ensemble learning algorithm. Then, the test dataset is applied to evaluate the classification performance. To minimize potential biases of the randomized sampling process, the partition process are repeated three times in each training and test process. Thus, ninety experiment results of the proposed

algorithm are obtained and the mean of those results is adopted to measure the classification performance. For designing a variation of parameter settings, the range of α values are investigated from 10% to 90% with a step size of 10% in the experiments. The tuning results indicate that when α equals 70%, the proposed method obtains the best performance. Thus, α is set as 70% for our subsequent experiments.

5.4. Results and discussion

Table 3 indicates experimental results when a subset is selected from 20 trained C4.5 decision trees by the proposed algorithm. In the Table 3, the sixth column shows average number of base decision trees using by our proposed method. The ensemble sizes of Boosting and Bagging equal 20.

Table 3: F_1 of the different methods on datasets (full ensemble size = 20)

Dataset	C4.5	Bagging	Boosting	Proposed	Number
white clover	0.578	0.701	0.685	0.713	5.5
grub damage	0.310	0.325	0.330	0.339	4.3
eucalyptus	0.592	0.641	0.642	0.656	7.6
pasture	0.682	0.823	0.674	0.871	8.1

As shown in Table 3, the proposed algorithm achieves considerable performance improvement over other algorithms on all datasets. For instance, the F_1 of the proposed method on white clover dataset equals 71.3%, which beats C4.5 by approximately 13.5%, Bagging by approximately 1.2%, and Boosting by approximately 2.8%. The ensemble size of the proposed mehtod is about only 27.5% (5.5/20).

Table 4: F_1 of the different methods on datasets (full ensemble size = 50)

Dataset	C4.5	Bagging	Boosting	Proposed	Number
white clover	0.578	0.672	0.684	0.692	5.9
grub damage	0.310	0.331	0.326	0.346	7.3
eucalyptus	0.592	0.632	0.655	0.668	6.2
pasture	0.682	0.825	0.707	0.856	11.2

Table 4 gives the results in terms of F_1 value when a subset is selected from 50 trained C4.5 decision trees by the proposed algorithm, and the ensemble sizes of Boosting and Bagging equal 50. the proposed method also achieves considerable performance improvement over other algorithms on all datasets. For instance, the F_1 of the proposed method on eucalyptus dataset is 66.8%, which beats C4.5 by approximately 7.6%, Bagging by approximately 3.6%, and Boosting by approximately 1.3%. The ensemble size of the proposed algorithm is about only 12.4% (6.2/50).

The performance curve of C4.5, Bagging, Boosting and the proposed method at different ensemble size on each dataset are shown in figures 5, 6, 7 and 8, respectively. From the four figures, it can be seen that the proposed algorithm beats C4.5, Bagging and Boosting on all four agricultural datasets.

AUC is widely used for judging the performance and discriminative ability of classifiers. Table 5 gives the AUC results of C4.5, Bagging, Boosting and the proposed method. The ensemble sizes of Boosting and Bagging equal 10. However, the proposed method selects a subset from 10 trained C4.5 decision trees to construct ensemble.

Table 5: AUC of the different methods on datasets (full ensemble size = 10)

Dataset	C4.5	Bagging	Boosting	Proposed
white clover	0.686	0.705	0.708	0.716
grub damage	0.592	0.640	0.604	0.672
eucalyptus	0.842	0.892	0.869	0.896
pasture	0.841	0.884	0.888	0.915

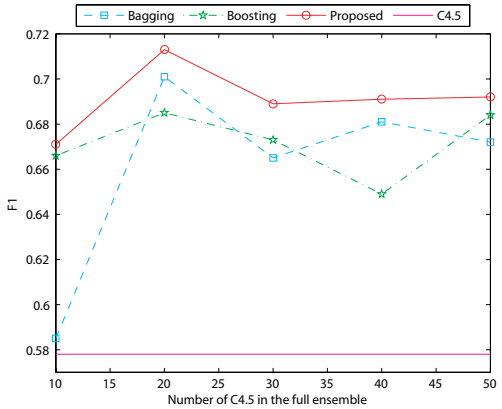

Figure 5: F_1 comparison of different algorithms on the white clover dataset

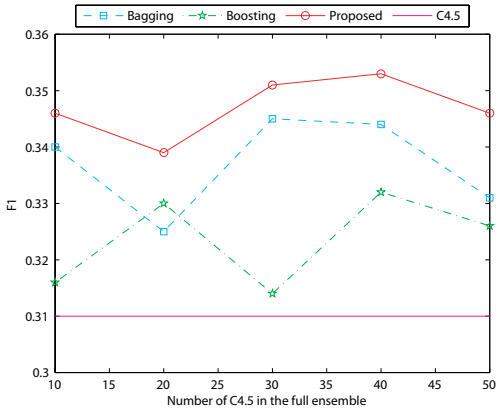

Figure 6: F_1 comparison of different algorithms on the grub damage dataset

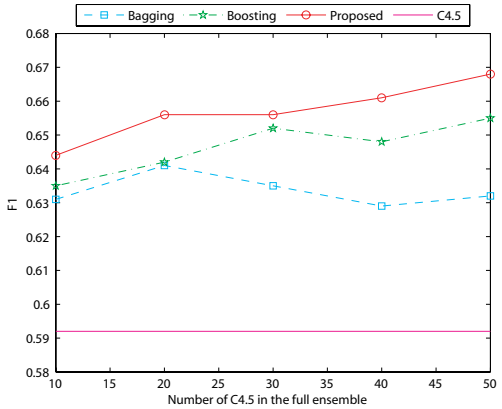

Figure 7: F_1 comparison of different algorithms on the eucalyptus dataset

Table 5 indicates that the AUC values of the proposed algorithm are improved significantly compared with the other algorithms on white clover, grub damage, eucalyptus and pasture datasets, respectively. For instance, the AUC value of the proposed algorithm on white clover dataset is 71.6%, which beats C4.5 by approximately 3.0%, Bagging by approximately 1.1%, and Boosting by approximately 0.8% . The number

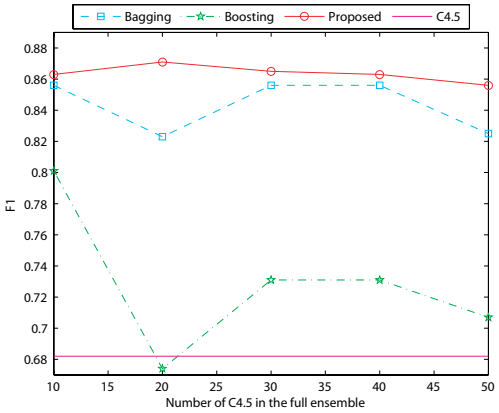

Figure 8: F_1 comparison of different algorithms on the pasture dataset

of base classifiers using by the Bagging and Boosting equal 10. However, the number of base classifiers using by the proposed algorithm is about 4.

Table 6: AUC of the different methods on datasets (full ensemble size = 20)

Dataset	C4.5	Bagging	Boosting	Proposed
white clover	0.686	0.694	0.715	0.730
grub damage	0.592	0.648	0.623	0.675
eucalyptus	0.842	0.904	0.879	0.909
pasture	0.841	0.885	0.893	0.906

Table 7: AUC of the different methods on datasets (full ensemble size = 30)

Dataset	C4.5	Bagging	Boosting	Proposed
white clover	0.686	0.703	0.727	0.739
grub damage	0.592	0.632	0.623	0.681
eucalyptus	0.842	0.904	0.888	0.912
pasture	0.841	0.894	0.911	0.918

Table 8: AUC of the different methods on datasets (full ensemble size = 40)

Dataset	C4.5	Bagging	Boosting	Proposed
white clover	0.686	0.713	0.729	0.732
grub damage	0.592	0.629	0.624	0.682
eucalyptus	0.842	0.904	0.884	0.918
pasture	0.841	0.898	0.898	0.929

Table 9: AUC of the different methods on datasets (full ensemble size = 50)

Dataset	C4.5	Bagging	Boosting	Proposed
white clover	0.686	0.742	0.734	0.745
grub damage	0.592	0.633	0.627	0.689
eucalyptus	0.842	0.906	0.889	0.913
pasture	0.841	0.895	0.899	0.935

When the ensemble sizes equal 20, 30, 40 and 50 on each dataset respectively, Tables 6, 7, 8 and 9 display the AUC value comparison of C4.5, Bagging, Boosting and the proposed method.

According to the experimental results presented in Tables 6, 7, 8 and 9, The AUC results of the proposed algorithm also achieve effective improvement compared with other methods.

Table 10: Comparison of different algorithms in efficiency on the eucalyptus dataset

Cost resource	C4.5	Bagging	Boosting	Proposed
Time	0.9	22.4	24.6	3.1
Memory	7	27	29	12

Table 10 summarizes the comparison of the efficiency on the eucalyptus dataset when a subset is selected from 50 trained C4.5 decision trees by the proposed algorithm, and the ensemble sizes of both Bagging and Boosting equal 50. All the algorithms are written in Java programming language. Times and memory used by different methods to classify the test set are shown in Table 10.

According to Table 10, it is shown that the proposed algorithm is much better than Bagging and Boosting methods, and it is comparable to the C4.5 algorithm. The memory that the proposed method used is 12 MB, which is about only 44.4% and 41.4% respectively of those costed by Bagging and Boosting techniques. The computational time that the proposed method used is 3.1 second, which is about only 13.8% and 12.6% of those used by Bagging and Boosting techniques respectively. On white clover, grub damage and pasture datasets, the proposed algorithm improves the memory consumption and computational cost too. Because of space limitations, those results are not listed here.

The following observations are given according to experimental results. The proposed method outperforms the C4.5, Bagging and Boosting classifiers in terms of F_1 and AUC on the four agricultural datasets. The proposed method is more efficient than Bagging and Boosting, and is comparable to the C4.5 algorithm.

6. Conclusion

Since the massive volume of agricultural data is available, agricultural data classification has become critical. A rough set based ensemble learning algorithm is presented in this paper, and the experimental results indicate that the proposed method obtains improvement obviously. In future research, the combination of the proposed algorithm and feature selection methods will be research so that the performance of agricultural data classification can be further improved.

References

[1] Rokach, L.: Ensemble-based classifiers. Artificial Intelligence Review 33(1-2),1-39(2010)
[2] Chen, W.C., Chen, M.H., Chou T.Y.: The Study of AdaBoost Algorithm to Enhance Support Vector Machine Application on Urban Rice Land Classification. In: Proceedings of the asian conference on remote sensing (2009)
[3] Mahajan, P., Kandwal, R., Vijay, R.: Rough set approach in machine learning: a review. International Journal of Computer Applications. 56(10),1-13(2012)
[4] Liu, Z.F., Pan, S.: Fuzzy-rough instance selection combined with effective classifiers in credit scoring. Neural Processing Letters.(4),1-10(2017)
[5] Shi, L., Zhang, Y.M., Zhao, J.Y.: Combining rough set and centroid classifier for text classification. Journal of Information and Computational Science. 7(1),79-84(2010)
[6] Ji, Z., Huang, Y., Xia, Y., Zheng, Y.: A robust modified gaussian mixture model with rough set for image segmentation. Neurocomputing. 266(29),550-565(2017)
[7] Bedi, J., Upadhyaya, S.: Attribute reduction approaches based on rough set theory: a analytical review. International Journal for Scientific Research & Development. (3)3,2321-0613(2015)
[8] Moshou, D., Vrindts, E., Ketelaere, B.D., et al: A neural network based plant classifier. Computers and Electronics in Agriculture. 31,5-16(2001)
[9] Ehret, D.L., Hill, B.D., Rawortha, D.A., et al: Artificial neural network modelling to predict cuticle cracking in greenhouse peppers and tomatoes. Computers and Electronics in Agriculture. 6,108-116(2008)
[10] Verdes, P.F., Granitto, P.M., Navone, H.D., Cecatto, H.A.: Frost prediction with machine learning techniques. In: Proceedings of the VIth Argentine Congress on Computer Science.1423-1433(2000)

[11] Camps-Valls, G., Gomez-Chova, L., Calpe-Maravilla, J.,et al: Support Vector Machines for Crop Classification using Hyperspectral Data. Lectures Notes on Computer Science. 2652,134-141(2003)
[12] Fagerlund, S.: Bird species recognition using support vector machines. EURASIP Journal on Applied Signal Processing. 64-64(2007)
[13] Karimi, Y., Prasher, S.O., Patel, R.M., Kim, S.H.: Application of support vector machine technology for weed and nitrogen stress detection in corn. Computers and Electronics in Agriculture. 51,99-109(2006)
[14] Chedad, A., Moshou, D., Aerts, J.M., et al: Recognition System for Pig Cough based on Probabilistic Neural Networks. Journal of Agricultural Enginnering Research. 79(4),449-457(2001)
[15] Elizondo, D.A., McClendon, R.W., Hoogenboom, G.: Neural Network Models for Predicting Flowering and Physiological Maturity of Soybean. Transactions of the American Society of Agricultural Engineers. 37(3),981-988(1994)
[16] Waheed, T., Bonnell, R.B., Prasher, S.O., et al: Measuring performance in precision agriculture: CART-A decision tree approach. Agricultural water management. 84,173-185(2006)
[17] Shi, L., Xi, L., Ma, X.M., et al: A novel ensemble algorithm for biomedical classification based on Ant Colony Optimization. Applied Soft Computing. 11(8),5674-5683(2011)
[18] Homayouni, H., Mansoori, E.G.: A novel density-based ensemble learning algorithm with application to protein structural classification. Intelligent Data Analysis. 21(1),167-179(2017)
[19] Shi, L., Duan, Q.G., Si, H.P., et al: Approach of hybrid soft computing for agricultural data classification. International Journal of Agricultural and Biological Engineering. 8(6),54-61(2015)
[20] Breiman, L.: Bagging predictors. Machine Learning. 24,123-140(1996)
[21] Freund, Y., Shapire, R.E.: A decision-theoretic generalization of online learning and an application to boosting. J. Comput. Syst. Sci. 55,119-139(1997)
[22] Jensen, R., Shen, Q.: A Rough Set-Aided System for Sorting WWW Bookmarks. Web Intelligence. 95-105(2001)
[23] http://www.cs.waikato.ac.nz/ml/weka/indexdatasets.html(2017)
[24] Yang, Y.: An evaluation of statistical approaches to text categorization. Information Retrieval. 1,69-90(1999)
[25] Hanley, J.A.: Receiver operating characteristic (ROC) methodology: the state of the art. Critical Reviews in Diagnostic Imaging. 29,307-35(1989)
[26] Fawcett, T.: An introduction to ROC analysis. Pattern Recognition Letters. 27,861-874(2006)
[27] Quinlan, J.: C4.5:Programs for Machine Learning. San Matteo, CA:Morgan Kaufmann Publishers (1993)