Approach of hybrid soft computing for agricultural data classification

Shi Lei[1], Duan Qiguo[3], Si Haiping[1], Qiao Hongbo[1], Zhang Juanjuan[1], Ma Xinming[1,2*]

(1. *College of Information and Management Science, Henan Agricultural University, Zhengzhou 450002, China*;
2. *Collaborative Innovation Center of Henan Grain Crops, Zhengzhou 450002, China*;
3. *Zhengzhou Commodity Exchange, Zhengzhou 450008, China*)

Abstract: Soft computing is an important computational paradigm, and it provides the capability of flexible information processing to solve real world problems. Agricultural data classification is one of the important applications of computing technologies in agriculture, and it has become a hot topic because of the enormous growth of agricultural data available. Support vector machine is a powerful soft computing technique and it realizes the idea of structural risk minimization principle to find a partition hyperplane that can satisfy the class requirement. Rough set theory is another famous soft computing technique to deal with vague and uncertain data. Ensemble learning is an effective method to learn multiple learners and combine their decisions for achieving much higher prediction accuracy. In this study, the support vector machine, rough set and ensemble learning were incorporated to construct a hybrid soft computing approach to classify the agricultural data. An experimental evaluation of different methods was conducted on public agricultural datasets. The experimental results indicated that the proposed algorithm improves the performance of classification effectively.

Keywords: agricultural data, soft computing, rough set, support vector machine, ensemble learning, classification
DOI: 10.3965/j.ijabe.20150806.1312

Citation: Shi L, Duan Q G, Si H P, Qiao H B, Zhang J J, Ma X M. Approach of hybrid soft computing for agricultural data classification. Int J Agric & Biol Eng, 2015; 8(6): 54－61.

1 Introduction

Soft computing is a consortium of methodologies to construct a computationally intelligent system which provides flexible information processing capability to reason and learn in an environment of uncertainty and imprecision[1]. Soft computing was first coined by Zadeh[2] and became a formal computer science area in early 1990's. Typically, soft computing consists of several computing paradigms, including artificial neural network[3], rough set theory[4], support vector machine (SVM)[5], genetic algorithms[6], simulated annealing[7], etc. The rapid adoption of information technology in agriculture production systems has resulted in the expanding amount of agricultural data being created and gathered. Because classifying agricultural data of an interesting class is an important step to mine the valuable information of agricultural data, agricultural data classification has attracted much attention in the agriculture and computer science communities. Recently, many soft computing approaches have been applied to the agricultural field[8,9]. Karimi et al.[10] employed SVM method for classifying hyperspectral data to identify weed and nitrogen stresses, which can aid in effective application of remedies to timely interventions in early growth stage. Chedad et al.[11] collected pigs

Received date: 2014-10-26 **Accepted date:** 2015-10-10
Biographies: Shi Lei, PhD, Associate Professor, Research interest: agricultural informatization, Email: sleicn@126.com; **Duan Qiguo,** PhD, Research interest: data mining, Email: dqgcn@126.com; **Si Haiping,** PhD, Associate Professor, Research interest: agricultural informatization, Email: pingsss@126.com; **Qiao Hongbo,** PhD, Associate Professor, Research interest: precision agriculture, Email: qiaohb@126.com; **Zhang Juanjuan,** PhD, Associate Professor, Research interest: agricultural informatization, Email: zhangjuan_2003@126.com;
***Corresponding author: Ma Xinming,** PhD, Professor, Research interest: agricultural informatization, College of Information and Management Science, Henan Agricultural University, Zhengzhou 450002. Tel: +86-371-56990030, Email: xinmingma@126.com.

sounds and using artificial neural network to distinguish cough sounds from other sounds for the detection of diseases. Schatzki et al.[12] used artificial neural network method as the classification technique to analyze the X-ray photographs of the fruit for estimating the probability that the fruit contains watercores.

SVM is a powerful soft computing technique, and it offers state-of-the-art performance. It realizes the Vapnik-Chervonenkis (VC) theory and the idea of structural risk minimization principle to constitute an objective function and then find a partition hyperplane that can satisfy the class requirement. An important advantage of SVM is that it can be analyzed theoretically using concepts from computational learning theory. As a kind of structural risk minimization based learning algorithms, SVM has better generalization abilities compared to other traditional empirical risk minimization based learning algorithms. Recently, it has also been successfully applied to a number of real world problems such as credit scoring[13], damage detection[14] and the classification of biomedical data[15]. Rough set theory, introduced by Z. Pawlak in the early 1980s, is an extension of set theory for study of the intelligent systems characterized by insufficient and incomplete information. As an effective tool to handle imprecise and inconsistent information in real world problems, rough set theory has become an important soft computing method. In the past decades, rough set theory has gained more and more attention and been applied in the areas of text classification[16], feature selection[17,18], etc. Ensemble learning is a machine learning paradigm that trains a set of component classifiers and then combines their predictions to output the final decision[19]. In recent years, ensemble learning is rapidly growing and enjoying a lot of attention from many various domains due to their potential to greatly increase the classification performance. As one of the most popular research directions of machine learning in the past years, ensemble learning has already been successfully applied to many areas, such as microarray data analysis[20] and face recognition[21].

With the growth of agricultural data available today, the technique for agricultural data classification with characteristics of both high accuracy and generalization is increasingly appreciated. Such characteristics indicate that a single method, such as SVM or rough set, is difficult to be implemented in such an aim because each method has its own limitations and weaknesses. In this study, a novel hybrid soft computing approach is proposed to efficiently deal with the classification problem of the agricultural data. The proposed approach incorporates the soft computing techniques, i.e., SVM, rough set and ensemble learning into the classifying system. Rough set theory is used as a preprocessor to reduce the redundant attributes in the data because of its reliability to reduce independent attributes with no information loss. Then, the obtained significant independent attributes are used as inputs and a collection of SVM classifiers is constructed to build a diverse ensemble. The experiments and evaluations of the different methods are performed with several publicly available agricultural datasets. The experimental results demonstrate that the proposed method not only reduced redundant attributes but also improve the classification performance of agricultural data effectively.

2 Methods

2.1 Rough set

Knowledge representation in rough set is used via information system, which is denoted as a 4-tuple $S=(U, A, V, f)$, where U is a nonempty finite set of objects, A is a nonempty finite set of attributes, V is the domain of attribute value and $f:U\times A\rightarrow V$ is an information function which associates a unique value of each attribute with every object belonging to U, such that $f(x,a)\in V_a$, for any $a\in A$ and $x\in U$. The information system is also called a decision table if $A=C\cup D$ and $C\cap D=\varnothing$, where, C is the set of condition attributes, D is the set of decision attributes.

With any $P\subseteq A$, there is an associated indiscernibility relation $IND(P)$:

$$IND(P)=\{(x,y)\in U\times U:\forall a\in P, f(x,a)=f(y,a)\}$$

where, $f(x,a)$ denotes the value of attribute a of object x. If $(x,y)\in IND(P)$, x and y are said to be indiscernible with respect to P. Indiscernibility relation is one of the most significant aspects of rough set theory.

The partition of U, determined by $IND(P)$ is denoted by $U/IND(P)$, and can be calculated as follows:

$$U / IND(P) = \otimes\{U / IND(\{a\}) : a \in P\}$$

where, $A \otimes B = \{X \cap Y : \forall X \in A, \forall Y \in B, X \cap Y \neq \varnothing\}$.

The equivalence classes of the P-indiscernibility relation are denoted by $[x]_P$ as follows:

$$[x]_P = \{y : (x, y) \in IND(P), y \in U\}$$

Given information system $S=(U, A, V, f)$, for any subset $X \subseteq U$ and attribute set $P \subseteq A$, X could be approximated by the lower approximation and upper approximation. The lower approximation of X is the set of objects of U that are surely in X, defined as:

$$\underline{P}X = \{x \in U : [x]_P \subseteq X\}$$

The upper approximation of X is the set of objects of U that are possibly in X, defined as:

$$\overline{P}X = \{x \in U : [x]_P \cap X \neq \varnothing\}$$

Then, the boundary region of X is defined as:

$$BND(X) = \overline{P}(X) - \underline{P}(X)$$

The set X is said to be rough if the boundary region $BND(X) = \varnothing$, otherwise the set X is crisp. Figure 1 provides a view of a rough set X within the upper and lower approximations.

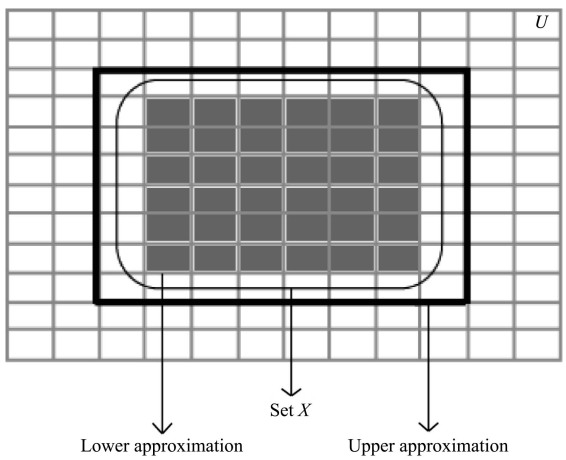

Figure 1 Rough set within the upper and lower approximations

Given any two subsets $P, Q \subseteq A$, the P-positive, P-negative and P-boundary regions of Q can be defined, respectively, as

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

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

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

The positive region is the set of objects that can be classified to classes of $U/IND(Q)$. The negative region is the set of objects that cannot be classified to classes of $U/IND(Q)$. The boundary region is the set of objects that can possibly, but not certainly, be classified in this way.

A discernibility matrix of a decision table is a symmetric $|U|\times|U|$ matrix with entries defined by

$$c_{ij} = \{a \in C \mid a(x_i) \neq a(x_j)\}, \quad i, j = 1, \ldots, |U|$$

The dependency is one of the most common measures to represent how much a set depends on another set. For any $P, Q \subseteq A$, it is said that attribute set Q depends on P with degree κ, denoted by $P \Rightarrow_\kappa Q$, where

$$\kappa = \gamma_P(Q) = \frac{|POS_P(Q)|}{|U|}$$

The dependency coefficient κ measures the degree of dependency between Q and P. If $\kappa=1$, Q depends totally on P, if $0<\kappa<1$, Q depends partially on P, and if $\kappa=0$ then Q does not depend on P. When P is a set of condition attributes and Q is the decision attribute set, $\gamma_P(Q)$ is called the quality of approximation of classification[4].

A reduct attribute set is a minimal set of attributes that provided that the object classification is the same as with the full set of attributes. For $R \subseteq C$, the set of all reducts can be defined by the following definition based on dependency degree:

$$Red = \{R \mid \gamma_R(D) = \gamma_C(D), \forall B \subset R, \gamma_B(D) \neq \gamma_C(D)\}$$

2.2 SVM

As a kind of maximum margin classifiers, SVM minimizes the training set error and maximizes the margin in order to achieve the best generalization ability and remain resistant to over fitting[5]. An illustration of the SVM for a linearly separable binary classification problem is shown in Figure 2.

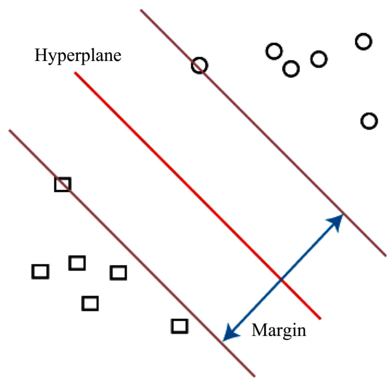

Figure 2 SVM for the linearly separable binary classification problem

Consider a binary classification task and let the training set be $\{x_i, y_i\}$, i=1,…, N, $y_i \in \{-1, +1\}$, $x_i \in R^d$, where, x_i is an input n-dimensional vector and y_i is the corresponding label of the class that the vector belongs to. The two classes of points are separated with a hyperplane given by

$$w^T x + b = 0$$

where, b is the offset from the origin; w is an n-dimensional coefficient vector which is normal to the partition hyperplane. Because the wider margin can acquire the better generalization ability, it is going to look for an optimal partition hyperplane to maximize the separating margin between the two classes of data. The task can be defined as follows.

Minimize $g(w)=\frac{1}{2}\|w\|^2$

So that:

$y_i(w^T x_i + b) \geq 1, \forall i$

The Lagrange function can be defined as:

$$L(w,b,\alpha_i) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{n} \alpha_i [y_i(w^T x_i + b) - 1]$$

where, α_i is Lagrange multipliers, hence subject to the following two conditions, i.e., $\sum_{i=1}^{n} y_i \alpha_i = 0$ and $\alpha_i \geq 0$. Then the following formula can be defined for seeking the minimum of Lagrange function.

$$H(\alpha)=\sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j$$

Further, the data points can be mapped into a Hilbert Inner Product space through a replacement for obtaining a better representation of the data:

$$x_i^T x_j \rightarrow \phi(x_i) \cdot \phi(x_j) = K(x_i, x_j)$$

where, $K(\cdot)$ is a kernel function. Then the kernel version of the formula can be given as follows.

$$H(\alpha)=\sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j K(x_i, x_j)$$

Given a kernel function, the SVM classifier can be described as follows.

$$F(x) = \mathrm{sgn}(f(x))$$

where, $f(x) = \sum_{i=1}^{n} \alpha_i y_i K(x_i, x) + b$ is the output hyperplane decision function of the SVM.

2.3 Ensemble learning

Ensemble learning is a technique to learn and retain multiple learners and combine their decisions during classification. In recent years, ensemble learning is intensively studied and has become one of the major advances in inductive learning. It is shown that an ensemble can provide significantly better classification performance compared to the best individual. Ensemble learning technique commonly comprises two key steps. The first step is to create the different models. The second step of an ensemble method is to combine the models by using some strategies, such as voting and weighted voting[19]. Compared with one single classifier, an ensemble classifier has advantages to handle a classification task and can achieve much higher prediction accuracy. In the past decades, ensemble learning has been proven to be quite versatile in a broad field of real applications such as bankruptcy prediction[22], sentiment classification[23], etc.

DECORATE (Diverse Ensemble Creation by Oppositional Relabeling of Artificial Training Examples) is one of the most famous ensemble learning methods, and it can use any strong learner to build a diverse ensemble in a fairly straightforward manner[24]. DECORATE is accomplished by adding different randomly constructed instances to the training set when building new ensemble members. These artificially constructed instances are given category labels that disagree with the decision of the current ensemble, thereby directly increasing diversity when a new learner is trained on the augmented data and added to the ensemble. It has been found that DECORATE produces highly accurate ensemble and outperforms bagging, AdaBoost and random forest[24].

3 Proposed approach

In this section, the proposed hybrid soft computing approach to classify the agricultural data is presented in detail. The proposed hybrid approach consists of two steps, i.e., attribute reducing step and ensemble constructing step. The basic framework of the proposed hybrid approach is shown in Figure 3.

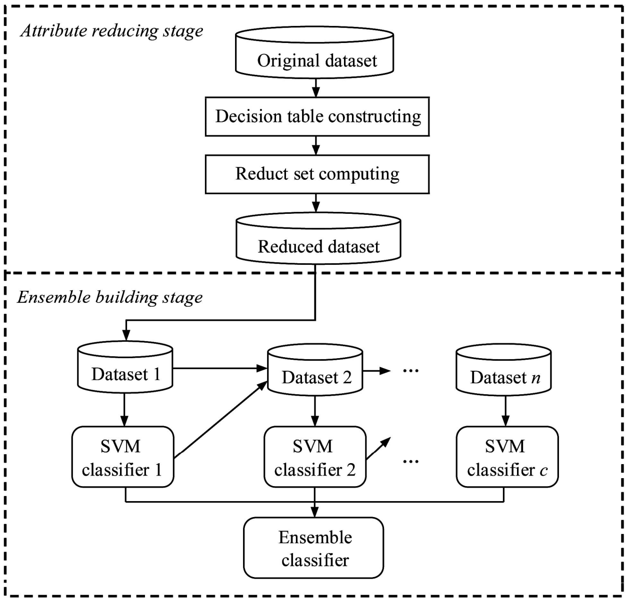

Figure 3 Conceptual framework of the proposed approach

In the attribute reducing step, the original agricultural dataset is used to construct the decision table and then the reduct set computation algorithm is employed to obtain the reduct of the decision table for removing redundant attributes from datasets[25]. The detailed algorithm is given in Algorithm 1.

Algorithm 1: Attribute reducing algorithm

Input: *TS*, original agricultural dataset

Output: reduced dataset

1) Read the agricultural training set *TS* and construct a decision table *IS*

2) Compute indiscernibility matrix $M(IS)=(C_{ij})$

3) Reduce M using absorbtion laws

4) Obtain d no-empty fields $C_1, C_2, \ldots, C_d$ of reduced M

5) Build families of sets $R_0, R_1, \ldots, R_d$ in the following way:

$R_0 = \varnothing$

For i=1 to d

$R_i = S_i \cup T_i$ where $S_i = \{R \in R_{i-1} : R \cap C_i \neq \varnothing\}$ and $T_i = (R \cup \{a\})_{a \in C_i, R \in R_{i-1} : R \cap C_i = \varnothing}$

End for

6) Remove dispensable attributes from each element of family R_d

7) Remove redundant elements from R_d

8) Obtain the reduct set $RED(IS) = R_d$

In the second stage, the proposed approach uses the reduced set as input and then constructs the ensemble by using the SVM and DECORATE. Initially, a SVM classifier is trained on the basis of the reduced agricultural training set obtained by attribute reducing step. Then, a SVM classifier is created by combining the reduced agricultural training set with some artificial data in successive iteration. The algorithm is described as Algorithm 2.

Algorithm 2: Ensemble constructing algorithm

Input: T, reduced agricultural training set consisting of N instances;

SVM, base learner;

C_{size}, desired ensemble size;

I_{max}, maximum number of iterations to build an ensemble;

R_{size}, a factor to determine number of artificial instances to generate.

Output: ensemble classifier

1) Let $i=1$ and $trials=1$

2) Provide the given training set T as the input of base learner SVM to obtain a classifier C_i

3) Initialize ensemble $C^*=\{C_i\}$

4) Compute ensemble error as

$$\varepsilon = \frac{1}{N}\sum_{i=1}^{N} I(C^*(x_i) \neq y_i)$$

5) While $i<C_{size}$ and $tirals<I_{max}$

6) Generate $R_{size}\times N$ training instances R based on the distribution of training data

7) Label instances in R with probability of class labels inversely proportional to those predicted by C^*

8) $T=T\cup R$

9) Apply base learner SVM to T to obtain a new classifier C'

10) $C^*=C^*\cup\{C'\}$

11) $T=T-R$, remove the artificial data

12) Compute training error, ε', of C^* as in step 4

13) If $\varepsilon'\leq\varepsilon$, let $i=i+1$ and set $\varepsilon=\varepsilon'$; Otherwise, remove C' from the ensemble set C^*, i.e., $C^*=C^*-\{C'\}$

14) $trials=trials+1$

15) End While

4 Experimental data and evaluation

4.1 Datasets

We performed extensive experiments on two benchmark agricultural datasets obtained from agricultural researchers in New Zealand, i.e., the white clover dataset and the grub damage dataset[26].

The white clover dataset consists of 63 instances and 32 attributes. The objective was to determine the factors to influence the persistence of white clover populations in summer dry hill land. In particular reference to the consequence of a severe summer dry period and how it impacted on the performance of three white clover cultivars.

The grub damage dataset consists of 155 instances and 9 attributes. Grass grubs are major insect pests of pasture in Canterbury. It can cause severe pasture damage and economic loss. Pastoral damage may occur periodically over wide ranging areas. Grass grub populations are often influenced by biotic factors and irrigation.

4.2 Performance measures

In the experiments, Accuracy, F_1 measure and AUC are used to measure the performance of classification. Four cases are considered as the result of classifier to the instance as shown in Table 1[27].

Table 1 Cases of the classification for one class.

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

TP (*T*rue *P*ositive): the number of instances correctly classified to that class.

TN (*T*rue *N*egative): the number of instances correctly rejected from that class.

FP (*F*alse *P*ositive): the number of instances incorrectly rejected from that class.

FN (*F*alse *N*egative): the number of instances incorrectly classified to that class.

Then, the Accuracy, Precision, Recall and F_1 measure can be defined as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F_1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

The Accuracy is used to measure the proportion of correctly classified instances. The F_1 measure is the harmonic mean of Precision and Recall and thus it is a more reliable and suitable measure.

The Area Under the receiver operating characteristic Curve (AUC) has been used widely as performance index in the data mining and machine learning communities[28]. The famous statistical meaning of AUC is that it is equivalent to the Wilconxon test of ranks. In this study, Accuracy, F_1 and AUC are used as performance measures for evaluating the performance of classification.

5 Results and discussion

To evaluate the performance of the proposed hybrid approach, three famous classification algorithms, i.e., decision-stump, zeroR and SVM, were implemented and used as benchmarks for comparison in the experiments. For SVM implementation, the LIBSVM[29] was used and radial basis function was set as default kernel function of SVM. For the proposed approach, the parameters C_{size} and I_{max} were set to 20, respectively. The statistics of classification performance was evaluated by 10-fold cross-validation approach to reduce the bias and variance of classification results. We split each dataset into ten parts. Then we used nine parts for training and the remaining tenth for test. We conducted the training-test procedure ten times and used the average of the ten performances as final result.

Figure 4 indicates the proposed hybrid approach yields a higher performance compared to other techniques for all datasets. The Accuracy value of the proposed algorithm is 63.7% on the white clover dataset, and it beats decisionstump by about 8.2%, zeroR by approximately 3.4% and SVM by about 3.3% on white clover dataset. The Accuracy value of the proposed algorithm is 47.6% on the grub damage dataset, and it beats decision-stump by about 14.2%, zeroR by approximately 16.1% and SVM by about 3.1% on white clover dataset.

The F_1 measured results of various techniques are shown in Figure 5.

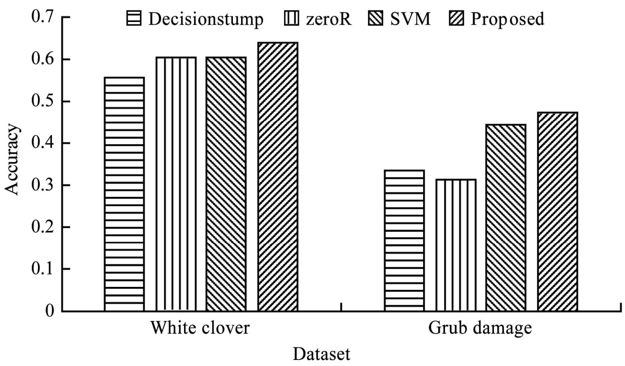

Figure 4 Comparison of accuracy value on two datasets

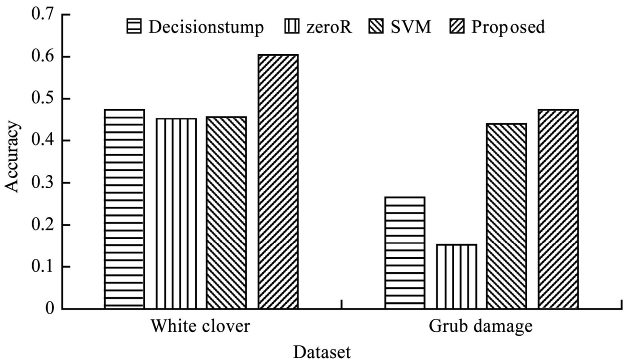

Figure 5 Comparison of F_1 value on two datasets

The proposed hybrid soft computing approach outperforms the other techniques on all datasets according to the experimental results presented in Figure 5. On the white clover dataset, the F_1 value of the proposed hybrid algorithm is 60.5%, and it is approximately 12.9% higher than that of decisionstump, 15.1% higher than that of zeroR and 15% higher than that of SVM. On grub damage dataset, the F_1 value of the proposed hybrid algorithm is 47.6%, and it is approximately 20.9% higher than that of decisionstump, 32.4% higher than that of zeroR and 3.5% higher than that of SVM.

The AUC measure results of various techniques are shown in Figure 6.

Figure 6 indicated that the AUC scores of the proposed approach were improved obviously. On the white clover dataset, the AUC value of the proposed hybrid algorithm is 68.4%, and it was improved 21.6%, 26.3% and 18.3% compared with the decision-stump, zeroR and SVM on white clover dataset, respectively. On the grub damage dataset, the AUC value of the proposed hybrid algorithm is 64%, and it was improved 11.9%, 16.7% and 2.9% compared with the decision-stump, zeroR and SVM on white clover dataset, respectively.

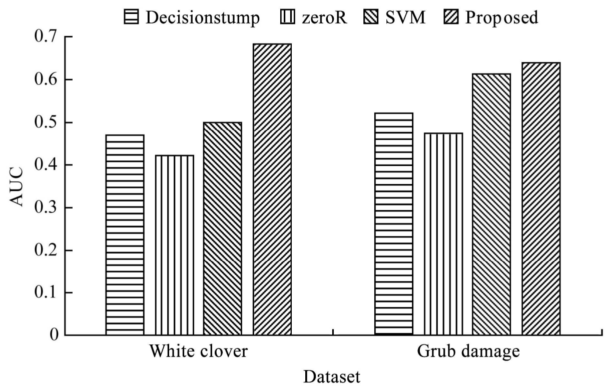

Figure 6 Comparison of AUC value on datasets

The numbers of the selected and used attributes by the proposed approach are shown in Table 2.

Table 2 Number of selected attributes for the algorithm

Dataset	Number of original attributes	Number of selected attributes
White clover	32	5
Grub damage	9	4

According to Table 2, the proposed approach reduces the large number of attributes effectively. For instance, the number of original attributes of white clover dataset is 32. Otherwise, the number of attributes selected and used by the proposed hybrid approach is 5, which is only 15.6% of the original attributes. Above all, the experimental results demonstrate that the proposed hybrid approach can improve classification performance and reduce the number of attributes effectively.

6 Conclusions

Soft computing is a paradigm to exploit the tolerance for uncertainty, imprecision and approximate reasoning for achieving robustness and tractability solutions. And it is currently attracting a great deal of attention and has already found a number of practical applications. With the growth of agricultural data available, building an effective agricultural data classification method with good performance are essential. A novel hybrid soft computing approach based on SVM, rough set and ensemble learning was proposed to classify the agricultural data effectively. Experimental results demonstrate that the proposed hybrid approach improves the performance of classification effectively.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (Grant No.31501225), the

National "Twelfth Five-Year" Plan for Science & Technology Support Program (Grant No.2014BAD10B06), the Key Science and Technology Project of Henan Province (Grant No.142102210054), and Henan Province Key Project of Science and Technology (Grant No.131100110400). We also thank to the editor's hard work and anonymous referees' comments and suggestions on this paper.

[References]

[1] Zadeh L A. Role of soft computing and fuzzy logic in the conception, design and development of information/ intelligent systems. Lecture Notes in Computer Science, 1998: 1–9.

[2] Zadeh L A. Fuzzy logic, neural networks, and soft computing. Communication of the ACM, 1994; 37(3): 77–84.

[3] Fieldsend J, Singh S. Pareto evolutionary neural networks. IEEE Transaction Neural Networks, 2005; 16 (2): 338–354.

[4] Pawlak Z. Rough sets. Int. Journal of Computer and Information Sciences, 1982; 11(5): 341–356.

[5] Joachims T, Finley T, Yu C N. Cutting-plane training of structural SVMs. Machine Learning, 2009; 77(1): 27–59.

[6] Srinivas M, Patnaik L. Adaptive probabilities of crossover and mutation in genetic algorithms. IEEE Transactions on System, Man and Cybernetics, 1994; 24(4): 656–667.

[7] Kirkpatrick S, Gelatt C D, Vecchi M P. Optimization by simulated annealing. Science, 1983; 220 (4598): 671–680.

[8] Huang Y, Lan Y, Thomson S J, Fang A, Hoffmann W C, Lacey R E. Development of soft computing and applications in agricultural and biological engineering. Computers and Electronics in Agriculture, 2010; 71(2): 107–127.

[9] Huang Y. Advances in artificial neural networks-methodological development and application. Algorithms, 2009; 2(3): 973–1007.

[10] 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, 2006; 51(12): 99–109.

[11] Chedad A, Moshou D, Aerts J M, Hirtum A V, Ramon H, Berckmans D. Recognition system for pig cough based on probabilistic neural networks. Journal of Agricultural Engineering Research, 2001; 79(4): 449–457.

[12] Schatzki T F, Haff R P, Young R, Can I, Le L C, Toyofuku N. Defect detection in apples by means of X-ray imaging. Transactions of the ASAE, 1997; 40(5): 1407–1415.

[13] Chen F L, Li F C. Combination of feature selection approaches with SVM in credit scoring. Expert Systems with Applications, 2010; 37(7): 4902–4909.

[14] Zhou Q F, Zhou H, Zhou Q Q, Yang F, Luo L K, Li T. Structural damage detection based on posteriori probability support vector machine and Dempster–Shafer evidence theory. Applied Soft Computing, 2015; 36: 368–374.

[15] Huang C L, Liao H C, Chen M C. Prediction model building and feature selection with support vector machines in breast cancer diagnosis. Expert Systems with Applications, 2008; 4(1): 578–587.

[16] Shi L, Ma X M, Xi L, Duan Q G, Zhao J Y. Rough set and ensemble learning based semi-supervised algorithm for text classification. Expert Systems with Applications, 2011; 38(5): 6300–6306.

[17] Shi L, Xi L, Ma X M, Weng M, Hu X H. A novel ensemble algorithm for biomedical classification based on ant colony optimization. Applied Soft Computing, 2011; 11(8): 5674–5683.

[18] Jensen R, Shen Q. Semantics-preserving dimensionality reduction: rough and fuzzy-rough-based approaches. IEEE Transactions on Knowledge and Data Engineering, 2004; 16(12): 1457–1471.

[19] Dietterich T G. Ensemble methods in machine learning. Springer Berlin Heidelberg, 2000; 1857(1): 1–15.

[20] Peng Y H. A novel ensemble machine learning for robust microarray data classification. Computers in Biology and Medicine, 2006; 36(6): 553–573.

[21] Raafat H M, Tolba A S, Aly A M. A novel training weighted ensemble (TWE) with application to face recognition. Applied Soft Computing, 2011; 11(4): 3608–3617.

[22] Kim M J, Kang D K. Ensemble with neural networks for bankruptcy prediction. Expert Systems with Applications, 2010; 37(4): 3373–3379.

[23] Wei C P, Chen H C, Cheng T H. Effective spam filtering: A single-class learning and ensemble approach. Decision Support Systems, 2008; 45(3): 491–503.

[24] Melville P, Mooney R J. Creating diversity in ensembles using artificial data. Information Fusion, 2005; 6(1): 99–111.

[25] Bazan J G, Nguyen H S, Nguyen S H, Synak P, Wróblewski J. Rough set algorithms in classification problems. Rough Set Methods and Applications: New Developments in Knowledge Discovery in Information Systems, 2000; 56: 49–88.

[26] http://www.cs.waikato.ac.nz/ml/weka/index_datasets.html. Accessed on [2015-3-10].

[27] Yang Y. An evaluation of statistical approaches to text categorization. Information Retrieval, 1999; 1(1-2): 69–90.

[28] Bradley A. The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recognition, 1997; 30(7): 1145–1159.

[29] Chang C C, Lin C J. LIBSVM: a Library for Support Vector Machines. Acm Transactions on Intelligent Systems and Technology, 2006; 2(3): 389–396.