Medical Diagnosis with C4.5 Rule Preceded by Artificial Neural Network Ensemble

Zhi-Hua Zhou, *Member, IEEE*, and Yuan Jiang

Abstract—**Comprehensibility is very important for any machine learning technique to be used in computer-aided medical diagnosis. Since an artificial neural network ensemble is composed of multiple artificial neural networks, its comprehensibility is worse than that of a single artificial neural network. In this paper, C4.5 Rule-PANE which combines artificial neural network ensemble with rule induction by regarding the former as a pre-process of the latter, is proposed. At first, an artificial neural network ensemble is trained. Then, a new training data set is generated by feeding the feature vectors of the original training instances to the trained ensemble and replacing the expected class labels of the original training instances with the class labels output from the ensemble. Additional training data may also be appended by randomly generating feature vectors and combining them with their corresponding class labels output from the ensemble. Finally, a specific rule induction approach, i.e. C4.5 Rule, is used to learn rules from the new training data set. Case studies on *diabetes*, *hepatitis*, and *breast cancer* show that C4.5 Rule-PANE could generate rules with strong generalization ability, which profits from artificial neural network ensemble, and strong comprehensibility, which profits from rule induction.**

Index Terms—**Artificial neural networks, Ensemble learning, Rule induction, Machine learning.**

I. Introduction

Machine learning techniques suiting computer-aided medical diagnosis should have good comprehensibility, i.e. the transparency of diagnostic knowledge and the explanation ability. This is because medical diagnosis, requiring reliable performance, can be profited if the diagnostic process can be checked and adequately explained.

In general, machine learning techniques can be categorized into two classes, that is, symbolic or connectionist learning techniques. Symbolic learning techniques such as rule induction are usually regarded as comprehensible techniques because the learned knowledge is expressed in forms such as production rules that are easy to be understood by the user. Rule induction has already been widely applied in medical diagnosis [5], [8], [22]. On the other hand, most connectionist learning techniques such as artificial neural networks are regarded as incomprehensible techniques because the learned knowledge is concealed in a lot of connections and is not transparent to the user. Although artificial neural networks have already been tried in several medical tasks [16], they have not yet been widely accepted in medicine [15]. Fortunately, during the last decade much work has addressed the issue of improving the comprehensibility of artificial neural networks [1], [30], and some results have already been applied to medical tasks [13], [25], [26].

Artificial neural network ensemble is a learning technique where multiple artificial neural networks are trained to solve the same problem. Since the generalization ability of learning systems based on artificial neural networks can be significantly improved with this technique, it has become a hot topic in both machine learning and neural computing communities [27] and has already been tried in several medical tasks [6], [28], [31]. However, since an ensemble is composed of multiple artificial neural networks, its comprehensibility is even worse than that of a single artificial neural network, which may hinder the wide acceptance of this technique in medicine.

In this paper, a novel learning approach named C4.5 Rule-PANE, i.e. C4.5 Rule Preceded by Artificial Neural Ensemble, is proposed. This approach is helpful for utilizing the power of artificial neural network ensembles in reliable applications where the comprehensibility is as important as the generalization ability. It should be mentioned that C4.5 Rule-PANE does not try to directly improve the comprehensibility of the artificial neural network ensembles. Instead, it employs artificial neural network ensemble as a pre-process for a specific rule induction approach, i.e. C4.5 Rule [20]. Case studies show that C4.5 Rule-PANE could generate rules with strong generalization ability and strong comprehensibility.

The rest of this paper is organized as follows. In Section 2, C4.5 Rule-PANE is proposed after the brief introduction of artificial neural network ensemble and C4.5 Rule. In Section 3, case studies on *diabetes*, *hepatitis*, and *breast cancer* are presented. In Section 4, some issues of C4.5 Rule-PANE that should be further investigated are discussed. Finally in Section 5, the main contribution of this paper is summarized.

Manuscript received February 7, 2002. This work was supported in part by the National Natural Science Foundation of China and the Natural Science Foundation of Jiangsu Province, China

Z.-H. Zhou is with the National Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093, China (phone: +86-25-359-3163; fax: +86-25-330-0710; e-mail: zhouzh@nju.edu.cn).

Y. Jiang is with the National Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093, China (e-mail: jy@ai.nju.edu.cn).

II. Methodology

A. Artificial Neural Network Ensemble

In the beginning of the 1990s, Hansen and Salamon [12] showed that the generalization ability of learning systems based on artificial neural networks can be significantly improved through ensembling artificial neural networks, i.e. training multiple artificial neural networks and combining their predictions. Subsequently there appears a hot wave in investigating artificial neural network ensembles [27], and this technique has already been successfully applied to diverse domains such as optical character recognition [7], [11], [17], face recognition [10], [14], scientific image analysis [4], seismic signals classification [29], *etc.*

In general, an artificial neural network ensemble is built in two steps, that is, generating component artificial neural networks and then combining their predictions. As for generating component networks, Breiman's *Bagging* [3] and Schapire's *Boosting* [24] are prevailing approaches. Bagging generates multiple training data sets from the original training data set via bootstrap sampling [9] and then trains a component network using each of those training data sets. Boosting generates a series of component networks whose training data sets are determined by the performance of the former networks. Training instances that are wrongly predicted by the former networks will play more important roles in the training of the later networks. As for combining component predictions, *voting* [12] is prevailing for classification while *averaging* [18], [19] is prevailing for regression. Voting regards the class label receiving the most number of votes as the final output of the ensemble. Averaging regards the average output of the component networks as the final output of the ensemble. Note that during the latest years much work has addressed the issue of designing effective and efficient ensemble approaches, which has been reviewed in some recent literature [32].

B. C4.5 Rule

C4.5 Rule [20] is a rule induction approach derived from Quinlan's famous C4.5 decision tree [20]. The rules generated by this approach is in conjunctive form such as "if *A* and *B* then *C*" where both *A* and *B* are the rule antecedents while *C* is the rule consequence. The rules are generated as follows.

At first, a C4.5 decision tree is trained. Then, every path from the root to a leaf is converted to an initial rule by regarding all the *test* conditions appearing in the path as the conjunctive rule antecedents while regarding the class label held by the leaf as the rule consequence. After that, each initial rule is generalized by removing antecedents that do not seem helpful for distinguishing a specific class from other classes, which is performed by a pessimistic estimate of the accuracy of the rule. In detail, the accuracy of the initial rule and that of its variant where an antecedent is removed are estimated. If the latter is not worse than the former then the initial rule is replaced by the variant of it. It is worth noting that usually there are several rule antecedents that could be removed. In such cases, C4.5 Rule carries out a greedy elimination, that is, the removal of the antecedent that produces the lowest pessimistic error rate of the generalized rule is kept, and such kind of removal is repeatedly performed until the rule could not be generalized further. After all the initial rules are generalized, they are grouped into rule sets corresponding to the classes respectively. All rule sets are polished with the help of the Minimum Description Length (MDL) Principle [21] so that rules that do not contribute to the accuracy of a rule set are removed. Then, the rule sets are sorted according to the ascending order of their false positive error rates. Finally, a default rule is created for dealing with instances that are not covered by any of the generated rules. The default rule has no antecedent and its consequence is the class that contains the most training instances not covered by any rule.

Note that with the above process, the comprehensibility of the rules generated by C4.5 Rule is better than that of the C4.5 decision tree. This is because the number of rules is usually less than the number of leaves in the tree, and the number of rule antecedents of a rule is usually less than the number of the *test* conditions appearing in the corresponding path in the tree. Moreover, in some cases the generalization ability of the rules may even be better than that of the tree [20].

C. C4.5 Rule-PANE

C4.5 Rule-PANE is helpful for utilizing artificial neural network ensemble in reliable applications such as computer-aided medical diagnosis where the comprehensibility is as important as the generalization ability. The motivation of this approach is the desire of combining the strong generalization ability of artificial neural network ensemble and the strong comprehensibility of rule induction. In detail, C4.5 Rule-PANE employs artificial neural network ensemble as a pre-process of C4.5 Rule. This is the reason why it is called as C4.5 Rule preceded by artificial neural ensemble.

Suppose we have a training data set $S = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$, where x_i and y_i are the feature vector and the expected class label of the i-th training instance, respectively. An artificial neural network ensemble is trained with S. Here Bagging is employed to generate the ensemble but note that other kinds of ensemble approaches such as Boosting can also be used. The pseudo-code of Bagging is shown in Fig. 1, where T bootstrap samples S_1, S_2, …, S_T are generated from the original training data set and a component artificial neural network N_t is trained using each S_t, an ensemble N^* is built from $N_1, N_2, \ldots, N_T$ whose output is the class label receiving the most number of votes.

```
Input: training set S, learner L, number of bootstrap samples T
Output: ensemble N*
Process:
    for t = 1 to T {
      S_t = bootstrap sample from S
      N_t = L(S_t)
    }
```

$$N^*(x) = \arg\max_{y \in Y} \sum_{t: N_t(x) = y} 1$$

Fig. 1. The Bagging approach

For each feature vector x_i (i = 1, 2, …, n), if it is fed to the trained artificial neural network ensemble N^*, a class label y_i' is the output from the ensemble. Note that y_i' may not be identical to the expected class label, i.e. y_i. By combining x_i and y_i', a new instance (x_i, y_i') is generated. Here we call the data set S' = {(x_1, y_1'), (x_2, y_2'), …, (x_n, y_n')} as the training data set processed by N^*. Since artificial neural network ensembles usually have strong generalization ability, we believe that some bad ingredients of S, such as the noise, may be depressed by the process of N^*. In other words, S' may be better than S for rule induction. This is one of the reasons why C4.5 Rule-PANE regards artificial neural network ensemble as a pre-process of C4.5 Rule.

Another reason for regarding artificial neural network ensemble as a pre-process of C4.5 Rule is that after obtaining a trained ensemble, more training data can be generated from the ensemble for the later rule induction so that the resource for rule induction is enriched. This is done by randomly generating some feature vectors and then feeding them to the ensemble. For each randomly generated feature vector x_j' (j = 1, 2, …, m), if it is fed to N^*, a class label y_i'' is the output from the ensemble. By combining x_i' and y_i'', a new instance (x_i', y_i'') is generated. Here we call the data set S'' = {(x_1', y_1''), (x_2', y_2''), …, (x_m', y_m'')} as the *additional training data set* generated by N^*. Since artificial neural network ensembles usually have strong generalization ability, we believe that some important characteristics of the problem to be solved, which has been encoded in S, may be captured by N^*. So, the additional training data set S'' generated by N^* may be useful in complementing S for rule induction.

After S' and S'' are obtained, they are united and used as the training data set for C4.5 Rule. It is obvious that the size of S' is equal to that of the original training data set, i.e. S. But the size of S'' is not certain. From the case studies presented in Section 3, we found that when the size of S'' is roughly treble of that of S, C4.5 Rule-PANE is able to achieve excellent performance.

```
Input: training set S, the number of additional training instance m
Output: rule set R
Process:
    N* = Bagging (S)        /* generate an artificial neural network
                               ensemble N* from S via Bagging */
    S' = ∅
    S'' = ∅
    for i = 1 to n {     /* process the original training data set by
                               the trained ensemble */
      y_i' = N*(x_i: (x_i, y_i)∈S)
      S' = S' ∪ {( x_i, y_i')}
    }
    for j =1 to m {      /* generate additional training data set by
                               the trained ensemble */
      x_j' = Random( )      /* generate a random feature vector */
      y_j'' = N*(x_j')
      S'' = S'' ∪ {( x_i', y_i'')}
    }
    S_R = S' ∪ S''      /* obtain the training data set for rule induction */
    R = C4.5Rule (S_R)
```

Fig. 2. The C4.5 Rule-PANE approach

In summary, the pseudo-code of C4.5 Rule-PANE is shown in Fig. 2.

III. Case Studies

Three case studies, on *diabetes*, *hepatitis*, and *breast cancer*, respectively, have been analyzed. The data set of *diabetes* is from the National Institute of Diabetes and Digestive and Kidney Diseases. There are 768 instances belonging to two classes, i.e. positive or negative for diabetes, described by 8 attributes among which 7 attributes are continuous and the remaining one attribute is categorical. The data set of *hepatitis* contains 80 instances belonging to two classes, i.e. positive or negative for hepatitis, described by 19 attributes among which 6 attributes are continuous and the remaining 13 attributes are categorical. The data set of *breast cancer* contains 683 instances belonging to two classes, i.e. malignant or benign, described by 10 attributes among which 9 attributes are categorical and the remaining one attribute is continuous. All the data sets are publicly available at the UCI machine learning repository [2]. Note that instances with missing values have been removed from the data sets of *hepatitis* and *breast cancer*.

10-fold cross validation is performed in each case study. In detail, the raw data set is divided into ten subsets with similar sizes and similar distribution of classes. Then, tests for the ten folds are performed, each with a different subset as the test data set and with the union of the other nine subsets as the training data set. In each fold, ten runs are performed and the average result of those ten runs is regarded as the result of the fold. The average result of those ten folds is recorded as the final result. In each fold, an artificial neural network ensemble comprising five BP networks, i.e. artificial neural networks trained with the Backpropagation algorithm [23], with one hidden layer containing ten hidden units is generated via Bagging. It should be mentioned that the architecture of the networks has not been finely tuned. The training data sets of the BP networks are bootstrap sampled from the training data set of the fold. Note that bootstrap sampling uses about 63% instances of an original data set to generate a new data set [3]. Therefore for each BP network, the instances that are not in its training data set but in the training data set of the fold can be used as its validation data. During the training process, the generalization error of the network is estimated in each epoch on its validation data set. If the validation error does not change in consecutive five epochs, the training of the network is terminated in order to avoid overfitting.

The predictive error rate and the number of the rules generated by C4.5 Rule-PANE are shown in Table 1. For comparison, Table 1 also shows the predictive error rate and the number of the rules generated by C4.5 Rule, and the predictive error rate of the artificial neural network ensemble. Note that here the size of the additional training data set employed by C4.5 Rule-PANE is treble of that of the original training data set. In other words, m in Fig. 2 is roughly equal to $3n$.

Although C4.5 Rule-PANE is expected to be more accurate than C4.5 Rule as it profits from the generalization ability of the

TABLE I
TEST RESULTS ON DIABETES, HEPATITIS, AND BREAST CANCER

	C4.5 RULE-PANE		C4.5 RULE		ENSEMBLE
	ERROR RATE	NUMBER OF RULES	ERROR RATE	NUMBER OF RULES	ERROR RATE
DIABETES	23.06%	57.3	27.26%	22.4	24.00%
HEPATITIS	10.34%	9.3	15.81%	4.5	14.96%
BREAST CANCER	4.60%	46.8	5.67%	21.4	2.96%

artificial neural network ensemble, it is still impressive that Table 1 indicates the generalization ability of C4.5 Rule-PANE is about 23% (((.2726-.2306)/.2726 + (.1581-.1034)/.1581 + (.0567-.0460)/.0567) / 3 = .2296) better than that of the well-established method C4.5 Rule on these three case studies. More amazingly, the generalization ability of C4.5 Rule-PANE is even better than that of the artificial neural network ensemble on *diabetes* and *hepatitis*. On the other hand, although the comprehensibility of C4.5 Rule-PANE is worse than that of C4.5 Rule since it generates more rules than C4.5 Rule does, its comprehensibility is obviously better than that of the artificial neural network ensemble because symbolic rules are easier to be understood than artificial neural networks.

For further exploring the performance of C4.5 Rule-PANE, more experiments are performed on it with different size of additional training data sets. The 10-fold cross validation results on *diabetes*, *hepatitis*, and *breast cancer*, are shown in Fig. 3 to Fig. 5, respectively. Note that the reported results of C4.5 Rule-PANE have been normalized according to that of C4.5 Rule. In detail, the relative predictive error rate (or number of rules) of C4.5 Rule-PANE shown in each figure is in fact the ratio against that of C4.5 Rule shown in Table 1 which is regarded as 1.0. Moreover, the size of the additional training data set is also shown in the form of the ratio against that of the original training data set.

Fig. 3a and Fig. 4a show that on both *diabetes* and *hepatitis*, even when there is no additional training data, the generalization ability of C4.5 Rule-PANE is better than that of C4.5 Rule. Such an improvement on the relative predictive error rate (about 10%) should owe to that the artificial neural network ensemble has depressed some bad ingredients of the original training data set in the way of correcting class labels. This observation supports our claim that the training data set processed by the artificial neural network ensemble may be better for rule induction than the original training data set. Fig. 5a shows that on *breast cancer*, when there is no additional training data, the generalization ability of C4.5 Rule-PANE is slightly worse than that of C4.5 Rule. We guess the reason may be that the artificial neural network ensemble has wrongly switched the class labels of some outliers so that some useful information has been lost.

Fig. 3a, Fig. 4a, and Fig. 5a also show that the generalization ability of C4.5 Rule-PANE can be improved further by utilizing additional training data. This observation supports our claim that the data generated by the artificial neural network ensemble may be useful in complementing the original training data set for rule induction. However, it should be noted that improvement on the generalization ability introduced by the additional training data is not consistent, which will be discussed in the next section.

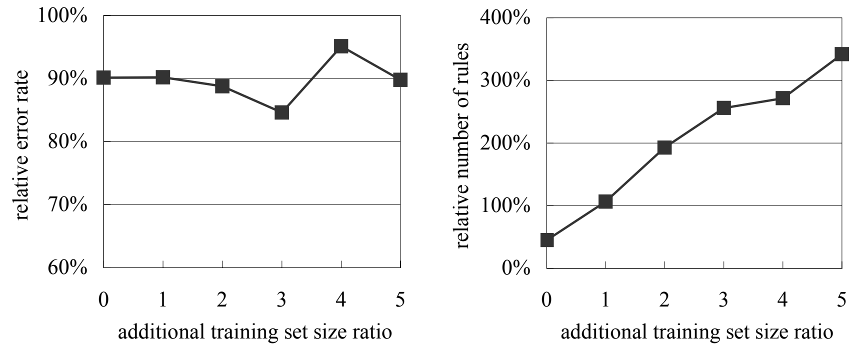
Fig. 3b, Fig. 4b, and Fig. 5b show that the number of rules generated by C4.5 Rule-PANE increases nearly linearly as the size of the training data set used for rule induction increases. However, when no additional training data is used, the number of rules generated by C4.5 Rule-PANE is roughly equal to or even less than that of C4.5 Rule. This means that the comprehensibility of C4.5 Rule-PANE may be better than C4.5 Rule when the size of the training data set is fixed. This also supports our claim that the training data set processed by the artificial neural network ensemble may be better for rule induction than the original training data set. Moreover, even when C4.5 Rule-PANE generates many rules due to the utilization of much additional training data, its comprehensibility is sure to be better than artificial neural network ensemble because symbolic rules are easier to be understood than artificial neural networks.

In summary, Table 1 and Fig. 3 to Fig. 5 reveal that C4.5 Rule-PANE could generate rules with strong generalization ability and strong comprehensibility on all the three case studies.

IV. DISCUSSION

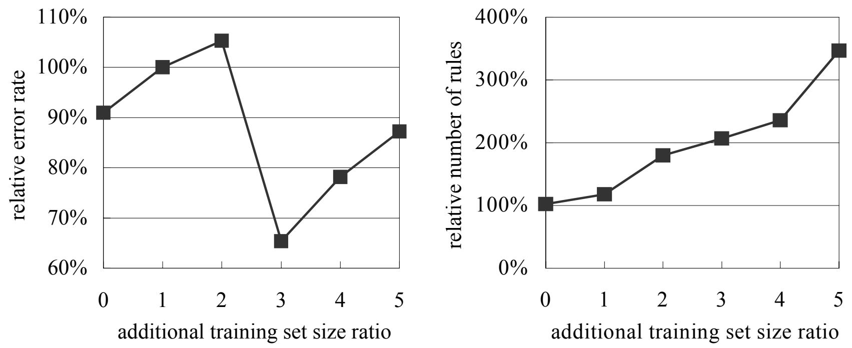
Case studies presented in Section 3 show that C4.5 Rule-PANE can achieve good performance, but they also show that the performance of C4.5 Rule-PANE is not very stable. The instability is exhibited in two aspects.

Firstly, on one hand, the generalization ability of C4.5 Rule-PANE on *diabetes* and *hepatitis* is better than that of C4.5 Rule even when no additional training data is used, but on *breast cancer* this does not hold. On the other hand, the generalization ability of C4.5 Rule-PANE can be improved by utilizing additional training data in all the three case studies. This suggests that for the first two case studies, the artificial neural network ensemble simultaneously contributes to the generalization ability of C4.5 Rule-PANE in two manners, that is, processing the original training data and generating the additional training data, but for the third case study, the artificial neural network ensemble only contributes in the second manner. It is worth noting that when the ensemble contributes in both manners (as on *diabetes* or *hepatitis*), the generalization ability of C4.5 Rule-PANE is even better than that of the ensemble. We believe that whether the ensemble can contribute to the generalization ability of C4.5 Rule-PANE in



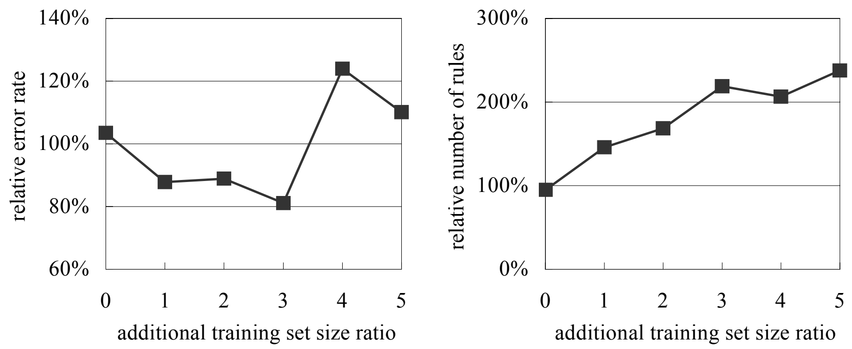
a) The relative error rate b) The relative number of rules

Fig. 3. The test results of C4.5 Rule-PANE with different size of additional training data set on *diabetes*



a) The relative error rate b) The relative number of rules

Fig. 4. The test results of C4.5 Rule-PANE with different size of additional training data set on *hepatitis*



a) The relative error rate b) The relative number of rules

Fig. 5. The test results of C4.5 Rule-PANE with different size of additional training data set on *breast cancer*

both manners or not depends on some properties of the problem to be solved, the most possible one of which may be the abundance of outliers. Finding out those properties is an interesting issue for future work.

Secondly, the improvement on the generalization ability of C4.5 Rule-PANE introduced by additional training data is not consistent. This may owe to two reasons.

The first reason may be that the generalization ability could not be infinitely improved through utilizing more additional training data. So, the generalization ability of C4.5 Rule-PANE may increase as the size of the additional training data set approaching some certain point, but decrease after that point. It is obvious that the ideal choice of the size of the additional training data set is just the size at the turning point. Exploring how to determine the turning point is another interesting issue for future work. It is likely that the turning point may not be independent of the problem to be solved. If it is the truth, we hope to devise some feasible approaches that could help the user to tune the performance of C4.5 Rule-PANE for specific problems.

The second reason for the inconsistency of the improvement of the generalization ability may be that at present the feature vectors of the additional training data are randomly generated. Ideal additional training data should serve to fill the gap between the actual distribution and the training data at hand, but random generation of feature vectors may create unrealistic cases of data. Thus, both the predictive error rate and the number of rules generated by C4.5 Rule-PANE may abruptly change as the additional training data changes. So, an important issue for future work is to develop some kind of guideline for the generation of the feature vectors of the additional training data so that unrealistic data will not be created. Note that such kind of guideline may not be obtained without the help of domain knowledge.

In addition, in the case studies presented in Section 3, the generalization ability of the artificial neural network ensemble is always better than that of the C4.5 Rule. Exploring the performance of C4.5 Rule-PANE when the generalization ability of the artificial neural network ensemble is not better than that of C4.5 Rule is also an interesting issue for future work.

V. Conclusion

Artificial neural network ensemble is a powerful learning technique that could significantly improve the generalization ability of neural learning systems, but its comprehensibility is even worse than that of a single artificial neural network, which may hinder the wide acceptance of this technique in medicine. In this paper, an approach named C4.5 Rule-PANE is proposed, which gracefully combines the advantages of artificial neural network ensemble and rule induction. Case studies show that C4.5 Rule-PANE could benefit medical diagnosis because it could generate rules with strong generalization ability and strong comprehensibility.

A prominent advantage of C4.5 Rule-PANE is that it is very easy to be used. This is because there are code or tools that could be easily obtained for C4.5 Rule and artificial neural networks (the code of C4.5 has been presented in Quinlan's book [20], and the software package C4.5 Release 8 is available at http://www.cse.unsw.edu.au/~quinlan/. Codes or tools of artificial neural networks could be easily obtained from the Internet, as indicated by ftp://ftp.sas.com/pub/neural/FAQ5.html#A_source), and the ensemble approaches such as Bagging is quite easy to be programmed. However, as discussed in Section 4, there are several problems on C4.5 Rule-PANE that should be investigated in the future. Exploring those problems may not only help us to finely tune C4.5 Rule-PANE, but also torch the way of designing novel learning approaches with strong comprehensibility as well as strong generalization ability.

Acknowledgment

The comments and suggestions from the anonymous reviewers greatly improved this paper.

References

[1] R. Andrews, J. Diederich, and A. B. Tickle, "Survey and critique of techniques for extracting rules from trained artificial neural networks," *Knowledge-Based Systems*, vol.8, no.6, pp.373-389, 1995.

[2] UCI Repository of Machine Learning Databases, C. Blake, E. Keogh, and C. J. Merz. (1998). [Online]. Available: http://www.ics.uci.edu/~mlearn/MLRepository.htm

[3] L. Breiman, "Bagging predictors," *Machine Learning*, vol.24, no.2, pp.123-140, 1996.

[4] K. J. Cherkauer, "Human expert level performance on a scientific image analysis task by a system using combined artificial neural networks," in *Proc. 13th AAAI Workshop on Integrating Multiple Learned Models for Improving and Scaling Machine Learning Algorithms*, Portland, OR, 1996, pp.15-21.

[5] D. Cosic and S. Loncaric, "Rule-based labeling of CT head image," in *Lecture Notes in Artificial Intelligence 1211*, E. Keravnou, C. Garbay, R. Baud, and J. Wyatt, Eds. Berlin: Springer-Verlag, 1999, pp.453-456.

[6] P. Cunningham, J. Carney, and S. Jacob, "Stability problems with artificial neural networks and the ensemble solution," *Artificial Intelligence in Medicine*, vol. 20, no.3, pp.217-225, 2000.

[7] H. Drucker, R. Schapire, and P. Simard, "Improving performance in neural networks using a boosting algorithm," in *Advances in Neural Information Processing Systems 5*, S. J. Hanson, J. D. Cowan, and C. L. Giles, Eds. San Mateo, CA: Morgan Kaufmann, 1993, pp.42-49.

[8] W. Duch, R. Adamczak, K. Grabczewski, G. Zal, and Y. Hayashi, "Fuzzy and crisp logical rule extraction methods in application to medical data," in *Computational Intelligence and Applications 23*, P. S. Szczepaniak, Ed. Berlin: Springer-Verlag, 2000, pp.593-616.

[9] B. Efron and R. Tibshirani, *An Introduction to the Bootstrap*. New York: Chapman & Hall, 1993.

[10] S. Gutta and H. Wechsler, "Face recognition using hybrid classifier systems," in *Proc. IEEE Int. Conf. Neural Networks*, Washington, DC, 1996, pp.1017-1022.

[11] L. K. Hansen, L. Liisberg, and P. Salamon, "Ensemble methods for handwritten digit recognition," in *Proc. IEEE Workshop on Neural Networks for Signal Processing*, Helsingoer, Denmark, 1992, pp.333-342.

[12] L. K. Hansen and P. Salamon, "Neural network ensembles," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.12, no.10, pp.993-1001, 1990.

[13] Y. Hayashi, R. Setiono, and K. Yoshida, "A comparison between two neural network rule extraction techniques for the diagnosis of hepatobiliary disorders," *Artificial Intelligence in Medicine*, vol.20, no.3, pp.205-216, 2000.

[14] F. J. Huang, Z.-H. Zhou, H.-J. Zhang, and T. H. Chen, "Pose invariant face recognition," in *Proc. 4th IEEE Int. Conf. Automatic Face and Gesture Recognition*, Grenoble, France, 2000, pp.245-250.

[15] I. Kononenko, "Machine learning for medical diagnosis: history, state of the art and perspective," *Artificial Intelligence in Medicine*, vol.23, no.1, pp.89-109, 2001.

[16] P. J. G. Lisboa, E. C. Ifeachor, and P. S. Szczepaniak, Eds. *Artificial Neural Networks in Biomedicine*. London: Springer-Verlag, 2000.

[17] J. Mao, "A case study on bagging, boosting and basic ensembles of neural networks for OCR," in *Proc. IEEE Int. J. Conf. Neural Networks*, vol.3, Anchorage, AK, 1998, pp.1828-1833.

[18] D. W. Opitz and J. W. Shavlik, "Actively searching for an effective neural network ensemble," *Connection Science*, vol.8, no.3-4, pp.337-353, 1996.

[19] M. P. Perrone and L. N. Cooper, "When networks disagree: ensemble method for neural networks," in *Artificial Neural Networks for Speech and Vision*, R. J. Mammone, Ed. New York: Chapman & Hall, 1993. pp.126-142.

[20] J. R. Quinlan, *C4.5: Programs for Machine Learning*. San Mateo, CA: Morgan Kaufmann, 1993.

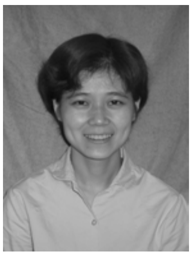
[21] J. R. Quinlan and R. L. Rivest, "Inferring decision trees using the minimum description length principle," *Information and Computation*, vol.80, no.3, pp.227-248, 1989.

[22] G. Richards, V. J. Rayward-Smith, P. H. Sönksen, S. Carey, and C. Weng, "Data mining for indicators of early mortality in a database of clinical records," *Artificial Intelligence in Medicine*, vol.22, no.3, pp.215-231, 2000.

[23] D. Rumelhart, G. Hinton, and R. Williams, "Learning representations by backpropagating errors," *Nature*, vol.323, no.9, pp.318-362, 1986.

[24] R. E. Schapire, "The strength of weak learnability," *Machine Learning*, vol. 5, no.2, pp.197-227, 1990.
[25] R. Setiono, "Extracting rules from pruned neural networks for breast cancer diagnosis," *Artificial Intelligence in Medicine*, vol.8, no.1, pp.37-51, 1996.
[26] R. Setiono, "Generating concise and accurate classification rules for breast cancer diagnosis," *Artificial Intelligence in Medicine*, vol.18, no.3, pp.205-219, 2000.
[27] D. Sharkey, Ed. *Combining Artificial Neural Nets: Ensemble and Modular Multi-net Systems*. London: Springer-Verlag, 1999.
[28] A. J. C. Sharkey, N. E. Sharkey, and S. S. Cross, "Adapting an ensemble approach for the diagnosis of breast cancer," in *Proc. Int. Conf. Artificial Neural Networks*, Skövde, Sweden, 1998, pp.281-286.
[29] Y. Shimshoni and N. Intrator, "Classification of seismic signals by integrating ensembles of neural networks," *IEEE Trans. Signal Processing*, vol.4, no.5, pp.1194-1201, 1998.
[30] A. B. Tickle and R. Andrews, "The truth will come to light: directions and challenges in extracting the knowledge embedded within trained artificial neural networks," *IEEE Trans. Neural Networks*, vol.9, no.6, pp.1057-1068, 1998.
[31] Z.-H. Zhou, Y. Jiang, Y.-B. Yang, and S.-F. Chen, "Lung cancer cell identification based on artificial neural network ensembles," *Artificial Intelligence in Medicine*, vol.24, no.1, pp.25-36, 2002.
[32] Z.-H. Zhou, J. Wu, and W. Tang, "Ensembling neural networks: many could be better than all," *Artificial Intelligence*, vol.137, no.1-2, pp.239-263, 2002.

Zhi-Hua Zhou (S'00–M'01) received the BSc, MSc and PhD degrees in computer science from Nanjing University, China, in 1996, 1998 and 2000, respectively, all with the highest honor. At present he is an associate professor of the Computer Science & Technology Department of Nanjing University. His current research interests are in machine learning, neural computing, evolutionary computing, pattern recognition, and data mining. In these areas he has published over 30 technical papers in referred journals or conferences. In 1999 he won the Mircrosoft Fellowship Award. He has chaired the organizing committee of the 7th Chinese Workshop on Machine Learning, and served as program committee members for many conferences. He is on the editorial board of the journal *Artificial Intelligence in Medicine*. He is an executive committee member of Chinese Association of Artificial Intelligence (CAAI) and CAAI Machine Learning Council, and is a member of IEEE and IEEE Computer Society.



Yuan Jiang received the MSc degree in computer science from Nanjing University, China, in 2001. At present she is a PhD candidate at the Computer Science & Technology Department of Nanjing University. Her current research interests are in machine learning, neural computing, and information retrieval.