Learning Classifier System Ensemble for Data Mining

Yang Gao
State Key Laboratory for Novel Software Technology, Nanjing University
Nanjing, China
gaoy@nju.edu.cn

Joshua Zhexue Huang
E-business Technology Institute, The University of Hong Kong
Hong Kong, China
jhuang@eti.hku.hk

Hongqiang Rong
Department of Computer Science, The University of Hong Kong
Hong Kong, China
hqrong@cs.hku.hk

Daqian Gu
State Key Laboratory for Novel Software Technology, Nanjing University
Nanjing,
Chinagudaqian@ai.nju.edu.cn

ABSTRACT

This paper proposes LCSE, a learning classifier system ensemble, which is an extension of the classical learning classifier system(LCS). The classical LCS includes two major modules, a genetic algorithm module used to facilitate rule discovery, and a reinforcement learning module used to adjust the strength of the corresponding rules while it receives the rewards from the environment. In LCSE we build a two-level ensemble architecture to enhance the generalization of LCS. In the first-level, new instances are first bootstrapped and sent to several LCSs for classification. Then, in the second-level, a plurality-vote method is used to combine the classification results of individual LCSs into a final decision. Experiments on some benchmark data sets from the UCI repository have shown that LCSE has better generalization ability than the single LCS and other supervised learning methods.

Categories and Subject Descriptors

I.2 [**Artificial Intelligence**]: Learning; I.2.6 [**Learning**]:

General Terms

Algorithms

Keywords

Learning classifier system, Ensemble

1. INTRODUCTION

The Learning Classifier System(LCS) is a machine learning technique that combines reinforcement learning, evolutionary computing and heuristics into an adaptive system. Each learning classifier system is a rule-based system in which the rules are in the form of "If conditions THEN action". Evolutionary computing techniques and heuristics are used to search for the possible rules, whilst the reinforcement learning techniques are used to assign utility to existing rules, thereby guiding the search for better rules. LCS had already been successfully applied into data mining domains, especially in medical data analysis. Bonelli et al.[3] have demonstrated that the learning classifier system is suitable in three medical domains. Holmes applied LCS to knowledge discovery in the clinical research databases and achieved some fruitful results in estimation of disease risk and epidemiologic surveillance [5][6]. Wilson [7][8] proposed the XCSR technique to adapt real-value attributes in LCS and used XCS in the oblique data set from the Wisconsin Breast Cancer data. Bernadó et al.[2] selected several medical data sets from the UCI repository, such as Pima-indians etc., to compare the performances of XCS and GALE. Similar works have been done by Bacardit et al. on comparison of XCS and GAssist[1].

Though the above researches have shown that LCS works well on the data mining domain. But, with smaller data sets, LCS clearly tends to over-fit the data[1]. Other problems are the noisy data (incorrect measured values) and the missing data which often occur in actual data sets. To solve these problems with LCS, we need to improve it in terms of generalization to avoid over-fitting and to increase accuracy of classification. So far, the ensemble method is one of most interesting and attractive learning systems with strong generalization ability. Leo Breiman[4] proved that generating multiple versions of a predictor and using them to get an aggregated predictor can improve the prediction accuracy of a weak predictor. The ensemble technology based on supervised machine learning has been studied in and applied to several domains, such as medical diagnosis, face recognition, scientific image analysis [10], etc. However, combination of LCS with ensemble learning has not yet been well tested.

In this paper, we propose a two-level LCS ensemble system(LCSE) and apply it into data mining. In the first-level of the ensemble, new instances are first bootstrapped and sent to several LCSs for classification. Then, in the second-level, a plurality-vote method is used to combine the classification results of individual LCSs into a final decision. Experiments on some benchmark data sets in UCI repository have shown that LCSE has better performance on incremental data learning and better generalization ability than the single LCS and other supervised learning methods.

The paper is organized as follows. In Section 2, a two-level learning classifier system ensemble is presented and the learning process is discussed in detail. We conduct some test experiments on the Pima Indians Diabetes data set and investigate the performance of the respective approaches in Section 3. Finally, we draw some conclusions and outline future works in Section 4.

2. LEARNING CLASSIFIER SYSTEM ENSEMBLE

The learning classifier system ensemble or LCSE we propose in this paper combines the learning classifier system with ensemble learning in order to improve the system generality. Fig. 1 shows the system architecture. Besides several sub LCSs, a bootstrap module and a voting module are added. The bootstrap module is used to distribute the inputs into different sub-LCSs and the voting module is used to combine all classification results of sub-LCSs to produce the final system output.

At initialization of the system, the sub-LCSs generate their population sets. In each episode, the bootstrap module inputs new samples to every sub-LCSs randomly with respect to the probability λ given in equation (1) where LCS_i denotes the ith sub-LCS. If the ith sub-LCS receives an input, it constructs the match set $[M]_i$. Then the wheel selecting algorithm is used to get the action set $[A]_i$. Finally, the ith sub-LCS outputs its classification result a_i. The voting module ensembles the sub-LCSs' outputs according to equation (2) to get the final system's classification result a_{LCSE}. The voting module uses the basic plurality voting method. We must emphasize that each sub-LCS that is participates in voting will receive a respective payoff r_i from the environment with respect to its input a_i. Every sub-LCS processes reinforcement learning of the rule strength and rule discovery based on its payoff. In figure 1, the solid line represents that the ith sub-LCS is activated. The dashed line indicates that the respective sub-LCS is not activated in the learning episode.

$$\begin{cases} if\ rand_i() \le \lambda & bootstrap(LCS_i) \leftarrow TRUE \\ else & bootstrap(LCS_i) \leftarrow FALSE \end{cases} \tag{1}$$

$$a_{LCSE} \leftarrow argmax_a \sum_i vote(a_i) \tag{2}$$

Essentially, the bootstrap module in LCSE aims at developing multiple sub-LCSs with different classification performances by means of inputting different samples, initiating different population sets, undertaking different rule learning and discovery processes. In other words, the classification results of every sub-LCS may be different even though the input sample is same. Therefor, the generality of the learning classifier system is improved when combing multiple sub-LCSs with ensemble learning.

Table 1: Results of stratified cross-validation test on the benchmark dataset Using LCSE with 7 sub-LCSs

	Correct Classification	Misclassification	Fraction Accuracy
Trial 1	58	19	0.7532
Trial 2	59	18	0.7662
Trial 3	54	23	0.7013
Trial 4	57	20	0.7403
Trial 5	58	19	0.7532
Trial 6	61	16	0.7922
Trial 7	63	14	0.8182
Trial 8	64	13	0.8312
Trial 9	56	20	0.7368
Trial 10	64	12	0.8421
Mean			0.7734

3. EXPERIMENTAL RESULTS AND ANALYSIS

The Pima Indians Diabetes data set from National Institute of Diabetes and Digestive and Kidney diseases has been used in benchmark performance test. The diabetes dataset consist of eight condition attributes.

In order to highlight the over-fitting problem when it occurs, the stratified tenfold cross-validation method was applied to the diabetes dataset in the first experiment. We randomly divided the diabetes dataset into 10 subsets (or folds). In each subset all actions had the same likelihood to be taken. In each trial, we uses one subset for testing and the remaining nine subsets for learning. Results of the stratified cross-validation test of LCSE with seven sub-LCSs are shown in table 1. Table 2 shows the comparison results of the average fraction correction between LCSE with seven sub-LCSs and other learning methods. Among these methods, we used the decision tree and neural network methods in the Weka toolbox from the University of Waikato in New Zealand [9].

From table 1 and table 2, we can see that all LCSE configurations performed better than LCS. And LCSE with 7 sub-LCSs performed the best of all LCS. Its average prediction accuracy reached 77.3%. We can also see from Tab. 2 that the average fraction accuracy increases as the number of sub-LCSs increases in LCSE. However, there is a limit in which the method approaches the maximal accuracy and further addition of sub-LCSs will no longer have effect.

We conducted the second experiment to investigate the on-line learning performance. In the second experiment, we also split the whole dataset into 10 subsets. Firstly, we used the first subset for learning and the second subset for testing. Then, we used the first two subsets for learning and the third subset for testing. This process continued till the step that the first nine subsets were used for learning and the last subset for testing. In fact, since both the learning classifier system LCS and the learning classifier system ensemble LCSE are on-line learning methods, it is unnecessary to separate the testing process from the learning process. The purpose that we designed the above learning and test-

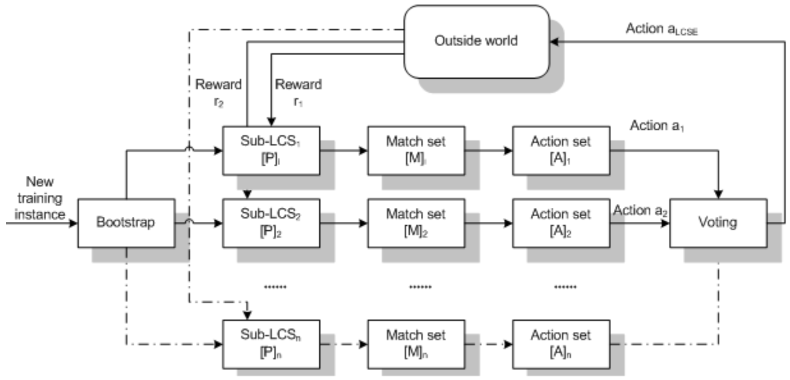

Figure 1: The Architecture of Two-level Learning Classifier System Ensemble.

Table 2: Comparing LCSE with LCS, Decision Tree and Neural Network

Learning Method	Correct Classification	Misclassification	Average Fraction Accuracy
LCSE with 7 sub-LCSs	594	174	0.7734
LCSE with 5 sub-LCSs	577	191	0.7513
LCSE with 3 sub-LCSs	564	204	0.7344
LCS	548	220	0.7135
Decision Tree(J48)	554	214	0.7213
Neural Network	585	183	0.7617

ing stages is to compare LCSE with decision tree and neural network methods.

We conducted 10 trials on each learning method. The results are shown in Fig.2 and Fig.3. We can see that the prediction accuracy has been improved when the systems learnt more instances. LCSE with 7 sub-LCSs performed better in the on-line learning than other LCSEs and LCS in almost all learning stages.

4. CONCLUSIONS

This paper has presented a Learning Classifier System Ensemble (LCSE) which combines the learning classifier system with ensemble learning to improve the generality of the single learning classifier system. The experiment results have shown that LCSE with sub-LCSs performs better than other learning methods such as decision tree and neural networks on the diabetes data set. Furthermore, the LCSEs with more sub-LCSs outperform the LCSEs with less sub-LCSs as well as LCS. These initial results have demonstrated the advantages of the LCSE learning system.

5. ACKNOWLEDGEMENT

The paper is supported by the Natural Science Foundation

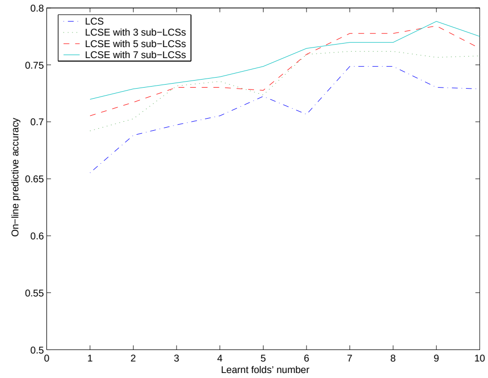

Figure 2: Comparison between LCSE and LCS.

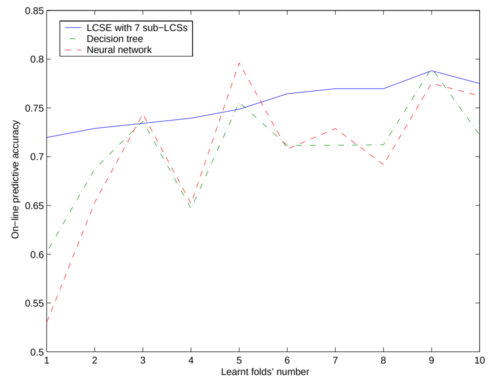

Figure 3: Comparison between LCSE and Neural Network and Decision Tree.

of China (No.60475026), the National Outstanding Youth Foundation of China (No.60325207), the National Grand Fundamental Research 973 Program of China (No.2002CB312002) and the Natural Science Foundation of Jiangsu Province, China(No.BK2003409).

6. REFERENCES

[1] Jaume Bacardit, Martin V. Butz, Data Mining in Learning Classifier Systems: Comparing XCS with GAssist, in L. Bull (ed), Applications of Learning Classifier Systems, Springer-Verlag LNAI Series, In press.

[2] Ester Bernadó, Xavier Llorà, Josep M. Garrell, XCS and GALE: A Comparative Study of Two Learning Classifier Systems on Data Mining. In: Pier Luca Lanzi, Wolfgang Stolzmann, Stewart W. Wilson, editors, Advances in Learning Classifier Systems. LNAI 2321, pages 115-132, Springer-Verlag, Berlin, 2002.

[3] P. Bonelli, A. Parodi, An Efficient Classifier System and Its Experimental Comparison with Two Representative Learning Methods on Three Medical Domains. In R.K. Belew and L.B. Booker, editors, Proceedings of the fourth international conference on Genetic algorithms(ICGA-4), pages 288-295. San Mateo, CA:Morgan Kaufmann, 1991.

[4] Leo Breiman, Bagging Predictors, Machine Learning, Vol.24, No.2, pp.123-140, 1996.

[5] John H. Holmes, Learning Classifier Systems Applied to Knowledge Discovery in Clinical Research Databases. In: Pier Luca Lanzi, Wolfgang Stolzmann, Stewart W. Wilson, editors, Learning Classifier Systems. From Foundations to Applications. LNAI 1813, pages 243-261, Springer-Verlag, Berlin, 2000.

[6] John H. Holmes, Applying a Learning Classifier System to Mining Explanatory and Predictive Models from a Large Database. In: Pier Luca Lanzi, Wolfgang Stolzmann, Stewart W. Wilson, editors, Advances in Learning Classifier Systems. LNAI 1996, pages 103-113, Springer-Verlag, Berlin, 2001.

[7] Stewart W. Wilson. Get Real! XCS with continous-valued inputs. In P. L. Lanzi, W. Stolzmann and S. W. Wilson(eds.), Learning Classifier Systems. From Foundations to Applications. LNAI 1813, pages 209-219, Springer-Verlag, Berlin, 2000.

[8] Stewart W. Wilson, Mining Obilque Data with XCS. In: Pier Luca Lanzi, Wolfgang Stolzmann, Stewart W. Wilson, editors, Advances in Learning Classifier Systems. LNAI 1996, pages 158-174, Springer-Verlag, Berlin, 2001.

[9] http://www.cs.waikato.ac.nz/ ml/weka/, Last visit at 22, Dec., 2004.

[10] Z.-H. Zhou, J. Wu, and W. Tang. Ensembling neural networks: many could be better than all. Artificial Intelligence, 137(1-2): 239-263, 2002.