A New Compaction Algorithm for LCS Rules
Breast Cancer Dataset Case Study

Faten Kharbat[1], Larry Bull[2] and Mohammed Odeh[2]
[1]*Emirates College of Technology, Abu Dhabi, U.A.E.*
[2]*University of the West of England, Bristol, U.K.*

Keywords: Learning Classifier System, Compaction Algorithm, XCS.

Abstract: This paper introduces a new compaction algorithm for the rules generated by learning classifier systems that overcomes the disadvantages of previous algorithms in complexity, compacted solution size, accuracy and usability. The algorithm is tested on a Wisconsin Breast Cancer Dataset (WBC) which is a well well-known breast cancer datasets from the UCI Machine Learning Repository.

1 INTRODUCTION

Learning Classifier Systems (LCS) (Holland, 1976) is a sample of the main issues that have been investigated in Artificial Intelligence (AI) over the last three decades. LCS is a rule-based system that uses evolutionary algorithms to facilitate rule discovery. It may be said that most current LCS research has made a shift away from Holland's original formalism after Wilson introduced XCS (Wilson, 1995).

XCS uses the accuracy of rules as their fitness and Genetic Algorithms (GA) (Holland, 1975) to evolve generalizations over the space of possible state-action pairs of a reinforcement learning task with the aim of easing the use of such approaches in large problems, (i.e., those with state-action combinations that are too numerous for an explicit entry for each). XCS can also avoid problematic over general rules that receive a high optimal payoff for some inputs, but are sub-optimal to other lower payoff inputs.

Illustrates the architecture of XCS, and readers who are interested in further details are referred to (Butz and Wilson, 2000) and (Butz et al., 2004).

LCS in general and XCS in particular have been applied to different data mining problems. It was shown that LCS could be effective for predicting and describing evolving phenomenon in addition to its modelling ability (e.g. Holmes et al., 2002). In particular, Wilson (2001a) (2001b) applied XCS to a medical dataset, namely the Wisconsin Breast Cancer Dataset (WBC), and showed that XCS can tackle real complex learning problems, in addition to its capability to deal with different representations. Also, XCS was tested on the Wisconsin Diagnostic Breast Cancer Dataset (WDBC) dataset in (Bacardit and Butz, 2004) and shown to have competitive performance in both training and testing phases.

However, in real environments, having generated descriptive rules, an LCS needs a further step in which the minimal number of rules can be found that can still describe this environment. In other words, this implies that a compaction process is required to run over the rules generated by the learning classifier system.

A number of approaches have been attempted to develop a sufficient compaction algorithm where a minimal subset of rules can be extracted with minimal run time required. In general, these attempts suffer from the same deficiency in terms of poor performance and difficult usability.

In this paper, a new compaction algorithm that overcomes the previous algorithms' rules compaction disadvantages is introduced. Evaluation of the results obtained is discussed briefly after applying the algorithm to a well-known breast cancer datasets: Wisconsin Breast Cancer Dataset (WBC) (Blake and Merz, 1998), followed by a conclusion and future directions.

Table 1 (Bernado et al., 2004) shows the prediction accuracy of XCS over the WBC (average and standard deviation) compared to other popular learning algorithms showing the efficiency and ability of XCS to tackle real complex problems. In this research, the WBC dataset has been used as the

Kharbat F., Bull L. and Odeh M..
A New Compaction Algorithm for LCS Rules - Breast Cancer Dataset Case Study.
DOI: 10.5220/0004167403820385
In *Proceedings of the International Conference on Knowledge Discovery and Information Retrieval* (KDIR-2012), pages 382-385
ISBN: 978-989-8565-29-7

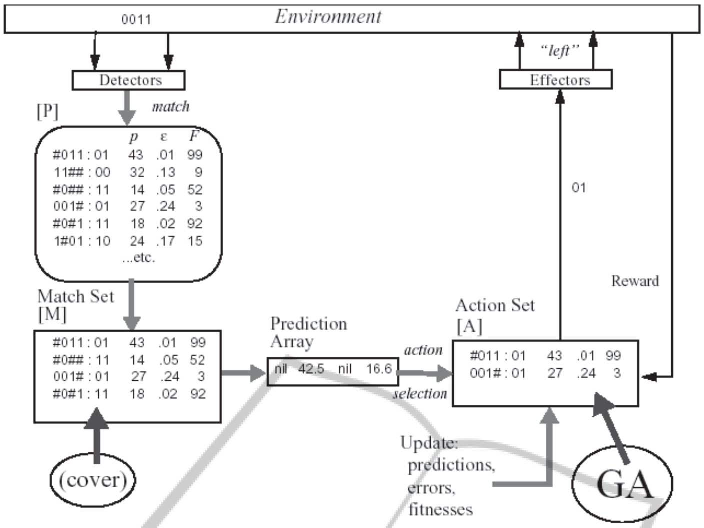

Figure 1: XCS Architecture (Wilson, 1995).

Table 1: Prediction accuracy of XCS and other learning algorithms on the WBC (Bernado et al., 2004).

DS	C4.5r8	PART	SMO	XCS
WBC	95.4 ±1.6	95.3±2.2	96.7±1.7	96.4±2.5

test bed to study and evaluate the outcomes of the new LCS compaction approach, namely Compaction using Recognise-Act Cycles (CRAC).

2 APPROACHES TO LCS RULE COMPACTION

2.1 Goals of LCS Rule Compaction

XCS has shown encouraging results in different domains in terms of its capability to produce a maximal, general, correct solution for a given environment. The huge size of the generated solution, however, may still be considered as a barrier to exploit its entire knowledge. For example, more than 2000 rules on average were generated when WBC dataset was applied to XCS.

The main objective of applying real problems to LCS is to provide the domain experts with a complete, minimal, readable, and usable solution with an organized underlying knowledge that has the ability to describe the given environment. "*Complete*" is one of the proved characteristics related to XCS (Kovacs, 1997) which implies that XCS is able to describe all regions of the input/action space (complete map) for a given environment. However, by increasing the number of rules describing the environment, overlapped patterns are allowed to exist, which conflict with the second term: *"minimal"*. In other words, there will be some regions in the environment that are described and covered by more than one rule (or pattern). Actually, some real problem domains require an overlapping solution as per their nature of complexity. But, the issue here relates to unnecessary overlapping that can be avoided.

One of the other main problems caused by large numbers of LCS rules is when they are presented to domain experts. This violates the third term *"readable"* due to the over expected number of rules that make it impossible to comprehend them smoothly or make the maximum benefit of them. For example, providing a breast cancer specialist with more than 2000 rules describing the 700 WBC cases is not easily comprehendible to make use of the underlying hidden knowledge for better understanding and enrichment of breast cancer knowledge.

Therefore, developing a compaction algorithm that addresses the above issues is essential to increase the level of rules readability, interpretation, and organization of the underlying knowledge held in them. We present below a brief description of the main algorithms attempted to compact LCS rules.

It is clear from the previous algorithms' descriptions and from experimental results (not shown) that Dixon et al. algorithm outperforms other algorithms although it may generate larger compacted solution in some cases. However, two main issues should be considered; first, the usability of the rules is essential which stands for the ability of the domain expert to utilize the generated solution without applying each new problem case to the prediction array calculation. In other words, the dependency on the prediction array calculation means that the solution should be resident in a computer system which makes no difference to store the full generated rule set from the beginning. Second, the quality of generated compacted rules needs to be sustained. While Dixon et al.'s algorithm performance is competitive, it can be seen that involving spoilers will adversely affect the quality of the compacted solution classifiers let alone the readability and usability of these rules.

3 THE NEW APPROACH

In an attempt to combat the main disadvantage of Dixon et al.'s algorithm, a new algorithm has been devised one additional step is added and another one is modified. The added step is to calculate for each rule its entropy as follows:
which represents its correct covering percentage and is affected by its incorrect ones. The entropy represents an accuracy measurement for each rule by which its potential can be evaluated. The higher the entropy, the higher its weight will be.

$$\text{entropy}(\text{rule}_i) = \frac{\sum \text{correct matched cases} - \sum \text{wrong matched cases}}{\text{number of cases}}$$

The algorithm continues to perform for each fact (case) the prediction array calculation as usual and the higher fitness-weighted action is selected. The modification step is that the rule with the highest entropy is selected from the action set and then added to the final compacted set. However, if the prediction array calculation reveals an equal weight for the actions, the rule with the highest entropy in the match set - instead of the action set - is to be added to the final compacted set. This step insures that the added rule covers correctly the largest number of facts which guarantees, to some extent, the generality of the final compacted set. Moreover, the rules in the final compacted set could be used without the need for a prediction array calculation. This new algorithm is summarized below.

```
1. For each rule in the ruleset find its entropy.
2. For each fact in the dataset
    2.1. Create its match set and prediction array.
    2.2. Select the best action which is represented by the highest fitness-weight calculation.
    2.3. Add the rule that has the highest entropy to the final compacted set if it does not yet exist.
3. End for
```

Although the main aim of introducing the new approach was to overcome the problem of the dependency on the prediction array calculation, the algorithm seems to produce more compacted solutions. Table 2 reveals a brief comparison between Dixon et al.'s algorithm and the newly proposed one, in which it clearly demonstrates that the latter approach has the ability to tackle the problem of generating a more compacted solution while sustaining its accuracy. Note that in Table 2 the idea of the spoilers is not implemented so as to keep the solution as compacted as possible.

In summary, the importance of the compaction step has been addressed as an essential post-phase in LCS computations. The simplest algorithm was of Dixon et al. (2003) which has a polynomial run-time complexity rather than exponential as in the algorithms of Fu et al. (2002) and Wilson (2001a).

In contrast, the contribution of Wyatt et al.'s (2004) modifications can be considered as a performance improvement over the latter ones. However, since the above algorithms use a simple match algorithm (mainly the XCS one), the acceptance of these algorithms is expected to be adversely affected by the excessive low match performance.

Table 2: comparison between Dixon et al. (2003) and the proposed approach.

Dixon compacted ruleset		The proposed approach compacted ruleset	
Accuracy	Size	Accuracy	Size
98.9%	44	98.1%	36.7

4 CONCLUSIONS AND FUTURE WORK

A new compaction algorithm has been proposed and implemented that overcomes the disadvantages of previous LCS rules compaction algorithms in terms of their poor performance and dependency on prediction array calculations. The results obtained will pave the way for a reflective approach that respects the quality of a rule's selection based on the expert's opinion.

REFERENCES

Holland J., 1976, Adaptation *in* R. Rosen and F. Snell *Progress in Theoretical Biology IV* Academic Press, pp.263-93.

Wilson S., 1995, Classifier Fitness Based on Accuracy. *Evolutionary Computation,* 3 (2): 149-175.

Holland J., 1975, *Adaptation in Natural and Artificial Systems*. Ann Arbor: University of Michigan Press.

Butz, M., Wilson, W., 2000, An Algorithmic Description of XCS, Lecture Notes in Computer Science, 1996: 253 - 272. Springer-Verlag.

Butz, M, Tim, K., Lanzi, L., and Wilson, W., 2004, Toward a Theory of Generalization and Learning in XCS. *IEEE Transactions on Evolutionary Computation*, 8(1): 28—46.

Holmes J., Lanzi P., Stolzmann W., and Wilson S., 2002, Learning Classifier Systems: New Models, Successful Applications. *Information Processing Letters,* 82 (1): 23-30.

Wilson S., 2001a, Compact Rulesets From XCSI. *In Fourth International Workshop on Learning Classifier Systems* (IWLCS-2001). pp. 197-210, San Francisco, CA

Wilson S., 2001b, Mining Oblique Data With XCS. *in* Lanzi, P. L. Stolzmann W. and S. W. Wilson *Advances in Learning Classifier Systems. Third International Workshop (IWLCS-2000),* pp. 253-272, Berlin Springer-Verlag .

Bacardit, J, Butz, M., 2004, Data Mining in Learning Classifier Systems: Comparing XCS with GAssist, in *Advances in Learning Classifier Systems (7th International Workshop, IWLCS 2004),* Seattle, USA, LNAI, Springer.

Blake C., Merz C., 1998, UCI Repository of Machine Learning Databases [online]. Irvine, *CA: University of California, Department of Information and Computer Science*. Available from: http://www.ics.uci.edu/~mlearn/MLRepository.html [Accessed 2/2004].

Bernado E., Llora X., Garrell J., 2002, XCS and GALE: a Comparative Study of Two Learning Classifier Systems on Data Mining *in Advances in Learning Classifier Systems, 4th International Workshop, volume 2321 of Lecture Notes in Artificial Intelligence,* Springer, pp.115-132.

Kovacs, T., 1997, XCS Classifier System Reliably Evolves Accurate, Complete, and Minimal Representations for Boolean Functions. *In* Roy, Chawdhry and Pant (Eds)*, Soft Computing in Engineering Design and Manufacturing (WSC2)*, pp. 59-68. Springer-Verlag.

Dixon, P., Corne, D., Oates, M., 2003, A Ruleset Reduction Algorithm for the XCS Learning Classifier System, in Lanzi, Stolzmann, Wilson (eds.), *Proceedings of the 3rd International Workshop on Learning Classifier Systems*, pp.20-29. Springer LNCS.

Fu, C., Davis L., 2002, A Modified Classifier System Compaction Algorithm. *GECCO-2002*. pp 920-925.

Wyatt, D., Bull, L., Parmee, I., 2004, Building Compact Rulesets for Describing Continuous-Valued Problem Spaces Using a Learning Classifier System. *In* I. Parmee (ed) *Adaptive Computing in Design and Manufacture VI*. Springer, pp235-248.