Intelligent Exploration Method for XCS

Ali Hamzeh
Iran University of Science and Technology
School of Computer Engineering
Room # 32, CE School, IUST, Narmak,
Tehran, Iran
Hamzeh@iust.ac.ir

Adel Rahmani
Iran University of Science and Technology
School of Computer Engineering
Room # 32, CE School, IUST, Narmak,
Tehran, Iran
Rahmani@iust.ac.ir

Abstract
Exploration/Exploitation equilibrium is one of the most challenging issues in reinforcement learning area as well as learning classifier systems such as XCS. In this paper[1], an intelligent method is proposed to control exploration rate in XCS to improve its long-term performance. This method is called Intelligent Exploration Method (IEM) and is applied to some benchmark problems to show advantages of adaptive exploration rate for XCS. It

Categories and Subject Descriptors
I.2.6 [**Artificial Intelligence**]: Learning- *Parameter Learning*.

General Terms
Algorithms, Theory, Performance.

Keywords
Learning Classifier Systems, XCS, Exploration, Exploitation, Adaptive Exploration Rate.

1. Introduction
Learning Classifier System (LCS) is a Machine Learning technology that was introduced by John Holland in the paper "Cognitive Systems based on Adaptive Algorithms" [1] for the first time. In this system, an agent learns its environment by applying some actions and gathering relevant rewards or punishments as a guideline for its internal environmental model that has been designed as a rule based system.

2. LCS and XCS in Brief
In this system, agent detects environment state by its sensors and then it chooses the more suitable action due to its current and previous states and applies it to the environment. After applying this action, the environment may change and some kind of reward or punishment may be waiting for the agent.

[1] This paper is the short version of a full paper that will appear in the post workshop proceeding of IWLCS 2005.

One of the major weaknesses in LCS is its Credit Assignment and Fitness Calculation Method. One modification of LCS to solve some problems of these two parts is proposed by Stewart W. Wilson: "Accuracy based Classifier System" (XCS) [2]. The main characteristic of XCS is its new method to calculate individual's fitness and distributing environmental rewards.

3. Exploration/Exploitation Relationship
It is notable that LCS's are from the Reinforcement Learning area. One of the most important advantages of Reinforcement Learning techniques, such as XCS, over other learning approaches is the ability of exploring problem environment.

To illustrate this issue, we refer to a famous AI problem, k-armed bandit, which described in many references such as [3]. This problem shows Exploitation/Exploration Relationship (EER) very clearly. Agent may believe that a specific arm has high payoff probability. Shall it choose this arm or choose to explore other arms to find a better one?

Answer to this question depends on many parameters, such as game duration, amount of exploration in past experiments and so on. This problem also exists in action selection procedure of XCS. Due to online performance measuring in XCS, major unanswered question is to creating equilibrium between selecting winner action with respect to agent's previous experiments or let the agent to explore its environment to find better rules for further actions.

It seems that agent must create equilibrium between these two strategies, but major problem is to create this balance and minor but important one is winner selection strategy in exploration phase. Should this strategy be pure random selection or must utilize some of gathered knowledge of the agent's experiments?

Let us inspire the current architecture to find its answer to above questions: This implementation based on "Algorithmic Description of XCS" [4] by M. Butz and S.W. Wilson. In this implementation balance between Exploration and Exploitation is created using a P_{exp} constant.

This P_{exp} is used to determine probability of exploring environment. This probability would be set static during agent's life cycle and commonly is equal to 0.5. Therefore, using only a constant parameter creates this balance. Considering second question, [4] uses pure random policy to select action winner in exploration phase and no other parameter such as *Fitness* or *Strength* is involved in selection procedure. In this paper, we propose an adaptive intelligent technique to create the balance between Exploration and Exploitation.

4. Intelligent Exploration Method

Proposed method is based on this theorem that static rate of exploration could not be accurate in the agent's life cycle. For example in the beginning of the learning procedure due to the lack of experience in the environment, acting with respect to previous experiences has no difference with pure random strategy. Therefore, in the beginning phase, existence or lack of exploration has no effect on online and long-term performance. However, in the middle phase, exploration helps us to find more information on the search space and may cause better long-term performance and at last, in the final phase, exploration can help us to escape from local optima but also it may cause some kind of disturbance in agent's knowledge about the world and can reduce agent's performance.

It seems that adaptive changes in exploration probability can improve XCS performance. Due to this theorem, we designed a system called IEM. It tries to propose suitable exploration rate with according to its information about the agent's performance and the environmental changes. IEM tries to distinguish beginning, middle and final phases to propose proportionate exploration probability for each phase.

IEM is a controller, which designed to process some input data and tries to create EER equilibrium by proposing proper exploration probability (P_{exp}). In figure 1, the architecture of IEM is shown.

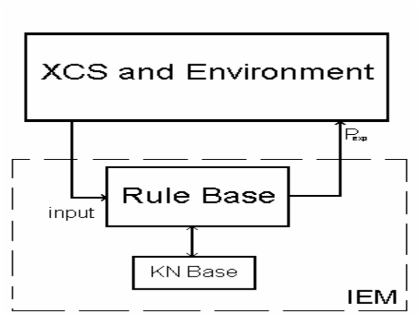

Figure 1. IEM Architecture

IEM is a controller, which designed to process some input data and tries to create EER equilibrium by proposing proper exploration probability (P_{exp}).

Different parts of IEM are as follows:

- Inputs: which consist of:
    - o XCS success ratio from the beginning. (*PERF*)
    - o Number of exploration epochs / Total number of epochs. (N_{exr})
    - o Number of exploitation epochs / Total number of epochs. (N_{exp})
    - o Best classifier fitness in XCS rule base. (F_{best})
    - o Mean fitness of XCS rule base individuals. (F_{mean})
    - o Fitness variance of XCS rule base individuals (F_v).
    - o Mean *Hamming Distance* of all individuals in XCS rule base with the fittest one. (D_{mean})
    - o *Hamming Distance* variance. (D_v)[2]

These inputs are chosen due to these reasons:

- First parameter determines overall picture of XCS success.
- Second and third parameters show XCS age and state of EER equilibrium.
- Forth to eighth parameters are selected to determine the state of XCS internal evolutionary process based on proposed parameters in [5].

This method has a experimentally gathered rule base which is explained in full paper.
Finally IEM is applied to some benchmark problems such as MP6, MP11 and MP20 [6] and its performance is compared by XCS with static exploration rate. As we expect, this modification can improve XCS's performance significantly in hard problems (such as MP20). It is notable that results are shown and discussed in full version of this paper.

5. Conclusion

As we described before, it seems that static exploration rate is not suitable for XCS and an adaptive one can improve XCS performance. To inspire this idea, we proposed an intelligent system called IEM. It is a fuzzy controller designed to propose exploration rate. IEM is added to XCS and the new system, called XCSI, is applied to some benchmark problems. Gathered results approve our theorem about usefulness of adaptive exploration rate for XCS performance. The main weakness of proposed method is its static rule base. Some other researches are ongoing to add learning capability to IEM.

6. References

[1] J. H. Holland, J. S. Reitinan. (1998) .*Cognitive Systems Based on Adaptive Algorithms*. In D. A. Waterman and F. Hayes-Roth, editors, Pattern-directed inference systems. New York: Academic Pr 1978. Reprinted in: Evolutionary Computation. The Fossil Record. David B. Fogel (Ed.) IEEE Press, 1998. ISBN: 0-7803-3481-7

[2] S. W. Wilson. (1995). Classifier Fitness Based on Accuracy. .Evolutionary Computation 3(2):149—175, 1995.

[2] All of these values are normalized.

[3] L. P. Kaelbling, M. L. Littman, A. W. Moore. (1998). *Reinforcement Learning: A Survey*, Artificial Intelligence Research, vol. 4, pp. 237—285.

[4] M. V. Butz, S.W. Wilson. (2002). *an Algorithmic Description of XCS.* Journal of Soft Computing, 6(3-4):144-153, 2002.

[5] N. Parsa. (2005). *A Fuzzy Learning System to Adapt Genetic Algorithms*, MSc. Thesis, Iran university of Science and Technology.