# A Hybrid Intelligent Learning Algorithm in MAS

SHUJUN ZHANG , QINGCHUN MENG , WEN ZHANG , CHANGHONG SONG
Computer Science Department
Ocean University of China
Computer Science Department of Ocean University of China, 266071
CHINA

*Abstract:* Machine learning is a major branch of AI and a main research direction of multi-agent systems (MAS). With the emergence of various new complex systems, agent's individual ability and the system's intelligence level are urgently to be improved. Agent learning methods and agent architecture are studied in this paper and a new hybrid intelligent learning algorithm based on multi-agent is proposed. The application and results in the soccer robot simulation system show the feasibility and efficiency of this algorithm.

*Key-words:* multi-agent systems (MAS); hybrid intelligent learning; robot soccer

## 1. Introduction

Machine learning has been a main research direction in multi-agent systems (MAS). For a long time, some research work has been carried out on a single agent learning technique and now with the development of such disciplines as Artificial Intelligence and Intelligent Control, many mature agent learning techniques have been proposed, such as Artificial Neural Network (ANN), Genetic Algorithms (GAs), Reinforcement Learning (RL) [1，2] and so on. But these techniques are mainly used for individual agent learning; each agent's learning is separate even in multi-agent environment. In order to solve more complex problems, collaborative learning in MAS[3], as a new focus, has appeared. A system requires all the agents to cooperate to complete some complex tasks that can't be finished by a single agent. Therefore the following two points become the major research content to MAS:how to take effective actions through learning when multi agents work together;how this cooperative behavior evolves over time to adapt the dynamically changing environment[4].

At present, the robot soccer system is the most challenging subject in AI field, especially the simulator, a good research platform to MAS and machine learning[5]. Such collaborative, adversarial, real-time and complex systems require every agent to show effective performance not only autonomously but also as a member of a team[6]. Therefore, each of them should have both individual skills and collaborative learning ability. This paper studies individual learning and group learning, proposes a hybrid intelligent learning algorithm based on multi-agent, the corresponding agent architecture and applies it to the robot soccer system.

## 2Individual learning and group learning

Each agent in MAS has its cognition and behavior. In general, it is at two correlative layers at the same time: individual layer and group layer. Every agent is a balanced combination of both layer roles. Generally, agents' learning methods fall into two categories,

- ***Individual learning:*** independent learning for a single agent. The agent learns the ability of interacting with the environment, such as basic actions, commands and communication.

- ***Group learning:***The agent learns group behaviors, i.e. the interactive and cooperative ability with other agents in the same system. The training procedure is carried out in a complex environment made up of multiple agents.

Reasoning based on a single agent has many shortcomings because independent solving ability of any single agent is limited and there're too many factors to be considered simultaneously. In general, individual learning only handles local objectives and local information relating to the agent itself to perform autonomously, so there're no connections between the agents. If the system only focuses on the improvement of a single aim (agent's individual ability) while ignoring interaction of agents, collaboration can't be achieved. However, interaction of agents is the essential terms of learning in MAS.

On the other hand, simplex group learning sometimes can't take agent's particular needs into account. For example, when improvement of individual ability and integral performance conflict, group learning may result in sacrifice of agent's individual ability, because it stresses the efficiency of the whole system.

In order to conquer their respective shortcomings, a hybrid intelligent learning algorithm based on multi-agent is proposed, allowing for interaction between agents, without the loss of the personality and feature of each agent. This is a multi-agent interactive learning method and enables every agent to bring its intelligence and autonomous behavior into full play so as to cooperate better with other agent in the system[7].

# 3 Hybrid intelligent learning algorithm (HILA) based on multi-agent

## 3.1 Preliminaries

There're two kinds of behaviors in MAS: individual behaviors and group behaviors. Only when an agent's belief and objective relate to other agents and its behavior is exactly based on such belief and objective, should the behavior be called a group behavior. That is, whether a behavior of an agent belongs to a group behavior doesn't depend on external description of the behavior itself but on whether it involves and affects other agents[8]. Otherwise, the behaviors that only involve but don't take any effect on other agents belong to individual behaviors. New algorithms should distinguish individual behaviors and group behaviors in order to carry out learning at different levels and complexities. According to their different objects, group behaviors can also divide into cooperative behaviors and adversarial behaviors.

In addition, there're more problems to be considered:

**(1) Difference and interaction between agents**

Agents in distributed open environment are usually heterogeneous, dynamic and unpredictable. These features render their collaboration much difficult and complex. As the whole performance and efficiency of the system have close relationship with agents' division of labor and coordination, learning process must allow for the difference between agents and dispersivity of concerted control. Many independent agents gathering together don't necessarily form a system because a system must be an organism in which the way of agents' interaction is defined by the environment and agents' behavior modes.

One of agents' cooperative methods in MAS is distributed control based on local knowledge. It enables each agent to have certain autonomy. In comparison with centralized control, this method increases flexibility and alleviates the bottleneck of control. But if every agent's performance is constrained by local and incomplete information (such as local objective and local plan), global concert is hard to be achieved. Therefore, it's very important to embed collaborative knowledge in each agent. This knowledge includes abilities, objectives, plans, interests, behaviors of other agents and mutual dependent information. Through information exchange, interactive agents can change the environment where they exist and influence other agents' learning. In particular, when many agents as a team try to perform a learning task which a single agent can't, that is, to carry out group learning, interaction becomes the key problem[9].An effective learning algorithm must consider the difference and interaction between agents to assign learning tasks better and adapt to the dynamically changing complex environment.

**(2) Multi-agent negotiation and automatic negotiation**

Negotiation is a concrete form of agent interaction. Since every agent in the MAS is autonomous, there may be conflicts if each of them executes systematic tasks only according to its own objective, knowledge and ability. Therefore, learning algorithms must take account of negotiation and automatic negotiation between agents in order to enable the agent to reason based on other agents' believes, further more, to learn from their behaviors and improve their mutual understanding and cooperation. Embodiment in the algorithm is that the effect on the environment and feedback the agent gets from the environment after it has executed an action at a certain time step in a learning process are related to not only the agent's own behavior but also other agents' behaviors and states.

If different agents have different viewpoints about the present optimal strategic action of the system based on everyone's perceptive information and reasoning, negotiation is urged[3].The process of agents' negotiation and state transformation is as follows. At first, the agent is in the state of waiting for operation. When the operating conditions are satisfied or it receives other agents' requests, it will negotiate with them according to certain rules and go to negotiation state. There're two kinds of results of negotiation. One, terms or rules can't be met with so the agent will refuse to take the action negotiated and go to refusal state. Then after a certain time it'll return to the original state. The other, with all the terms and rules satisfied the agent will take according action and go to operation state. During the operating process, if the agent comes across an exception it will restore to operation. If failing to restore, it'll go back to the original state; otherwise, it'll continue to operate until the task is finished.

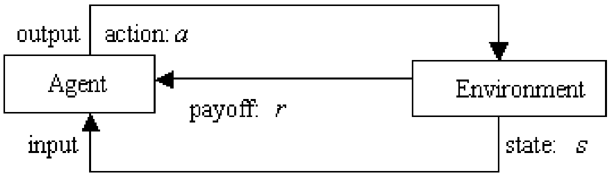


Fig.1 Principle of Reinforcement Learning

### 3.2 Implementation of HILA

The hybrid intelligent learning algorithm proposed in this paper is based on reinforcement learning (RL). Fig.1 illustrates the basic principle of RL. The agent perceives the state information from the environment, chooses proper actions according to its own strategy, thereby changes the environment and gets reward or stimulus as a contribution function of the new state relative to the agent[9]. The agent's target is to search out a strategy for action choice that can get it a reward as large as possible on the basis of exploring the state space of the environment.

Our learning system is layered and the agent uses RL directly to learn individual behaviors, so the following mainly discusses group-behavior learning. The algorithm extends the idea of RL, considering and making use of multi-agent interaction and its changes while inheriting the same method of adjusting the decision-making probability with feedback. So it's a dynamic collaborative model in which "individual optimization" loses its sense because every agent 's feedback not only depends on itself but also on other agents' decisions. The learning target of the system is to look for an appropriate strategy, maximizing the overall reward the system can get in the future.

Now we give mathematic description of the algorithm. Suppose MAS is made up of n agents, expressed: MAS = {$Agent_1$, $Agent_2$, ⋯, $Agent_n$}. Let $E$ denote the external environment state set of the system. An agent is denoted in the form of $< S, A, f >$,where $S$ is the agent's state set; $A$ is the action set; $f$ is its planning function. Let $P(s, a, s')$ denote the state transformation probability, i.e. the probability of moving to state $s' \in S$ after the agent executes action $a \in A$ in the state $s \in S$. Agents' collaborative knowledge, i.e. correlation of the interactive behaviors between the agents and relationship between individual targets and integral targets, are denoted by fuzzy matrixes $H(t), H'(t)$ respectively. Because the two are both matrix functions of time $t$, they can reflect the dynamic changes of multi-agent interaction and changes of the relationship between every agent and continuously changing targets.The hybrid intelligent learning algorithm during a given cycle $t$ is as follows:

(1) Set the external environment state $E(t)$ randomly;

(2) Agent$_i$ $(i = 1,2,\cdots,n)$ observes $E(t)$ and accordingly updates its own state $S_i(t)$ based on $H(t)$ and $H'(t)$;

(3) All the agents learn concurrently: Agent$_i$ uses state transformation probability $P(S_i(t), a_i, s'_i)$ and its own decision-making function $f_i$ to choose an action $a_i$;

(4) The environment produces an external reinforcement signal $R(t)$ after all the agents actions;

(5) Agent$_i$ figures out $W_i(t)$ according to $H(t)$ and $H'(t)$ ,where $W_i(t)$ denotes the weight associated with Agent$_i$ relative to the whole system at time t and it satisfies:

$$\sum_{i=1}^{n} W_i(t) = 1 \quad ;$$

(6) Agent$_i$ calculates its own effective reinforcement signal $r_i(t) = W_i(t) * R(t)$ , where $r_i(t)$ represents the reward Agent$_i$ gets at time $t$ when it sets off from state $S_i(t)$ and executes the action the decision-making function selected;

(7) Update $H(t)$ , $H'(t)$ ;

(8) Search out a systemic strategy with which every agent's sub-strategy can enable the system to get the largest discount reward sum (total return), that is, to maximize $\sum_{i=1}^{n} W_i(t) \sum_{j=0}^{\infty} \gamma_j r_i(t)$ ,where $\gamma$ is discount factor, reflecting the degree of the time's influence on the repayment, and usually it is assigned a value slightly smaller than 1.0.

This algorithm adopts layered idea and separates individual-behavior learning from group-behavior learning. Through the collaborative knowledge among agents in the same system, using the weighted method to adjust agents' interaction and mutual effects with the environment, the agent can learn with different emphasis, their strong points developed and also they're trained as a whole. Consequently, agents' individual skill and the system's performance are both improved.

## 4 Application in robot soccer system

Robot soccer has become a standard research platform of MAS[7]. Agent learning in MAS can also be applied to this field. In the following part, we introduce HILA based on multi-agent to the robot soccer simulation system and present the experimental results.

First, based on layered idea, agent's tasks are decomposed into different layers (Fig.3) to use HILA. From the lower to the top: (1) is individual learning, the agent learns individual skills from the world model; (2)(3) are group learning: the agent learns group behaviors based on the world model and skills it has already acquired. In details, individual behaviors mainly include positioning, dribbling, kicking etc. Collaborative behaviors (group behaviors in the same team) include passing and interception. Adversarial behaviors include shooting, goaltending, defending and so on.

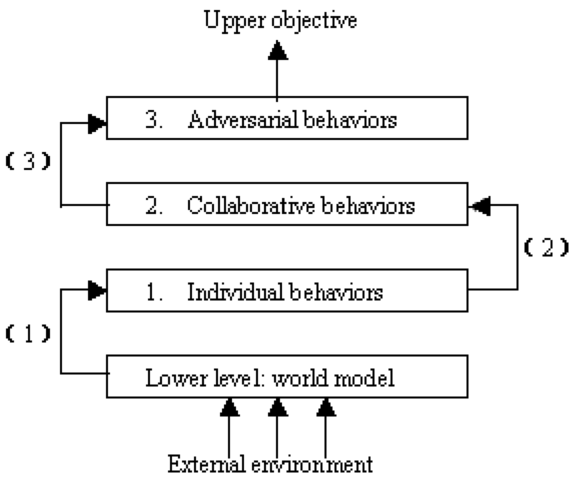


Fig.2 Task decomposition in layered learning[6]

## 4.1 Application samples

➢ **Passing:** to all the players

Factors such as the passing object, trajectory and choice of the passing opportunity are to be considered. In a soccer game passing is the major cooperation for the players. Both teams want to keep control of the ball in the front field to help them goal. Passing involves the whole situation of the field and there's hardly a criterion to judge whether a pass succeeds or not. In our algorithm, let fuzzy matrix H(t) denote the probability with which the player who's handling the ball may pass it to others; fuzzy matrix H'(t) denote the correlation between the agent and the global target. Environmental state vector is make up of the positions and velocities of 22 players and the ball; all the basic commands constitute action set, to perform multi-agent learning of team behavior. If the pass succeeds, external reinforcement signal R(t) is assigned 1.0; the weights connected to the controlling-ball player and the receiving teammate Agenti are relatively larger; so are the specific gravities of their valid reinforcement signals with respect to the overall reward of the system. Therefore, they two are the major learners on this occasion.

➢ **Goaltending:**

The goalie should observe and make statistics of the shooting habit of the opponent forwards, predict their shooting tendency and adjust his own interception point to the ball. There're four kinds of geometric defending strategies. We use HILA to perform learning and train the goalie in various environments. In the beginning, only one forward is used. Training is carried out for many times with the feedback given by the environment each time: successful payoff is given when the goalie catches the ball, failure payoff is given when the opponent goals. Then add the number of the opponent forwards and improve the difficulty of defending. Learning steps of the goaltending strategy are as follows:

(1). The forward is placed randomly in the field and the goalie is set at his home-position;

(2). The forward dribbles and waits for the opportunity to shoot (on the basis of having acquired the shooting skill);

(3). The goalie chooses a strategy to defend according to present decision-making probability;

(4). If the ball is caught, the goalie succeeds and positive reward is given; if the opponents goal, negative penalty signal is given; otherwise (for example, the ball doesn't enter the goal-area) the signal is zero;

(5). Adjust the decision-making probability according to the feedback. Repeat the learning process until the optimal strategy with the largest state evaluation function is found out.

## 4.2 Experimental results

In order to inspect the optimization of the control strategy, we make statistics of the times the goalie succeeds in defending every 100 learning cycles and get the experimental data as shown in Fig.4 (the abscissa denotes the time order number in learning). We can see that the goalie's success times are increasing with the learning time. (The curve's oscillation is due to the uncertainty of the environment and the randomization of the algorithm.)

In order to inspect the improvement of the whole system's performance we use simulator team A without the new algorithm to play against simulator team B adopted HILA. The scores of the 10 games randomly selected are shown in Table 1.

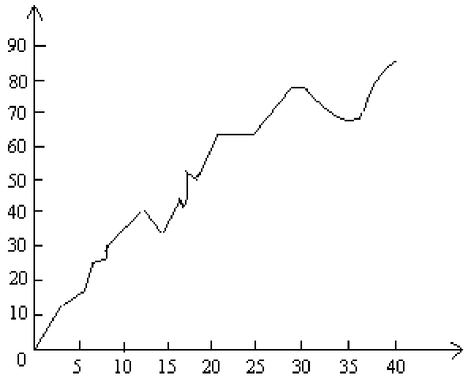


Fig.3 Relationship between the times of success in goaltending and the learning time

Table 1 Results of 10 simulated games between A and B

| times / score | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| A: B | 3 : 11 | 4: 12 | 5 : 13 | 4 : 12 | 3: 14 | 3 : 12 | 2 : 14 | 3 : 13 | 4: 13 | 3: 15 |

## References


1. Zhuang Xiao-dong, Meng Qing-chun, Wei Tian-bin, Robot Path Planning in Dynamic Environment Based on Reinforcement Learning, Journal of Harbin Institute of Technology, 2001.9, Vol.8, No.3
2. Wang Zhi-jie, Fang Jian-an, Shao Shi-huang, A Self-learning Fuzzy Logic Controller Using Reinforcement, Control and Decision, 1997, Vol.12, No.2
3. Wang Li-chun, Li Hong-bing, Chen Shi-fu, Multi-agent Negotiation and Learning in AODE, Pattern Recognition and Artificial Intelligence, 2001, Vol.14, No.3
4. Wu Jian-bing, Yang Jie, Wu Yue-hua, An Evolutionary Multi-agent Robotic System, Journal of China University of Science and Technology, 1998, Vol.28, No.5
5. Li Shi, Xu Xu-ming, Ye Zhen, International Robot Soccer Tournament and Correlative Technique, Robot,2000, Vol.22, No.5
6. Peter Stone , Layered Learning in Multi-Agent Systems, doctor thesis , Carnegie Mellon University, Pittsburgh, 1998
7. Meng Wei, Hong Bing-rong, Han Xue-dong,, Application of Reinforcement Learning to Robot Soccer, Application Research of Computers, 2002
8. Shuai Dian-xun, Gu Jing, A New Algebraic Modeling for Distributed Problem-Solving of Multi-agent Systems (Part 1): Social Behavior, Social Situation and Social Dynamics, Chinese Journal of Computers, 2002.2, No.2
9. Cai Qing-sheng, Zhang Bo, An Agent Team Based on Reinforcement Learning Model and Its Application, Journal of Computer & Development, 2000.9, Vol.37, No.9