Evaluating Learning Automata as a Model for Cooperation in Complex Multi-agent Domains

Mohammad Reza Khojasteh[1] and Mohammad Reza Meybodi[2]

[1] AI & Robotics Laboratory, Computer Engineering Department
Shiraz Islamic Azad University, Shiraz, Iran
mrkhojasteh@persianrobotics.net
[2] Soft Computing Laboratory, Computer Engineering Department
Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran
meybodi@ce.aut.ac.ir

Abstract. Learning automata act in a stochastic environment and are able to update their action probabilities considering the inputs from their environment, so optimizing their functionality as a result. In this paper, the goal is to investigate and evaluate the application of learning automata to cooperation in multi-agent systems, using soccer simulation server as a test bed. We have also evaluated our learning method in hard situations such as malfunctioning of some of the agents in the team and in situations that agents' sense/act abilities have a lot of noise involved. Our experiment results show that learning automata adapt well with these situations.

1 Introduction

As a model for learning, learning automata act in a stochastic environment and are able to update their action probabilities considering the inputs from their environment, so optimizing their functionality as a result.

Also, as a test-bed, we have used the simulated robotic soccer, "SoccerServer2D" [3] in this paper. Robotic soccer is an example of a complex environment that some agents should cooperate with each other, in order to achieve the team's goal [2][3]. In fact, in this paper we have focused on the systems composed of some autonomous agents that can act in real-time, noisy, collaborative and adversarial environments [1].

To do so, we implemented teams composed of 11 agents that learn using learning automata and compared them to similar teams that have no learning capability or use other learning methods such as Q-learning.

In this paper, we have used Learning Automata [4] as our machine learning method. In the coming sections of this paper, we first present our use of learning automata in a complex multi agent domain with presenting some results of our simulations.

Then, we evaluate our learning method by running some simulated plays with minor and major changes in environment parameters. Also, we evaluate our learning method in hard situations such as malfunctioning of some of the agents in the team and in situations that agents' sense/act abilities have a lot of noise involved.

The goal is to investigate our learning method's adaptation with these changes.

G. Lakemeyer et al. (Eds.): RoboCup 2006, LNAI 4434, pp. 410–417, 2007.

2 Cooperation in a Team Using Learning Automata

Our goal in this section is to use learning automata for cooperation among the members of a simulated soccer team with 11 players in order to achieve the team's goal. By now, various machine learning methods such as Q-learning, genetic algorithms, decision trees, behavioral learning, to mention a few, have been used for training the soccer player agents [1]. To our knowledge, this research is the first attempt to use learning automata in cooperation in multi-agent systems.

Because of the large state space in such a complex multi agent domain, it is vital to have a method for environmental states' generalization. In this paper we have used the technique called the "Best Corner in State Square" for generalizing the vast number of states in agent's domain environment to a few number of states by building a virtual grid in that environment [5]. Our experiments in [6] show that by using the "Best Corner in State Square" technique, each agent performs well in determining its own state and consequently, in determining the proper action in that state.

We have also used 8 learning automata for each agent (one automata for each corner in the "Best Corner in State Square" [6]). Also, we have defined 8 actions for each learning automata; sending the ball to the center of one of agent's 8 immediate squares as defined in the "Best Corner in State Square" technique [6].

In our simulations we used 4-3-3 formation for each team for organizing the eleven players in the field. We implemented some teams using fixed structure learning automata, some teams using variable structure learning automata, and a team using Q-learning (as a team that uses another method of learning). Each of the above teams played against the "without learning" team. Note that the "without learning" team in our simulations is like the "learning" teams from every aspect (architecture, states, actions, and even team formation), except that it can't learn from its previous experiences.

In our "learning automata" teams (fixed or variable), the agents determine their current state by the "Best Corner in State Square" technique [5]. Then the agent that possesses the ball performs the action that is advised by the corresponding automata in its state. The agent then percepts its action's result, and gives itself a reward or a penalty depending on that result.

In fact, our agent simply gives itself a reward if the ball has gone toward opponent's goal and one of its teammates (or even itself, in case of a dribble) has chased the ball (as the result of its action). Similarly, the agent gives itself a penalty if the ball has gone toward its team's goal and one of the opponent players has chased the ball (again as the result of its action). In all other cases, the agent does not give itself any reward or penalty and leaves its learned values unchanged.

Note that we have simulated our agents to learn from zero (i.e. without any previous knowledge of the environment before starting the simulation). Also, we have used the agent itself for the judgement about its action's results and this let us have what we call "distributed judgement", again a multi-agent approach.

We should point that we have used a memory depth of 3 for our fixed structure learning automata teams ($L_{2N,2}$, $G_{2N,2}$, Krylov, and Krinky). Also, for our variable structure learning automata teams, we have used ($a = b = 0.1$) for L_{rp}, ($a = 0.1$, $b = 0.0$) for L_{ri}, and ($a = 0.1$, $b = 0.01$) for L_{rep} and Full_ L_{rep}. For our Q-learning team, we have used the TPOT_RL introduced in [1].

The simulation results show that the learning automata teams could defeat the "without learning" team after a few number of training plays. Figure 1, shows one of the results of our simulations.

Note that we have give our results based on the number of games played and each player in our team has a chance of 5 ball kicking (by average) in each game. More results are presented in [6].

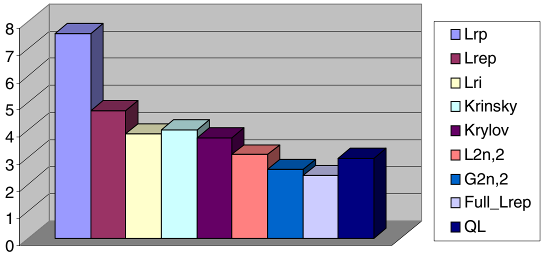

Fig. 1. The ratio of the average goals scored to the average goals received for each team (during 3 test plays), after 15 training plays against the "without learning" team

This figure shows that the variable structure learning automata teams have a better performance in defeating the "without learning" team and so exhibit higher ability of learning in this domain.

We have also discussed the speed of convergence of our learning automata algorithms used by the agents and suggested some techniques for increasing the speed of convergence in [6].

Our simulations [5][6][7] show that learning automata perform well in order to have a cooperative team of agents in a complex multi-agent domain.

3 Evaluation Tests for Teams Using Learning Automata

In previous section, we investigated the efficiency of using learning automata in doing teamwork [5][6][7]. In this section we evaluate our learning method by running some simulated plays with one of the teams that had taken part in previous world RoboCup competitions. Also we investigate the efficiency of learning automata by changing some of the RoboCup SoccerServer2D parameters and observing their effects on our teams' performance. It is necessary to note that our base code is the code of CMUnited98 team [1].

In doing so, we used the team Saloo 2001 [10] that was similar to our team from the agent individual behaviors' (shooting the ball, etc.) point of view. We let our learning automata teams play against this team and observed the results in detail [6].

In this section, we use the teams Full_ L_{rep} and also L_{rep} (both based on L_{rep} automata) for our simulations. Selecting these automata is because of the good results obtained for them in our previous simulations. For more simulations the reader may refer to [5][6][7].

Also in this section, we investigate the performance of our learning method in the environment situations that are more difficult for the agents to adapt with (comparing the situations we have considered so far).

At the end of this section, we have investigated the effects of different formations on the teamwork.

In the first series of simulations, ten consecutive plays between team L_{rep} (learning from zero) and the "without learning" team were simulated. Our goal was to investigate the efficiency of our learning methods and to observe how our team performs against "without learning" team in the presence of noise.

There are several parameters in the RoboCup SoccerServer2D that can be changed [3]. In the first series of simulations conducted, we study the effect of the "rand" parameters (which indicate the amount of noise values) in the RoboCup SoccerServer2D. We changed the parameter "player_rand" from 0.1 to 0.2, the parameter "ball_rand" from 0.05 to 0.1, and at last the parameter "kick_rand" from 0.0 to 0.1. The first parameter mentioned enables us to add noise to the players' movements and the second and the third parameters, add noise to the ball movement and kicking the ball, respectively.

Figure 2 shows the cumulative results of this experiment. As the figure shows, by playing more games, the learning team, adapts itself to the situation more and more and increases its gap with the "without learning" team. This figure shows the efficiency of the proposed learning method when the environment is noisy.

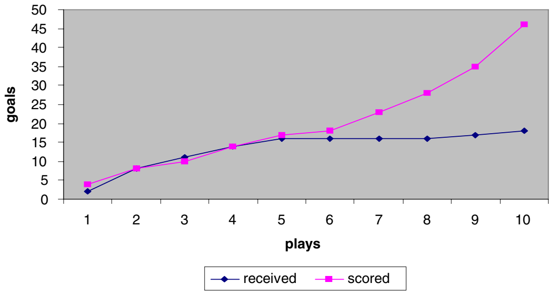

Fig. 2. The number of goals scored versus the number of goals received by the learning team during 10 consecutive training plays versus the team "without learning", as the noise increases

In the second series of simulations, we studied the effects of not using (although single-channel, crowded, and unreliable) communication facility provided by RoboCup SoccerServer2D on the performance of the proposed learning method. We simulated 10 consecutive plays between team L_{rep} (learning from zero) and the "without learning" team. Figure 3 shows the efficiency of our learning method and indicates that by playing more games, the learning team, adapts itself with the situation and increases its gap with the "without learning" team.

In third series of simulations, we eliminated 3 players from the left side of our team. They were player number 2 from the defense line (the left defense), player number 6 from the middle line (the left piston), and player number 10 from the forward line (the left forward) of our team [6]. Our goal was to evaluate the function of our team in the case of failure in some of our agents.

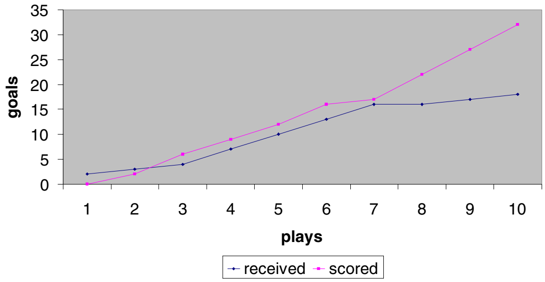

Fig. 3. The number of goals scored versus the number of goals received by the learning team during 10 consecutive training plays versus the team "without learning", without the communication facility between the agents

We simulated 15 consecutive plays between team L_{rep} (learning from zero) and the "without learning" team (with 11 players). Figure 4 shows the results of this experiment. Since each player in our team has a limited freedom around its special post in the field, our simulations show that this elimination causes our team's left side to malfunction. Note that we have used a 4-3-3 formation for our teams.

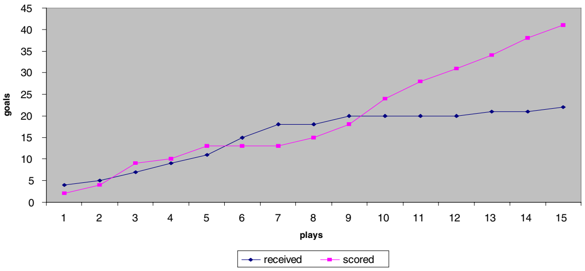

Fig. 4. The number of goals scored versus the number of goals received by the learning team during 15 consecutive training plays versus the team "without learning", with eliminating 3 players from the left side of the learning team

As figure 4 shows, our learning team was able to overcome the absence of its players and adapted itself to fewer number of players and finally defeated the "without learning" team.

In next series of simulations, we made some simulations between team L_{rep} (learning from zero) and team Saloo 2001 and the results of the first seven plays are gathered in table 1. As the table shows, team Saloo 2001 could win all the first games with relatively high goal average (average scored goals of 5.7 versus average received goals of 0.3 in each play), and has an absolute better performance comparing to our team. We conducted more training plays in order to be able to defeat the team Saloo 2001. So, we simulated 150 consecutive plays (equal to 25 hours) between the two teams.

Table 1. The statistics for the first 7 plays between team L_{rep} (learning from zero) and team Saloo 2001

The percentage of possession of the ball for the opponent team (Saloo 2001)	52.6
The percentage of possession of the ball for the our team (L_{rep})	47.4
The percentage of ball movement in opponent's 1/3 of the field	10.5
The percentage of ball movement in the middle 1/3 of the field	47
The percentage of ball movement in our 1/3 of the field	42.5
The maximum continuous time that opponent team has the ball in possession	185
The maximum continuous time that our team has the ball in possession	112
The maximum number of continuous passes between the members of the opponent's team	14
The maximum number of continuous passes between the members of our team	8

Table 2. The statistics The statistics for the last 7 plays (after 25 hours training) between team L_{rep} (learning from zero) and team Saloo 2001

The percentage of possession of the ball for the opponent team (Saloo 2001)	45
The percentage of possession of the ball for the our team (L_{rep})	55
The percentage of ball movement in opponent's 1/3 of the field	24.5
The percentage of ball movement in the middle 1/3 of the field	42
The percentage of ball movement in our 1/3 of the field	33.5
The maximum continuous time that opponent team has the ball in possession	112.7
The maximum continuous time that our team has the ball in possession	134.2
The maximum number of continuous passes between the members of the opponent's team	8.8
The maximum number of continuous passes between the members of our team	12

Table 3. Average wrong actions' percentage for each agent of the team L_{rep} in the first 7 plays versus the last 7 plays (after 25 hours of training) when played against team Saloo 2001

Wrong actions' percentage	
40.1	In the first 7 plays
24.6	In the last 7 plays

During these 150 simulated plays (that their overall results are shown in tables 1-3), our team could improve its performance and gradually move toward “not losing" and finally to continuously “win”.

The statistics of the first 7 plays are given in table 1. Also, the statistics of the last 7 plays are given in table 2. In these simulations, an average scored goals of 3.6 and an average received goals of 0.1 is obtained. In table 3, we give the average percentage of wrong actions done [6] by the learning team's players in the first 7 plays and in the last 7 plays, for the sake of comparison.

It is necessary to remind that our agents try to send the ball toward one of their 8 directions, whichever seems to be better for achieving the team's goal [5][6][7]. This action might (relative to position) seem as a pass, a dribble, a shoot, etc. In team CMUnited98, there are two layers for multi-agent behavior (pass evaluation that is trained offline using decision trees) and for team behavior (pass selection that is trained online using a method based on Q-learning which uses the output of the previous mentioned layer as the input) [1].

Our method for learning has combined the above two layers into one layer. We've done some offline training before our team plays against another team. We haven't separated "pass evaluation" from "pass selection". In fact we are dealing with actions that a player chooses and whether or not the action chosen is a right action. We aren't involved with "pass" as a separate problem to solve. Instead, we have looked at the problem of "cooperation between our agents".

As an end to this research, we investigated the effects of different team formations on the agents' cooperation. Note that for all the simulation presented so far we have used 4-3-3 team formation. We created similar teams (L_{rep}) but with different formations 4-4-2, 3-6-1, 4-3-3, 3-5-2, and 3-4-3 and then simulated a series of plays as a tournament between them. By tournament, we mean that we let each of the above mentioned teams play against all other teams and gathered the results. Figure 5 shows the results of these simulations [6].

As the results show it is very important to have a proper team formation in order to achieve a good team performance. Figure 5 shows that the highest number of goals scored was by the team with the formation 3-5-2 and the least number of goals received was by the team with the formation 4-4-2.

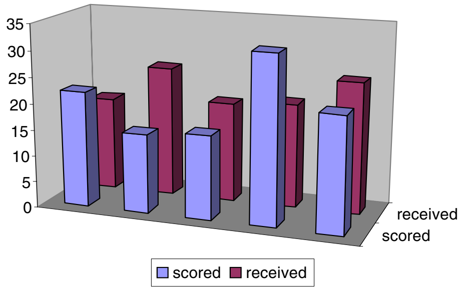

Fig. 5. Scored and received goals by the learning automata teams (L_{rep}) for different team formations playing against each other in a tournament (from left to right 4-4-2, 4-3-3, 3-6-1, 3-5-2, and 3-4-3)

For simulations in which other learning automata such as "Estimator algorithm" [8] and "Discretized Pursuit Learning Automata" [9] are used, the reader may refer to [6]. For a discussion about the speed of convergence of the proposed method and also methods to improve the speed of convergence the reader may again refer to [6].

4 Conclusion

We used learning automata for successful production of a series of actions for agents that were members of a team, such that the resulting team could act well in multi-agent, adversarial, noisy, real-time, and most important collaborative environments.

The methods introduced are general methods that can be implemented, applied, and used in other domains or other test-beds with minor changes.

At last, we evaluated the efficiency of learning automata in cooperation among agents that are seeking a common team goal by implementing some evaluation tests and observing the results.

Our experiments, showed that learning automata adapts itself well with major and minor changes in the environment parameters and also in hard situations such as malfunctioning of some of the agents in the team and in situations that agents' sense/act abilities have a lot of noise involved.

References

1. Stone, P.: Layered Learning in Multi_Agent Systems, PhD thesis, School of Computer Science. Carnegie Mellon University (December 1998)
2. Kitano, H. (ed.): RoboCup-97: Robot Soccer World Cup I. Springer, Heidelberg (1998)
3. Andre, D., Corten, E., Dorer, K., Gugenberger, P., Joldos, M., Kummenje, J., Navaratil, P.A., Noda, I., Riley, P., Stone, P., Takahashi, R., Yeap, T.: Soccer server manual, version 4.0, Technical Report RoboCup –1998-2001, RoboCup (1998)
4. Narendra, K.S., Thathachar, M.A.L.: Learning Automata: An Introduction. Prentice Hall, Inc., Englewood Cliffs (1989)
5. Khojasteh, M.R., Meybodi, M.R.: The Technique "Best Corner in State Square" for Generalization of Environmental States in a Cooperative Multi-agent Domain. In: Proceedings of the 8th annual CSI computer conference (CSICC' 2003), pp. 446–455, Mashhad, Iran (February 25–27, 2003)
6. Khojasteh, M.R.: Cooperation in Multi-agent Systems using Learning Automata, M.Sc. thesis, Computer Engineering Faculty, Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran (May 2002)
7. Khojasteh, M.R., Meybodi, M.R.: Using Learning Automata in Cooperation among Agents in a Team. In: Proceedings of the 12th Portuguese Conference on Artificial Intelligence, IEEE Conference Publication Program with ISBN 0-7803-9365-1 and IEEE Catalog Number 05EX1157, University of Beira Interior, pp. 306–312, Covilhã, Portugal (December 5th-8th, 2005)
8. Thathachar, M.A.L., Sastry, P.S.: A New Approach to the Design of Reinforcement Schemes for Learning Automata, IEEE Transactions on Systems, Man, and Cybernetics, Vol. SMC-15(1) (Janaury/February 1985)
9. Oomen, B.J., Lanctot, J.K.: Discretized Pursuit Learning Automata. IEEE Transactions on Systems, Man, and Cybernetics SMC-20(4) (July/August 1990)
10. Noda, I.: Team Description: Saloo, AIST & PREST, Japan (2001)