Nemesis Team Description 2009

Mehrab Norouzitallab[1,i], M.M Korjani[1], S.M.A Salehizadeh[1], Ahmad Pourshoghi[1],Danial Khashabi[1], Abbas Razaghpanah[1], Pooria Kaviani[2], and Meisam Vosoughpour[3]

[1] Amirkabir University Of Technology, Hafez Ave, Tehran-Iran
[2] Sharif University Of Technology, Azadi Ave, Tehran-Iran
[3] University Of Tehran, Enghelab Ave, Tehran-Iran

{m.norouzitallab, korjani, smasalehizadeh, pourshogi, danyal.khashabi, a.razzaghpanah, pooria.kaviani, vosoughpour}@Gmail.com

Abstract. In our study, we tried to develop our teams in such a way that machine learning techniques and advanced artificial intelligence tools have the main role in improving skills and increasing team performance. We consider soccer simulation platform as an uncertain and dynamic environment, so we develop learning algorithms according to this important feature and agent's partial observability.

1 Introduction

The Nemesis team was established in 2004. The team's been improving and consolidating with newer skills and sophisticated yet highly effective strategies.

In the beginning, the team was developing on Mersad source codes. But the problems with the massive time consumed by low-level related programmings along with revelation of speed and efficiency issues with the base code convinced the team to move to Hellios base as a base code with better low-level foundations and continuous project activity.

Hence, the members' researches to develop scientific projects involving use of MATLAB due to great deal of capacities provided ease of use caused by toolboxes and easy-to-code yet powerful language to implement complicated algorithms.

However, this is a shortened Team Description Paper (TDP) to try and display the team's main ideas and their efforts to put them in effect. The complete version of the TDP, Nemesis 2008 source codes, pseudo-codes and MATLAB codes of some implemented algorithms are available to download at the team's homepage:
http://mnt.ir/nemesis

2 POMDP Framework

2.1 Kick Skill

Finding a good kick routine is a very important task. Within our RL framework, this is done by the agent in a partially observable environment. According to the fact that the agents cannot often directly access actual states of the environment, but can only get observations, which may be partial, from them, serious computational difficulties arise in estimating unobservable states. It is provided with 500 parameterized instances of the kick command (direction discrete in 100 steps and power discrete in 5 steps) together with 36 instances of the turn command. This makes an overall of 536 actions, from which the agent can choose one per cycle according to its observations. The learning agent is controlled based on one-step-ahead prediction using opponent agents' models. It is difficult, however, to apply this method without any approximation because soccer 2D is a large-scale multi-player environment, and then the utility prediction should involve intractable integrations. To overcome this intractability, sampling is performed over a subspace under the assumption that each opponent player will perform the action which is detrimental to the learning agent [3]. The utility function is given by the following expectation with respect to the predictive distribution:

$$U(a_t, H_t) = \left\langle R(a_t, u_t) + V(s_{t+1}) \right\rangle \quad (1)$$

$$\left\langle f(s_{t+1}) \right\rangle \equiv \sum_{s_{t+1} \in S_{t+1}} P(s_{t+1} \mid a_t, H_t) f(s_{t+1}) \quad (2)$$

$R(a_t, u_t)$ denotes an immediate reward, which is $R(a_t, u_t) = 0.5 - n$ when the agent gets n penalty points (n may be 0) after the *t-th* cycle. The constant bias 0.5 is attached to make the sum of all agents' rewards zero. In our study, a normalized Gaussian network (NGnet) is used for approximating the value function V, we first define a *consistent state* as follows: for a given action sequence, a state which may realize the action sequence is said to be a consistent-state. State can be represented with five elements: the ball's relative position, velocity, the length of the desired ball speed, the player's velocity, the player's body facing. For more details readers are referred to the 2008 Nemesis team description [13].

3 Positioning and Team Formation

A new framework is proposed for formation strategy where every agent is able to extract the expert knowledge via observing his behavior [11]. We use Fuzzy ARTMAP as knowledge based neural network for extraction of expert knowledge. Thereby, an intelligent model of expert behavior can be built for formation strategy by combination of expert knowledge and low-level behavior.

The time required to develop a model of high-level behaviors for such agents could be significantly reduced. Therefore, the only time consuming part is execution of low-level behaviors. This method allows agents to learn knowledge from unwilling experts and/or experts who are unable to convey their knowledge explicitly to a third party. So we can extract other teams' formation strategy.

The target of the formation is to reach to the desired position. Each agent must first find its position and then consider the position of the ball and other agents. After it gathers the local information about the environment, the agent must conclude the desired position and implement some low-level behaviors to reach the goal. The position of each agent and the ball is normalized to the interval [0,1] and the context of learning algorithm is selected. Then the Fuzzy ARTMAP neural network is used to learn the structure of the data. ARTMAP-based framework has a number of additional advantages over other model, such as transparency, no more catastrophic forgetting, fast learning, higher performance and more robustness. It also stabilizes with fewer training exemplars.

4 Satisficing Applied to Soccer Agents

4.1 Brief introduction to satisficing

In many real life situations, it becomes more of a challenge to coordinate multi-agent systems when each agent has a different view or understanding than the others. A good decision-making procedure in autonomous multi-agent systems should be able to take into consideration the preferences of others when forming its own preferences.

In the context of decision making, satisficing, introduced by the economist Herbert Simon, allows for choosing options that are good enough. According to Simon, as one is searching the range of options, the first option that is good enough should be chosen if it is above a certain aspiration level. No other option will be considered after that point [9]. Stirling's version of satisficing is a promising approach to decision making in autonomous multi-agent systems for it includes the concept of "good enough" and allows for consideration of other's preferences, both of which are needed for a good cooperative/competitive system. In many competitive situations, an agent or group of agents can win by just a fraction of a second or by one point. It therefore seems that real-time optimization is important in competitive situations, such as simulated soccer. In order to incorporate real-time optimization with satisficing, we applied a novel optimization algorithm "Individual Particle Optimization" which has recently been submitted by Salehizadeh et al [10]. IPO in conjunction with satisficing can be shown to be a promising approach to global optimization.

4.2 Satisficing and Soccer Players

In RoboCup Soccer Simulation, each soccer player is run as a separate process that uses lower-level functions to send to and receive from the RoboCup server through a sockets protocol.

After moving to initial positions, each player during each simulator step (cycle) decides many options (kick the ball, intercept, pass, shoot, going to a particular position …).

The Rejectability of each option in our implementation is always based on the player's current stamina. The Rejectability is initially set to

$$1 - \frac{(curStam - recStam)}{(\max Stam - recStam)}$$

If the player's stamina drops below "*recStam*", the player will be unable to recover its full stamina. The variable "*maxStam*" is the maximum amount of stamina a player can possess. The variable *"curStam"* is the player's current level of stamina.

The important thing to notice is that the Rejectability is inversely proportional to the current stamina. After calculating all the selectabilities and rejectabilities and discarding non-satisficing options, if an option is satisficing, its Rejectability is multiplied by this option's importance factor Each option has its own boldness factor such that the Rejectability becomes more or less important in the utility function.

5 Coverage Control

The method we are describing here is a part of a hierarchical decision system, so the output of this part is just optimal coverage of agents and can be affected by outputs of other parts of system such as man marking, offside trap etc.

We expect our defenders to cover a/n, midfielders to cover b/n and forwards to cover c/n of the field at time t. Terms a, b, c can be constant during the match, can be changed when ball possession is changed or even can be a (t), b (t) and c (t) so at each moment have different amounts during the match due to manager orders or be adapted due to opponents' style of playing.

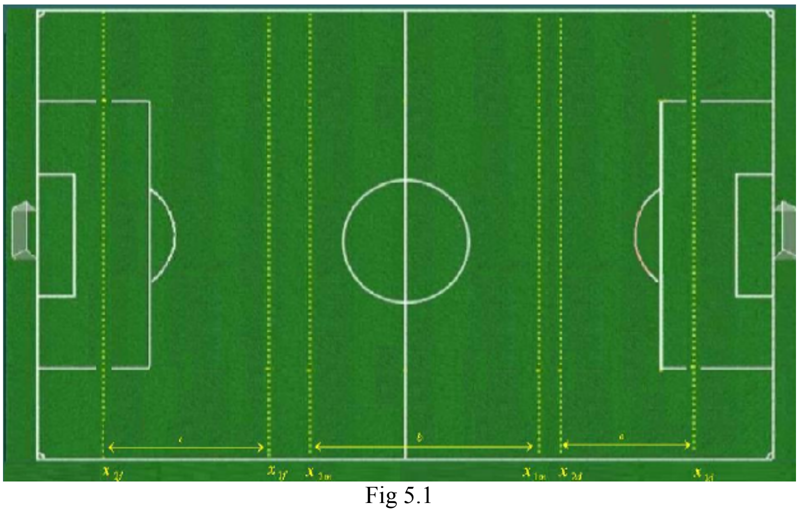

Fig 5.1

As you can see in Fig 5.1, Each section (a, b, c) is limited by two vertical line x_1, x_2 and two horizontal line of the field. Now according to initial team system, players of each line should cover their specific section. For example in a (3-4-3) system the section between x_{1d}, x_{2d} should be covered by 3 defense players and section between x_{1m}, x_{2m} by 4 midfielders and section between x_{1f}, x_{2f} with 3 forward players.

It should be mentioned that different arrangements of these vertical lines would result in three types of coverage: exact coverage (with no gap), coverage with gap and overlapped coverage. For example if $x_{1m} = x_{2d}$ there will be a no-gap coverage between defenses and midfielders, if $x_{1m} \succ x_{2d}$ there will be a gap area between defenders and midfielders with size $|x_{1m} - x_{2d}|$ and if $x_{1m} \prec x_{2d}$ section a and b will have an overlap and an area with a size of $|x_{1m} - x_{2d}|$ will be covered by both defenders and midfielders.

Let $P = (p_1, p_2, \ldots, p_n)$ be the location of n players in each section. Because of noise and loss of resolution, the covering performance at point q taken from i-th player at the position p_i degrades with the distance $\|q - p_i\|$.

We consider the task of minimizing the locational optimization function

$$H(P,V) = \sum_{i=1}^{n} \int_{V_i} f\left(\|q - p_i\|\right) dq$$

$$f\left(\|q - p_i\|\right) = \|q - p_i\|^2$$

The i-th player is responsible for covering over its dominance region V_i. The function H is minimized with respect to the player's location and the partition of Q. The optimal partition of Q is the Voronoi partition: $V(p) = V_1, V_2, \ldots, V_n$ generated by the points $(p_1, p_2, \ldots, p_n)$,

$$V_i = \{q \in Q \; \|q - p_i\| \le \|q - p_j\|, \forall j \ne i\}$$

for each Voronoi region Vi.

6 Conclusion and Future Activity

In future, the team's program is to add and put in effect many new features and research results and simultaneously import MATLAB codes of implemented algorithms to C++ in order to use in the team.

Another issue currently in debate to be considered in near future, is the implementation of dynamic positioning according to the positions of opponents.

References

1. S. P. Singh, T. Jaakkola, and M. I. Jordan., Model-free reinforcement learning for non-Markovian decision process problems. In *Proceeding of the Eleventh International Conference on Machine Learning,* pages 284-292. San Francisco, California, 1994.
2. L. P. Kaelbling, M.L. Littman, A. R. Cassandra. Planning and acting in partially observable stochastic domains. Jaunuary, 13, 1997.
3. A. Stroupe, M. C. Martin, and T. Balch., Distributed sensor fusion for object position estimation by multi-robot systems. In *IEEE International Conference on Robotics and Automation*., IEEE, May 2001.
4. D. H. Wolpert and K. Tumer. Optimal payoff functions for members of collectives. *Advances in Complex Systems*, 4(2/3):265–279, 2001.
5. A. K. Agogino, K. Turner.Team formation in partially observable multi-agent systems.
6. C.H.Lee, C.C Teng, "Identification and Control of Dynamic Systems Using Recurrent Fuzzy Neural Networks," IEEE Transactions on fuzzy systems, vol. 8, No. 4, 2000.
7. S. M. A. Salehizadeh, P. Yadmellat, A. R. Pourshoghi, M. J. Aein, "A Hybrid Algorithm for Training Recurrent Fuzzy Neural Network". In Proc: Artificial Intelligence and Pattern Recognition, Florida, USA, pp. 181-186, 2008.
8. Stirling. W. C, A Primer for Selectability and Rejectability. Brigham Young University, Provo 2000.
9. Stirling, W. C, Satisficing Games and Decisions. Brigham Young University, Provo 2000.
10. S.M.A. Salehizadeh, P. Yadmellat, M.B. Menhaj, " Transition from Particle Swarm Optimization to Individual Particle Optimization.", Submitted in: 24th IEEE International Symposium on Intelligent Control (ISIC), Part of 3rd IEEE Multi-conference on System and Control, Saint Petersburg, Russia, July 2009.
11. Mohammad Mehdi Korjani, Ahmad Afshar, Mehrab Norouzitallab, Mohammad Bagher Menhaj, "Dynamic Autonomous Agent Positioning Based on Computational Intelligence", Accepted in IAENG Conferences - IMECS 2009.
12. B.S. Stensrud and A.J. Gonzalez, "Context-Based Reasoning: A Revised Specification", Florida Artificial Intelligence Research Society (FLAIRS) Conference2004.
13. Mehrab Norouzitallab, S.M.A Salehizadeh, Ahmad Pourshoghi: "Nemesis Team Description 2008" Proceedings of Robocup 2008, Suzhou, China, 2008