United-2002 Team Description

Peter Stone

pstone@cs.cmu.edu
http://www.cs.cmu.edu/ pstone

Note: New permanent affiliation will be known prior to the tournament.
Please allow me to revise before publication.

1 Introduction

United-2002 is a successor of the CMUnited and ATTUnited teams: CMUnited-97, CMUnited-98, CMUnited-99, ATT-CMUnited-2000, and ATTUnited-2001. It is built mainly upon CMUnited-99 [8]. It also incorporates the team action architecture from ATT-CMUnited-2000 [7] (but not the dynamic planning of set-plays capability [3]) and the heterogeneous player management from ATTUnited-2001 [6]. The main research foci of United-2002 are:

- Scaling up temporal difference reinforcement learning;
- On-line learning of opponent player types by the on-line coach;
- Task decomposition with heterogeneous agents;
- Single-channel, limited-bandwidth communication.

2 Foundations

The United family of teams is the result of a long-standing and on-going research program on multiagent learning in dynamic, real-time domains. The three main research contributions are described in detail in a published monograph [4].

First, the book presents a flexible team structure which allows agents to decompose the task space into roles and formations. Team organization is achieved by the introduction of a "locker-room agreement" as a collection of shared conventions, including team plans and pre-defined conditions for agents to switch roles in coordination [10].

Second, it introduces layered learning, a general-purpose machine learning paradigm for complex domains in which learning a mapping directly from agents' sensors to their actuators is intractable [12]. Given a hierarchical task decomposition, layered learning allows for learning at every level of the hierarchy, with learning at each level directly affecting learning at the next higher level.

Third, it introduces a new multiagent reinforcement learning algorithm, namely team-partitioned, opaque-transition reinforcement learning (TPOT-RL) [11]. TPOT-RL is designed for domains in which agents cannot necessarily observe the state changes when other team members act.

Empirical results validate all three contributions in controlled experiments. The generality of the contributions is verified by applying them to the real robotic soccer [15] and network routing [5] domains.

3 Temporal Difference Learning

Our research in recent years has been focused mainly on temporal difference learning [13] in a subproblem of the full RoboCup task.

RoboCup simulated soccer presents many challenges to reinforcement learning methods, including a large state space, hidden and uncertain state, multiple agents, and long and variable delays in the effects of actions. In our on-going research [9], we have described our application of episodic Sarsa(λ) with linear tile-coding function approximation and variable λ to learning higher-level decisions in a keepaway subtask of RoboCup soccer. In keepaway, one team, "the keepers," tries to keep control of the ball for as long as possible despite the efforts of "the takers." The keepers learn individually when to hold the ball and when to pass to a teammate, while the takers learn when to charge the ball-holder and when to cover possible passing lanes. Our agents learned policies that significantly out-performed a range of benchmark policies. We demonstrated the generality of our approach by applying it to a number of task variations including different field sizes and different numbers of players on each team.

In United-2002, we aim to scale up the approach further towards the entire robotic soccer task. To date, our methods have worked surprisingly well given the large state space of involved in RoboCup domains. However, they have only been applied in situations involving very limited numbers of action choices (e.g. choosing to which player to pass). Scaling reinforcement learning techniques to large action spaces remains an important open challenge. Policy gradient methods [14] are promising candidates for meeting this challenge. As such, we plan to incorporate such techniques in United-2002.

4 Coach: On-line Learning of Player Types

Our main focus for the coach competition will be learning to identify opponent player types from observation. Due to the short time-scale of simulation league games, past opportunities for on-line learning have been extremely limited. With the introduction of heterogeneous players, the opportunity arises for the on-line coach to attempt to deduce the opponent player types from observing their movements on the field. Since the players are drawn from a pool of just 7 player types, the problem reduces to a 7-way classification problem with constraints (e.g. no more than 3 players of any given type). The coach will either transmit player types to its team members and expect them to self-organize, or actively assign team members to roles so as to best counteract/exploit the opponents' capabilities.

5 Task Decomposition with Heterogeneous Agents

Until the last RoboCup competition, the players have all been homogeneous in terms of low-level capabilities. The introduction of heterogeneous players introduces several challenges:

– The need to select a team from a pool of players;
– The need to organize the players based on their differing capabilities;
– The opportunity to exploit/counteract the opponents' heterogeneous capabilities (see above)

In our research, we aim to address these challenges and to investigate and quantify the potential benefits of heterogeneous teams over homogeneous teams. We will incorporate this line of research both into our team and into our coach entry.

6 Single-Channel, Limited-Bandwidth Communication

In most multiagent systems with communicating agents, the agents have the luxury of using reliable, multi-step negotiation protocols (see [1] for instance). They can do so primarily when communication is reliable and the cost of communication relative to other actions is small. For example, in Cohen's convoy example [2], the communication time required to form and maintain a convoy of vehicles is insignificant compared to the time it takes the convoy to drive to its destination. Similarly, message passing among distributed information agents is typically very quick compared to the searches and services that they are performing. Thus, it makes sense for agents to initiate and confirm their coalition while guaranteeing that they will inform each other if they have trouble fulfilling their part of the joint action.

In previous research, we have considered multiagent environments with unreliable, high-cost communication. For example, if there is only a single, low-bandwidth, unreliable communication channel for all the agents, and if the agents must sacrifice valuable resources in order to communicate, then although inter-agent communication may be beneficial, the agents' behaviors must not *depend* upon it.

The RoboCup SoccerServer is one clear example of an environment. In previous research, we have presented techniques for dealing with the obstacles to inter-agent communication in such environments, particularly those with several *teams* of agents [10].

This year, the SoccerServer has drastically reduced the limits on message length. As such, the United-2002 will investigate the applicability of previous methods to the even more constrained environment while simultaneously searching for new methods of coordinating and communicating in such challenging environments.

Acknowledgments

The author thanks collaborators on previous United teams and research in the RoboCup SoccerServer including Patrick Riley, Manuela Veloso, David McAllester, and Richard Sutton.

References

1. Mihai Barbuceanu and Mark S. Fox. COOL: A language for describing coordination in multi agent systems. In *Proceedings of the First International Conference on Multi-Agent Systems*, pages 17–24, Menlo Park, California, June 1995. AAAI Press.
2. Philip R. Cohen, Hector J. Levesque, and Ira Smith. On team formation. In J. Hintikka and R. Tuomela, editors, *Contemporary Action Theory*. Synthese, 1999. to appear.
3. Patrick Riley and Manuela Veloso. Adaptive team coaching using opponent model selection. In *Fifth International Conference on Autonomous Agents*, 2001.
4. Peter Stone. *Layered Learning in Multiagent Systems: A Winning Approach to Robotic Soccer*. MIT Press, 2000.
5. Peter Stone. TPOT-RL applied to network routing. In *Proceedings of the Seventeenth International Conference on Machine Learning*, 2000.
6. Peter Stone. ATTUnited-2001: Using heterogeneous players. In Andreas Birk, Silvia Coradeschi, and Satoshi Tadokoro, editors, *RoboCup-2001: Robot Soccer World Cup V*. Springer Verlag, Berlin, 2002. To appear.
7. Peter Stone and David McAllester. An architecture for action selection in robotic soccer. In *Proceedings of the Fifth International Conference on Autonomous Agents*, 2001.
8. Peter Stone, Patrick Riley, and Manuela Veloso. The CMUnited-99 champion simulator team. In M. Veloso, E. Pagello, and H. Kitano, editors, *RoboCup-99: Robot Soccer World Cup III*, pages 35–48. Springer Verlag, Berlin, 2000.
9. Peter Stone and Richard S. Sutton. Scaling reinforcement learning toward RoboCup soccer. In *Proceedings of the Eighteenth International Conference on Machine Learning*, 2001.
10. Peter Stone and Manuela Veloso. Task decomposition, dynamic role assignment, and low-bandwidth communication for real-time strategic teamwork. *Artificial Intelligence*, 110(2):241–273, June 1999.
11. Peter Stone and Manuela Veloso. Team-partitioned, opaque-transition reinforcement learning. In *Proceedings of the Third Annual Conference on Autonomous Agents*, pages 206–212. ACM Press, May 1999.
12. Peter Stone and Manuela Veloso. Layered learning. In Ramon López de Mántaras and Enric Plaza, editors, *Machine Learning: ECML 2000*, pages 369–381. Springer Verlag, Barcelona,Catalonia,Spain, May/June 2000. Proceedings of the Eleventh European Conference on Machine Learning (ECML-2000).
13. R. S. Sutton. Learning to predict by the methods of temporal differences. *Machine Learning*, 3:9–44, 1988.
14. R.S. Sutton, D. McAllester, S. Singh, and Y. Mansour. Policy gradient methods for reinforcement learning with function approximation. In *Advances in Neural Information Processing Systems 12*, pages 1057–1063. MIT Press, 2000.
15. Manuela Veloso, Peter Stone, and Kwun Han. The CMUnited-97 robotic soccer team: Perception and multiagent control. In *Proceedings of the Second International Conference on Autonomous Agents*, Minneapolis, MN, May 1998.