Team Description for Baltic Lübeck 2002

Martin Haker[1], André Meyer[1], Behboud Kalantary[1],
Daniel Polani[1], and Thomas Martinetz[1]

Institute for Neuro- and Bioinformatics, University of Lübeck, Germany
`polani@inb.mu-luebeck.de`
`http://www.inb.mu-luebeck.de/robocup/robocup-e.html`

Abstract. A concept is presented for the development of mid-term game strategies that go beyond reactive play and attempt to model more far-sighted game playing. The relations of the concept to TD-learning models and to strategy identification from optimized behavior patterns are emphasized. Furthermore we model opponents to enable prediction of their behavior outside the view cone and to adapt positional and behavior patterns to the specific behavioral properties of the opponents.

1 Introduction

Game-playing is a classical scenario to evaluate AI programs. The field of automated strategy game playing has seen significant successes in recent time, most notably the defeat of a Chess world champion by a computer which has long been seen as the ultimate pinnacle of game-playing AI. In some respect this success, however, is not completely satisfying. It is due — to a not insignificant part — to the massive number-crunching capabilities of modern hardware which allows the explicit evaluation of a very large number of possible constellations. It is believed that believed expert human players do not do that, at least not explicitly.

The computerized evaluation of a large number of positions can be seen as an inherently static view of situations. The dynamics of the strategy is implicit to the evaluation function. It is well known that the typical human player, however, studies a much smaller number of positions but includes goals and plans in his consideration; in other words, the view of a human player is an inherently dynamic one. There exist AI models that strive to capture this dynamic view [1–4]. At the same time, expert human players use a long-term "condensed" view of situations [5, 6].

This differs considerably from the way computers play strategy games. From a philosophical point of view, perhaps the the closest approximation to that type of mixture of situational and operational view is a type of work which is best exemplified by TD-gammon [7]. Here a backgammon strategy is automatically learned by temporal-difference (TD) learning. The TD element corresponds to a short-term dynamic view, the value function of the different board states corresponds to the condensed state information. The success of TD-gammon can be attributed to a significant degree to the probabilistic structure of the game.

It causes the strategic correlation length to remain relatively short and thus to be compatible to the learning time horizon in which the TD approach operates best. One of the goals underlying our work is trying to extend this short-term positional and dynamical view of strategy development to more complex and more long-term scenarios. RoboCup soccer provides a scenario with intermediate strategic correlation length.

2 The Connection to RoboCup

One of the phenomena that can be observed in the behavior of many RoboCup teams is a clear separation between reactive and positional play, apart from relatively long-term formational changes. This holds even for strong teams. The effect is that, even when very good teams are playing, it is not always clear whether their players follow a middle-term strategy or just switch between reactive play and an positioning strategy according to prior criteria. Pre-planned middle-term strategies (locker-room agreements) like in [8] or situation-based positioning as in [9] are used by strong teams.

Our goal is a concept that takes a middle ground between a position-oriented reactive mechanism with externally defined formation and positioning and long-term strategic plans. Purely reactive mechanisms which have been used before in our teams [10] lacked the capability of foresight. We find it necessary to construct a mechanism that explicitly introduces such a concept. A TD approach learning couples static and dynamical sight [11], but for large state spaces, it takes a long time to search the strategy space. In addition, the result of a TD approach does not provide an explicit concept of middle-term strategy.

Instead, we argue that the high-level concept of strategy can be seen as an representation emerging from the condensation of characteristic behavior classes. This view of things is related to [12], where behavior classes are extracted from existing player behaviors. We intend to develop this view further. Instead of just derive them from pre-programmed player behaviors, it is one of our mid-term goals to identify strategy structures which crystallize from TD-like optimizations of player behavior. On the other hand, we believe that this view also offers a suitable basis for the introduction of external knowledge, which motivates our present work.

3 Strategy Horizon and Opponent Model

To extend the strategy horizon of our players, we provide the players with explicit micro-strategies. The most important application is in situations where the players do not need to react immediately, but have a certain degree of freedom to prepare longer-termed moves. These micro-strategies can be seen as puzzle pieces that describe possible movements of players. These puzzle pieces are adapted into the players' current situation context if they fulfil certain properties. Such a puzzle piece can be the preparation of a flank, or certain move patterns to escape marking. However, other applications also include the exclusion or preferred execution of certain pass patterns. In addition, we intend to combine these puzzle pieces with evaluations that allow to adapt the preferences of making use of these pieces using TD-learning.

This centerpiece of our concept is supported by two further mechanisms. First, a static overall model that governs formations and positioning of the players similar to [9] is used. This serves to ensure that the puzzle piece mechanism does not break up the whole formation after some playing.

The second mechanism is a fully dynamical one. It is related to the mechanism presented in last year's qualification material and uses a specialized function to evaluate game situations and to derive the tactical maneuvres from that function. We couple the minimax approach from [13] with a distributed Dynamic Programming evaluation scheme. A given game situation is evaluated according to this scheme. In this scheme, the probability of a pass success for the player with the ball and a pass failure is estimated. In the case of success, the ball can be passed or dribbled on. In the case of failure, it is assumed that an opponent catches the ball and then will attempt to perform that move which is most advantageous to it. The resulting value is then estimated after an iteration of 2-3 steps (longer iterations lose accuracy because of the decrease of the strategy correlation after these passes).

Based on these values, a strategic positioning for agents can be calculated. This can be done in addition to the static or the puzzle piece positioning. The disadvantage that makes the fully dynamical positioning insufficient to be used alone is that its time-range is relatively limited because small errors in the pass probabilities can add up to significant errors after few iterations. However, we consider it as useful to enhance and fine-tune the action decisions suggested by the static and the puzzle piece model.

A second aspect of our code is an enhancement of prior world model work [14, 15]. There, the questions of adequate representation of the world at the right time and a improved prediction quality of the movement of predictable objects (i.e. the ball) have been addressed. Here, we include into the model the opponents. The field of influence which is detected during the game and their assumed intentions is modeled to allow exploitations of tactical weaknesses of their behavior. Since tournament games are too short for a systematic learning, predefined models are provided with which the opponent behavior is matched. This opponent model serves to better predict pass probabilities and the usefulness of strategic positioning. In addition, due to the recently introduced limitation of bandwidth it will no more be possible for an agent to obtain a halfway global world view by world model broadcasts of its team members. Thus, the opponent model will be useful to improve the localization of opponents outside a player's view cone.

Acknowledgements

We base our work upon a port of the public available part of the Tsinguaeolus 2001 code to Linux.

References

1. Rao, A.S., Georgeff, M.P.: BDI Agents: from theory to practice. In Lesser, V., ed.: Proceedings of the First International Conference on Multi–Agent Systems, San Francisco, CA, MIT Press (1995) 312–319
2. Richards, N., Moriarty, D., McQuesten, P., Miikkulainen, R.: Evolving Neural Networks to Play Go. In: Proceedings of the 7th International Conference on Genetic Algorithms, East Lansing, MI (1997)
3. Georgeff, M., Pell, B., Pollack, M., Tambe, M., Wooldridge, M.: The Belief-Desire-Intention Model of Agency. In Müller, J., Singh, M.P., Rao, A.S., eds.: Proceedings of the 5th International Workshop on Intelligent Agents V : Agent Theories, Architectures, and Languages (ATAL-98). Volume 1555 of LNAI., Berlin, Springer (1999) 1–10
4. Burkhard, H.D., Hannebauer, M., Wendler, J.: Belief-Desire-Intention Deliberation in Artificial Soccer. The AI Magazine **1998** (1998) 87–93
5. Nichelli, P., Grafman, J., Pietrini, P., Alway, D., Carton, J.C., Miletich, R.: Brain activity during chess playing. Nature **369** (1994) 191
6. Amidzic, O., Riehle, H.J., Fehr, T., Wienbruch, C., Elbert, T.: Pattern of focal gamma bursts in chess players. Nature **412** (2001) 603
7. Tesauro, G.: Temporal Difference Learning and TD-Gammon. Communications of the ACM **38** (1995)
8. Stone, P., Veloso, M., Riley, P.: The CMUnited-98 Champion Simulator Team. In Asada, M., Kitano, H., eds.: RoboCup-98: Robot Soccer World Cup II. Springer Verlag, Berlin (1999) To Appear.
9. Reis, L.P., Lau, N.: FC Portugal Team Description: RoboCup 2000 Simulation League Champion. In Peter Stone, T.B., Kraetzschmar, G., eds.: RoboCup-2000: Robot Soccer World Cup IV. Springer, Berlin (2001)
10. Polani, D., Uthmann, T.: Between Teaching and Learning: Development of the Team Mainz Rolling Brains for the Simulation League of RoboCup '99. In: Team Descriptions Simulation League, RoboCup '99, Stockholm. (1999)
11. Buck, S., Riedmiller, M.: Learning situation dependent sucess rates of actions in a robocup scenario. In: Proceedings of PRICAI '00, Melbourne, Australia, 28.8.-3.9.2000. (2000) 809
12. Wünstel, M., Polani, D., Uthmann, T., Perl, J.: Behavior Classification with Self-Organizing Maps. In Stone, P., Balch, T., Kraetzschmar, G., eds.: RoboCup-2000: Robot Soccer World Cup IV. Springer Verlag, Berlin (2001) Winner of the RoboCup 2000 Scientific Challenge Award.
13. Littman, M.L., Szepesvári, G.: A Generalized Reinforcement-Learning Model: Convergence and Applications. In Saitta, L., ed.: Proceedings of the 13th International Conference on Machine Learning (ICML-96), Bari, Italy, Morgan Kaufmann (1996) 310–318
14. Polani, D., Weber, S., Uthmann, T.: A Direct Approach to Robot Soccer Agents: Description for the Team Mainz Rolling Brains, Simulation League of RoboCup '98. In Asada, M., Kitano, H., eds.: RoboCup-98: Robot Soccer World Cup II. Springer Verlag, Berlin (1999)
15. Haker, M., Meyer, A., Polani, D., Martinetz, T.: A Method for Incorporation of New Evidence to Improve World State Estimation. In: Proc. RoboCup 2001 Symposium, Seattle (2001)