Reading the Play - Adaptation by Prediction of Agent Motion

David Ball[1,2], Gordon Wyeth[2]
[1]Queensland Brain Institute
[2]School of Information Technology and Electrical Engineering
The University of Queensland
St Lucia, Queensland, 4072 Australia
{d.ball, g.wyeth}@uq.edu.au

Abstract

An adaptive agent improves its performance by learning from experience. This paper describes an approach to adaptation based on modelling dynamic elements of the environment in order to make predictions of likely future state. This approach is akin to an elite sports player being able to "read the play", allowing for decisions to be made based on predictions of likely future outcomes. Modelling of the agent's likely future state is performed using Markov Chains and a technique called "Motion and Occupancy Grids". The experiments in this paper compare the performance of the planning system with and without the use of this predictive model. The results of the study demonstrate a surprising decrease in performance when using the predictions of agent occupancy. The results are derived from statistical analysis of the agent's performance in a high fidelity simulation of a world leading real robot soccer team.

1 Introduction

The most successful sports players are often described as having the ability to "read the play" – to instantly evaluate the likely future state of the game and choose an action that will benefit themselves or their team. Can an agent's planner "read the play" by predicting the dynamic agents and adapt? While an agent may achieve acceptable performance by representing the dynamic agents as static elements, the agent may improve its performance by inferring the dynamic agent's state from multiple observations.

This research is relevant to applications ranging from the autonomous operation of robots to the design of non-player characters in computer games. For these applications, it is typical to assess the state of the environment to select appropriate actions for the agent(s).

1.1 Adaptive Autonomous Agents

An autonomous agent is adaptive if it is capable of improvement in achieving its goals with experience [Maes, 1994] either by increasing its performance within the current situation or learning to handle new tasks or changes in the environment. Adaptation is also a method for an agent to increase its robustness and effectiveness. A review of adaptation methods for autonomous agents can be divided into the following three categories.

Techniques in the adaptation by reinforcement category directly connect performance evaluations of the agent's actions with a mechanism that updates the way the agent thinks. For approaches in this category, the agent does not try to understand the cause of adaptation; however, it is powerful because the agent is not limited by preconceived notions of the environment or task.

Techniques in the adaptation by classification category attempt to have a predefined response or strategy for all possible situations that might occur. This approach can work well when the responses cover and match the possible range of situations. Approaches in this category contrast with the first approach as the learning or designing phase occurs before deployment.

The third category consists of approaches that model or predict the environment for use as a resource for future planning. These approaches combine sensing of the current state with online learning to make predictions about the future state. It is this third category that is the focus of this paper.

Adaptation by modelling and prediction in spatial and dynamic environments is a difficult and relatively unexplored research area. There has been significant successful work in modelling the static features of the environment, for example in SLAM navigation systems [Thrun, 2002]. However, modelling and predicting the dynamic features of an environment, such as the other agents, is a difficult and relatively unexplored research area. The key is to identify patterns and tendencies in the behaviour of the dynamic features, and then to make plans that exploit the understanding of the patterns. Neither pattern identification nor integration of this new resource into an agent's planning system is typically easy. However, an agent's planner that can integrate learnt patterns of dynamic elements with its other resources will potentially be able to select more effective actions. Integrating learnt patterns of the other agents has been used in 'simple' games and environments to improve the performance of algorithms such as Minimax [Neumann and Morgenstern, 1944] and research has shown this gives better results [Carmel and Markovitch, 1996]. This paper investigates whether modelling and prediction in spatial and dynamic environments can similarly improve adaptive behaviour.

Multi-agent environments increase the difficulty in identifying patterns in the dynamic elements due to the possibility of interaction and interference between agents. Many environments require coordination of multiple agents using either centralised or distributed planning systems. There are many different techniques for designing a planning system; the most flexible ones allow integration of new information, such as learnt exploitable patterns of the dynamic elements, without modifying their design. Some agents may not be under the control of the planner, and these agents may not have the same goals or may have conflicting or opposing goals. In a competitive domain where agents have opposing goals the opponent agents are likely to hide their strategy. An example of such an environment is where teams of agents play a modified version of human soccer, such as in the RoboCup soccer leagues.

1.2 RoboCup

This paper presents its results within the Small Size League (SSL) of the international organisation RoboCup [Kitano, *et al.*, 1995]. RoboCup fosters intelligent agent and robotic research using the well known domain, soccer. In the SSL both teams have five robots that each must physically fit inside a cylinder with a diameter of 180mm and a height of 150mm. The rules are similar to the FIFA rules for the human version of the game, but without the offside rule and with other changes required to make sense for wheeled robots. The robots are fully autonomous in the sense that the human operators do not input any strategy or control during play although during half time and timeouts they may make changes. Humans referee the matches.

The introduction of the coach competition in the RoboCup Simulation League and the development of Coach LANGuage (CLANG) has had a significant effect on promoting the development of systems that adapt to the behaviour of other agents [Chen, *et al.*, 2001].

Bowling *et al.* [2003; 2004] applied adaptation by reinforcement to the online selection of team strategy within the Small Size League environment. The basis of the strategy system is the selection of a play from a playbook.

There are several successful adaptation by classification approaches; notably Takahashi's [1999] and Visser's *et al.* [2001] adaptation of formation and Riley and Veloso's [2001] and Steffens [2004] predefined responses to recognised situations.

There are several successful adaptation by modelling approaches, from Habibi's *et al.* [2002] adaptation of formation, Riley and Veloso's [2004] and Kuhlmann's *et al.* [2005] integration with CLANG.

1.3 Paper Outline

The remainder of the paper is organised as follows:

Section 2 describes the research platform, in particular, the RoboRoos' Multi-Agent Planning System. The experiments will compare its performance with and without the predictions.

Section 3 provides an overview of previous research that models and predicts an agent's likely future state as a probability distribution [Ball and Wyeth, 2005].

Section 4 describes the methodology behind the approach in this paper, before listing the details of the experiments including the different predictive approaches tested. A case study demonstrates why the predictions result in poor performance.

Section 5 discusses the results before Section 6 concludes the paper.

2 Research Platform

The research platform, the RoboRoos, was The University of Queensland's entry into the annual RoboCup Small Size League competitions. The experiments presented in this paper use the RoboRoos' system as it was at RoboCup 2003, operating under the 2003 rule set.

The RoboRoos are more than just a research platform; they are a successful and respected robot soccer team [Ball and Wyeth, 2004; Ball, *et al.*, 2003; Wyeth, *et al.*, 2002]. The RoboRoos team retired in 2004 after seven years of international competition in which they came 2nd three times. These results highlight the high performance of the team.

2.1 Multi-Agent Planning System

The most relevant module of the RoboRoos' system to this paper is the Multi-Agent Planning System (MAPS) which is the highest level planner in the system, responsible for distributing the overall goal of the team amongst the individual robots [Tews, 2002]. MAPS is responsible for multi-robot coordination by selecting a role and role parameters for each agent. MAPS uses potential fields as the primary mechanism for determining roles. The potential fields can model the suitability of a role for the different agents, or find suitable role parameters. MAPS can represent features of an agent's behaviour space by overlaying different potential field components.

Figure 1 shows the potential field generated to determine the location to dribble a ball to during a robot soccer match. The opponent's goal is on the right side and the white circles represent the opponents. There are several overlapping fields at work here.

- Basefield: This field is ramped towards the opponent's end of the field and off the walls. This encourages dribbling towards the opponent's goal.
- Clear Path: This field represents clear paths from the opponent's goal and this encourages movement towards a clear shot on goal.
- Distance From: Locations further from the ball have higher values. This encourages shorter dribble distances.

- Object Regions: The opponent's positions have regions with low values. This encourages dribbling to a location away from the opposition.
- Object Region: The dribble to grid cell from the previous frame is biased. This is to help prevent oscillations between grid cells with similar values.

The white regions in the figure represent the most desirable areas to dribble the ball. The most desirable area in this case is to the right of the goal where there is an opportunity for a clear shot on goal that is well away from other players ('in space') but is not too far from the goal mouth.

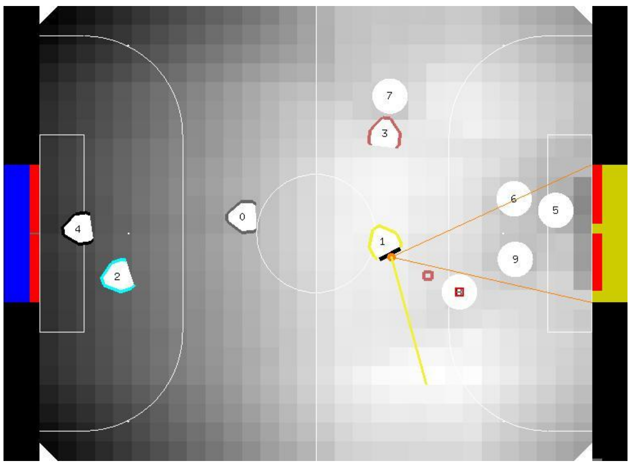


Figure 1 - This figure shows a potential field that determines where to dribble the ball to. The opponent's goal is on the right side of the field and the opponent agents are the circles. Agent 1, who is located in the centre of the field, has the ball and is assigned a dribble to command. The lightest point in the potential field represents the desired location to dribble to; the thin line from agent 1 indicates this point. Note the strong effect of the Basefield and the clear path to the opponent's goal.

3 Prediction of Agent Motion

Motion and Occupancy Grids are an extension to the Occupancy Grid method required to deal with dynamic agents in the environment. The Occupancy Grid method [Elfes, 1990; Moravec and Elfes, 1985] provides a probabilistic approach for representing how a region of space is occupied. Representing obstacles for robot navigation is one typical use. The occupancy grid is also suitable for representing the probability distribution of the space that an agent will potentially occupy in the future. Each grid cell will have a probability of the agent occupying that particular location. To predict the future state of the agent the module will model the probability of transitioning between grid cell states. However, this ignores the agent's motion within the grid cell and cannot account for different motion paths through the same grid cell. The Motion and Occupancy Grid approach extends this method so that it can capture and represent the motion of occupied space or in this case, dynamic agents.

The Motion Grid captures the agent's current direction of motion so that it can distinguish between different motion types through a grid cell. For example, an agent may have two motion paths through a cell, one from bottom to top and the other from left to right. If the motion of an agent is not captured the probability distribution will show the predicted motion to the right and the top. The Motion Grid allows for separation of the two motion types through the grid cell. Figure 2 shows a graphical representation of the Motion and Occupancy Grid cells. The centre circle represents a stationary agent which will maintain the same occupancy grid state. The other cells represent directions of motion, with the domain of possible directions divided into equal amounts.

This approach uses the Markov Chain methodology to connect states in time. A Markov Chain consists of a finite number of states and a State Transition Matrix (STM) which contains probabilities of moving between states. Typically an agent's behaviour is dependent on other agents or other dynamic elements of the environment. For the Markov assumption to hold true, the state vector must represent these elements.

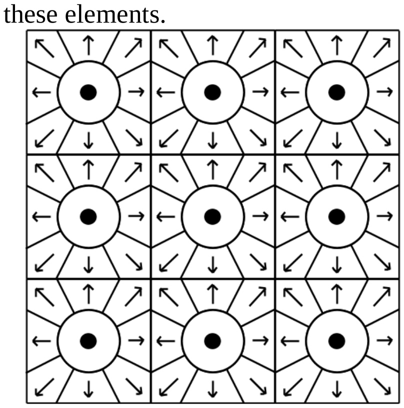


Figure 2 - This figure shows how the state of an agent is represented using the Motion and Occupancy Grid. The centre circle represents stationary and the cells surrounding the circle represent the possible directions of motion.

3.1 Experiment

To test the prediction of agent motion approach the agents play a standard game lasting twenty minutes using most of the rules of the Small Size League. The opponent agent's motion is modelled and predicted relative to the motion of the ball.

The time between the current and next state, one time step, is 250 milliseconds (fifteen frames). The model is updated with the pose information from every global external state update (60Hz). The agents are predicted forward one second into the future (a significant time in robot soccer) representing four steps in the Markov chain. At the agent's top speed of 1.5 m/s they can move up to approximately eight grid squares in the prediction time.

Each Occupancy Grid cell is approximately the same size as an agent. This gives 15 x 12 grid squares. The approach represents the agent's motion by a 3 x 3 matrix where the centre cell indicates no motion and the other cells represent motion in eight directions. The agent's motion will be set to stationary if its velocity is below a threshold. This threshold is the velocity where an agent is just as likely to remain in its current Occupancy Grid cell as it is to change to an adjacent cell, where the uncertainty comes from quantisation of position. This is half of the cell width divided by the time step time, giving a threshold velocity of 0.36 m/s.

The results include a First Order prediction of the agent's motion. The First Order prediction is the grid cell that the agent will occupy if it continues with its current motion (based on the non-quantised position and velocity). If the agent is stationary then the prediction is that it will maintain its current Occupancy Grid cell. When the agent is in a state that it has not previously experienced, the model will be unable to give a prediction. The results also include a Stationary prediction, which is that the agent will stay in its current grid cell. The results include another set of predictions, termed Always Predicting, which uses the First Order prediction or Stationary prediction when the model is unable to give a prediction.

The results use the Euclidean distance between the actual and the predicted position (as given by a weighted average) to determine performance. Figure 3 shows the error for the full game.

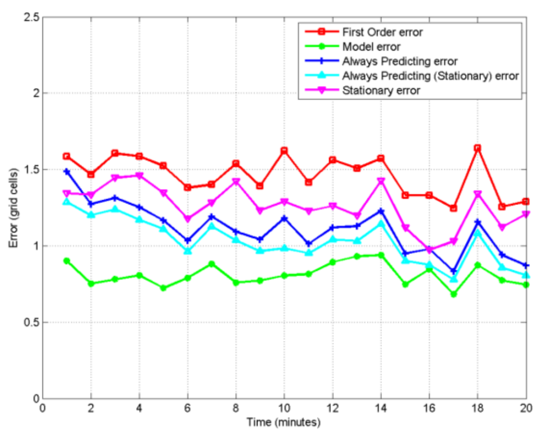


Figure 3 – This figure shows the average errors for the prediction methods for a full twenty minute game.

These results do not reflect the usefulness of knowing the likely paths that the agent will take. Figure 4 shows the predicted probability distributions for two instances. In the first instance, the distribution shows two possible paths that the agent may take to acquire the ball. The second instance shows that the agent has moved to the cell with the highest probability from the first instance. Note that this would have given approximately a 1.5 cell error in Figure 3 because of the difference between the agent's location and the average model error (as given by the concentric circles).

Exploiting the predictions will assess the usefulness of knowing the shape of the likely paths. The next section uses the predicted probability distributions to investigate adaptation by prediction of agent motion.

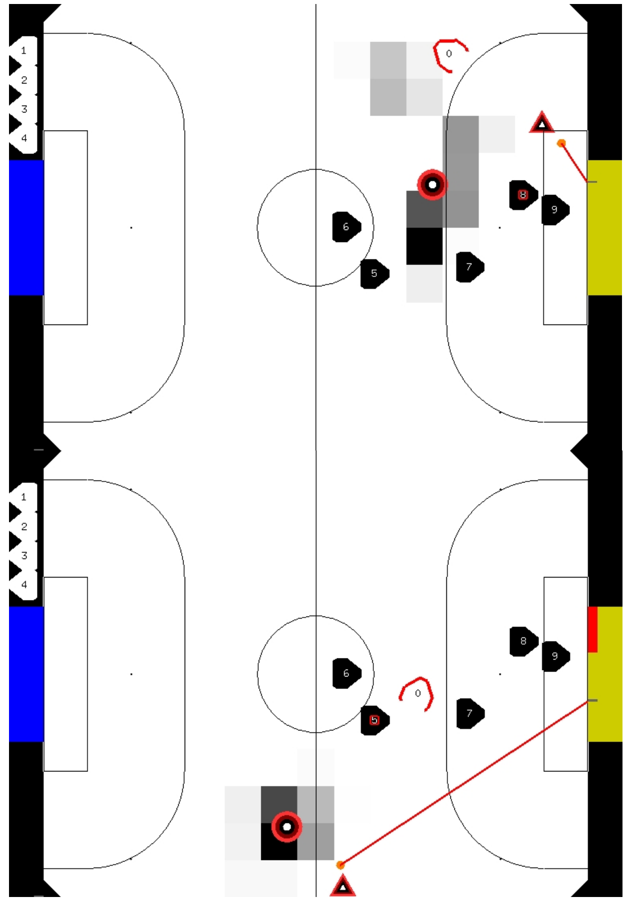


Figure 4 – This figure shows the agent's predicted occupancy for two frames separated by the prediction time. Darker grid cells represent a stronger prediction. The triangle represents the first order prediction; the circle represents the average model prediction. The second frame shows that the agent moved in one of the predicted paths.

4 Adaptation by Prediction of Agent Motion

The study involves integrating the previous sections distributions of future occupancy into the RoboRoos' Multi-Agent Planning System (MAPS). MAPS uses potential fields as a mechanism for determining the role and role location based on a library of potential field functions and abstractions to model the suitability of a role for an agent or to find a suitable role location. Up until this point, MAPS treated the effect of the opposition as small abstract fields centred at their current location. As MAPS uses the potential fields to determine long term plans (relative to path planning and control), the predicted future occupancy of the opposition seems a more appropriate representation of the effect of the opposition. This section investigates how worthwhile the predictions of occupancy are with regard to increasing the performance of MAPS.

4.1 Prediction Study Details

MAPS selects the role and role parameter for all the agents. For the offensive strategies used by the 2003 RoboRoos team, the most important parameter to determine is the desired location to dribble the ball, termed the Dribble field. Dribbling is the most obvious candidate for the use of prediction. Locations for shooting will not benefit from predictions as the ball reaches the goal in a matter of milliseconds. Defence roles are conservative, and tend to use the opponent's crisp location rather than a broader potential field to determine the target locations for defence. Dribbling is more speculative, and makes use of a broad potential field to model the opponent's influence. Changing this broad potential field to a more qualified prediction of opponent motion seems a strong candidate for improving the team's performance.

Section 2.1 describes an example Dribble field. Determining the clear paths to shoot the ball at the goal is a critical subfield to the Dribble field. There are two steps to build the clear path field. The first step is to add fields representing the opposition agents to a blank field. The second is to trace paths from regularly spaced intervals along the opponent's goal through the field built in the first step. MAPS also uses the oppositions' effect to determine clear paths to dribble the ball and object regions so that the ball is not dribbled too close to the opponents. In this study, instead of using the current location of the opposition, MAPS uses the discrete probability distributions that represent the opposing agents' likely paths.

First Order Prediction

The RoboRoos' system has a module to estimate the state of the robots from the noisy and incomplete information sent by the RoboRoos' vision system. The First Order prediction method uses the estimated velocities to linearly predict the pose of the agents. This is identical to the method employed in the previous section. The same size and shape object region is built at the predicted region as the non adapting team. The agents will sometimes have a predicted position outside the field using the First Order prediction, but the object region builder limits the effect to on-field locations. This method is included in the experiments as a baseline prediction method to compare to the other methods.

Model Learning

The Model Learning method integrates the discrete probability distributions from the Motion and Occupancy Grids module described in the previous section into the existing RoboRoos' planning system, MAPS. This method learns the model of the opponents' motion throughout the experiment at the same time as MAPS is using the predictions based on the model to adapt. This is the primary prediction method that would be implemented for a competition, where the model of the opponent's motion is unknown before the match. There are two implementations of this method for when the model is unable to provide a prediction. One uses a Stationary prediction, which assumes the agent will maintain its current grid location, the other uses a First Order prediction, which assumes the agent will continue with their current motion.

Model Learned *a priori*

Like the previous method, this method integrates the discrete probability distributions described in the previous section into the existing RoboRoos' planning system, MAPS. However, it differs in that the model of opponent motion is learned a priori by observing the two teams playing without either adapting and then the model is fixed for a game while one team adapts. Having the model learned a priori will provide a method that separates the issue of insufficient learning time with adaptation based on prediction performance. However, note that this method is only valid for obtaining results for this paper as there is typically not an opportunity to appropriately observe and model the opponents' motion before a game, especially against the RoboRoos' team.

4.2 Experimental Setup

The experiments use the RoboRoos' simulator operating under the 2003 Small Size League rules. Fifty games each lasting twenty minutes were used to test each prediction method. Both teams have an equal chance of winning the match; confirmed by comparing the mean goals scored for fifty games. The opponents' likely location is predicted 0.5 seconds into the future for all of the prediction methods. In the prediction approach where the model is learned a priori the learning time was one hour. The experiments also include a reference set of games where neither team adapted.

When integrating the predictions of the oppositions' effect into the MAPS system it must weight these fields relative to the other potential field components. (See Section 2.1 for more details.) For these experiments, the weights were only minimally tuned and were constant for an entire implementation testing. The minimal amount of tuning was to ensure that the field has some effect on the decision. Most notably the Clear Path and Opponent Region fields were increased in strength to account their more diffuse distribution in the adaptive case. Later investigations involved tuning the weights further as discussed later in the paper.

The three different prediction methods for adaptation were each tested using two different offences against two different defences (four of tests for each adaptation method). This is to test the generality of the adaptation approach. The following offensive strategies, that dictate the behaviour of an attacking agent that does not have the ball, were tested:

- Ball Player Screening (BPS): This offence has two modes dependent on which team has possession of the ball. When a team does not have possession of the ball, an attacking agent tries to prevent the closest opponent agent to the ball from reaching the ball. The attacking agent does this by moving between the ball and the opponent agent. When a team does have possession of the ball, an attacking agent tries to create more freedom for the agent that has possession of the ball. The attacking agent does this by 'screening' the closest opponent to the ball.
- Cover Screening (CS): In this offence one attacker

attempts to impede the motion of the opponent agents that are preventing direct shots on the goal. As the agent with the ball dribbles or passes the ball across the goal these defensive agents typically move laterally across the front of their Goal Keepers' area covering the goal from direct shots. The attacker tries to slow down how quickly they move into new covering positions by holding a position on their cover path.

The following defensive strategies, that dictate the behaviour of a defending agent that does not have the ball, were tested:

- Double Ball Player (DBP): MAPS assigns a tackling agent to each side of the goal. This prevents the screening type offences from holding the defence on one side of their goal which would expose the other side to a fast dribble or pass across the goal. This defence also spreads out the defence which helps avoid situations where the defensive agents interfere with each other.
- Goal Side (GS): MAPS assigns the floating defence agent to stay level with and goal side of the opponent to prevent them from moving across the goal face. This can be effective as it hinders the ball moving towards the centre of the field where there is more goal area to shoot at. This defence also 'encourages' the opponent agent who has the ball towards the edge and corner of the field where they can be contained and tackled.

A heteroscedastic (unequal variance) student's t-test is used to determine the whether the adaptation approach improved the chance of scoring goals for each of the offensive and defensive strategy combinations. The experiments use a confidence level of 5% to conclude a statistically significant change in the number of goals scored per game from the non-adaptive case. For this experiment the null and alternate statistical hypothesises are defined as:

H0 – The proposed integrated prediction approach has no effect on the number of goals scored per game.

H1 – The proposed integrated prediction approach increases the number of goals the adapting team scores per game.

4.3 Case Study

This section presents the results for the experiments where the Ball Player Screening (BPS) offence and Double Ball Player (DBP) defence are used. This particular combination of offence and defence was the most commonly used by the RoboRoos team in RoboCup 2003, including the grand final. It is useful to look at this set of results before the next section presents the results of all the experiments.

Table 1 shows the results for all the BPS and DBP experiments. The first column shows the particular combination of the prediction method, offence, and defence strategy. The second and third columns show the sample mean goals for both the team that was adapting and the team that was not. The next column shows the results of the Student's one tailed heteroscedastic t-test. Positive results indicate that the mean goals scored increased significantly by using the adaptation module. Negative t-test results indicate a significant decrease in the goals scored with the adaptation module. Zero indicates an inconclusive result. The last two columns show the number of games won by both teams. These may not necessarily add to the fifty games in the experiment due to tied games.

The results show that the team that is adapting performs significantly worse that the team that is not. This result can be verified in two ways: the t-test results significantly decrease the number of goals scored by the adaptive team and the win-loss ratio of the adapting team show a significant decline. Reading down the table, the mean goals scores drops when using; First Order prediction by approximately 24%, Model (Stationary) based prediction by approximately 28%, Model (First Order) based prediction by approximately 32%, and Model-based prediction learned a priori by approximately 43%. This trend tends to indicate that the more accurate the predictions, the worse the adaptation performance.

Table 1 - These results show the performance of adapting using the various prediction methods for the combination of the BPS and DBP strategy. The table shows the testing combination, the mean goals scored per game scored by both teams and the result of the 5% t-test. The table also shows the number of wins for both teams. Note that these may not necessarily add up to fifty games due to tied games.

Testing Combination	$\bar{x}_{ADAPT}$	$\bar{x}_{NONE}$	t-test	Wins Adaptive	Wins None
Stationary, BPS, DBP	5.36	5.34	0	23	23
First Order, BPS, DBP	4.06	5.56	-	14	28
Modelling (Stationary), BPS, DBP	3.88	6.58	-	10	37
Modelling (First Order), BPS, DBP	3.66	6.20	-	5	39
Modelled a priori, BPS, DBP	2.70	6.66	-	5	43

The following sequences shown in Figure 5 help demonstrate a fundamental problem with using the predictions to adapt. They show how using the predictions of occupancy cause the agent to not score a goal. The sequences begin just after the referee has restarted play with an indirect free kick given in the bottom left corner to the team attacking the right goal and with the agent dribbling the ball towards the opponent's goal. Note that the agent has already followed a similar attacking path once before.

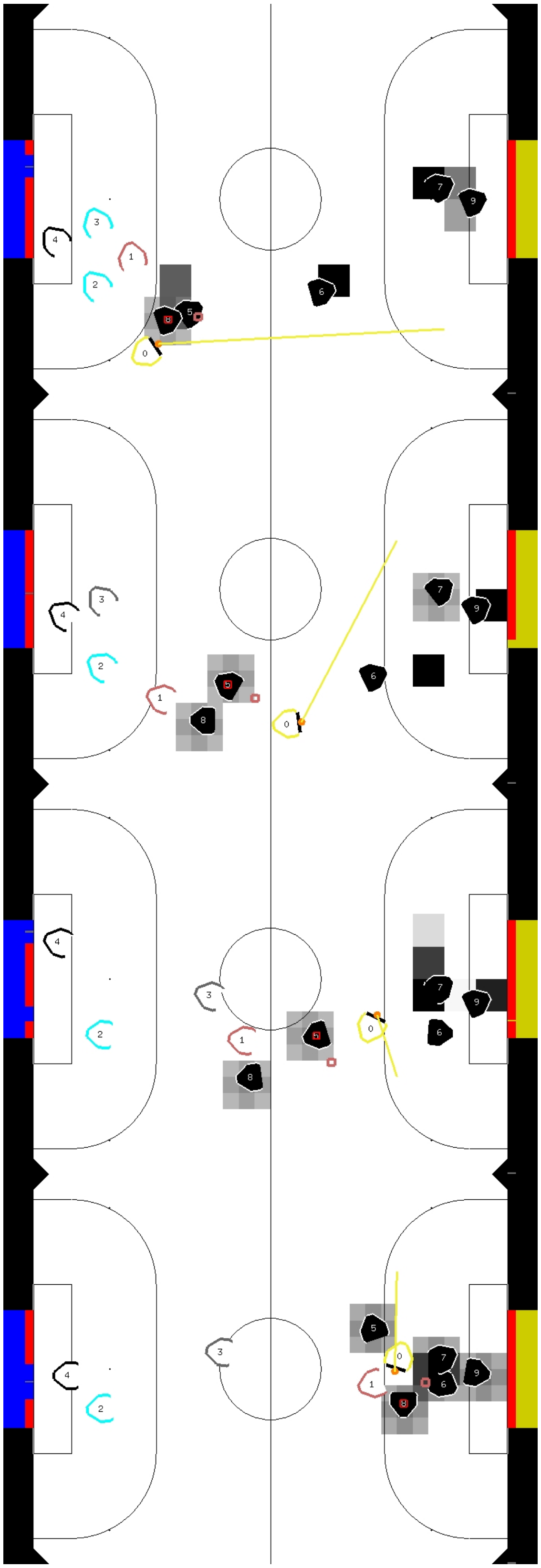
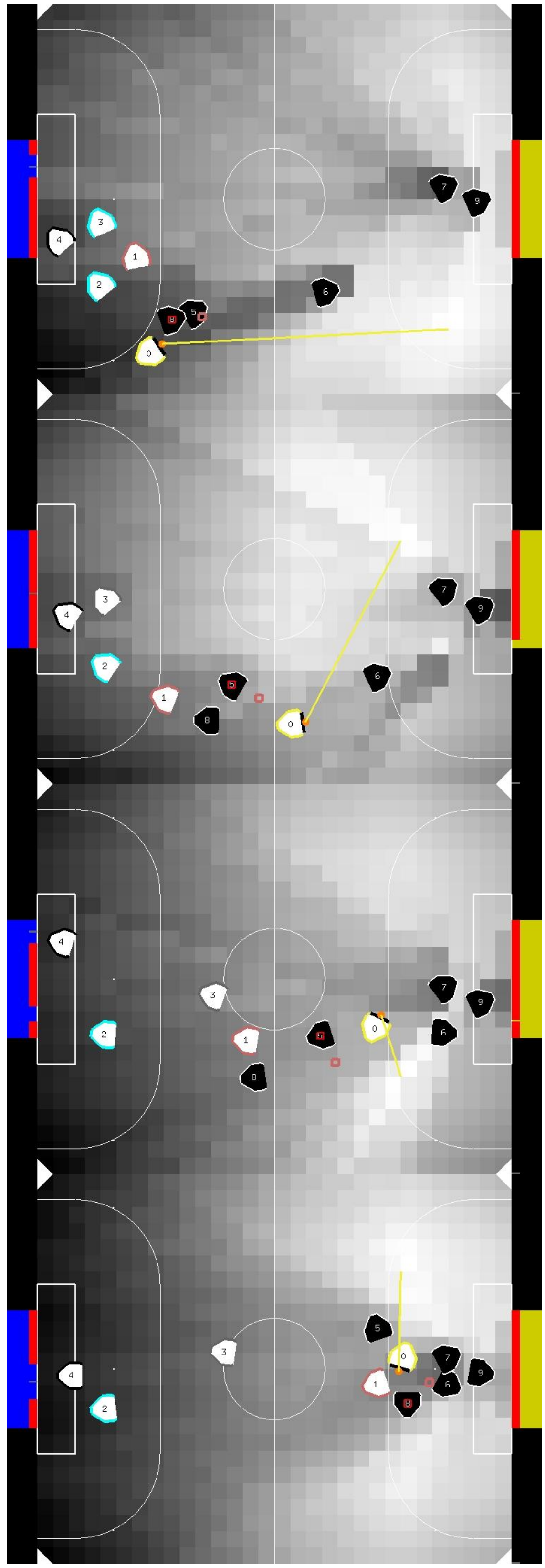


Figure 5 - This figure shows a sequence during a BPS and DK experiment demonstrating how the predictions affect the agent's ability to score a goal. The figures on the left show the opponent's predicted distributions, and on the right the resultant potential field. These frames are not separated by the prediction time.



Note in the third frame from the top how the opponent agent's dribble to location switches to the other side of the opponent's goal. This switch is in response to predicted motion of the defending opponent agents influencing the clear paths to goal potential field.

As the agent crosses the half way line the opponent's defence is predicted to move to the same side of the goal. However, as the agent moves across the goal the defensive agents are predicted to move across the goal. This has the effect of causing the clear paths to show that the other side of the goal is now a more appropriate location to dribble the ball to. This causes the agent to turn around as shown in and the ball is now in a state not previously experienced. As the ball is now in a new state the model can no longer predict the future state of the opponent agents. This means that the dribble to location has again switched to the other side of the goal.

This problem occurs continually throughout the experiments. It causes the dribble-to location to oscillate, resulting in the agent not making positive progress because of indecision. Note that the model will learn this newly experienced agent motion. Next time it will predict both paths for the opponent agents which will again change the choice of dribble-to location. This repeats throughout the experiment and potentially decreases the models usefulness for planning.

Figure 6 shows the average ball location across the field at the end of experiment. The left side shows the distribution for the team that is not adapting (attacking the left goal); the right side shows the distribution for the team that is adapting by predictions based on a model (attacking the right goal). This figure reinforces the discussion above by showing the result of the oscillations and indecision caused by the predictions. Instead of long paths across the goal that explore the action space and weaknesses in the opponents defence, the agent spends most of the time with the ball in the centre of the field unable to make positive motion.

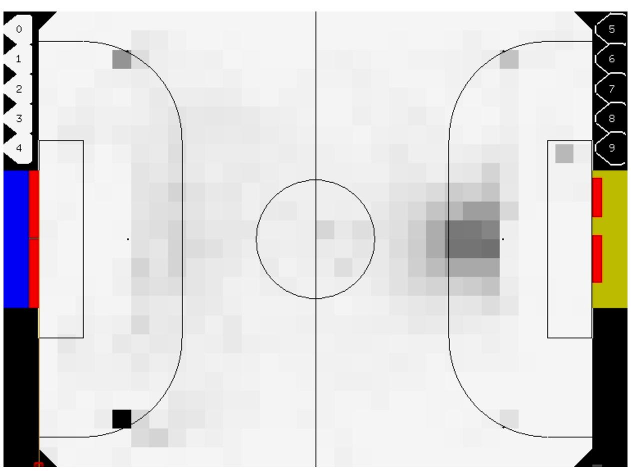


Figure 6 - These two figures show, at the end of a twenty minute experiment using the BPS offence and DBP defence, the average ball position in the opponent's half of the field for: (left) the team not adapting and (right) the team adapting by modelling and prediction. Note that the distribution is more spread out in (left) compared to (right) which indicates a greater exploration of the action space. The limited action space in (right) is a result of the indecision and oscillations discussed. The four dark squares near the corners are the restart points for indirect free kicks.

4.4 All Experiments Results

Table 2 shows the results for all the experiments. The first column shows the particular combination of the prediction method, offence, and defence strategy. The second and third columns show the sample mean goals for both the team that was adapting and the team that was not adapting.

These results show that the performance of the adaptation approach across the different combinations of offensive and defensive strategies are similar to those in the case study above. They show that this adaptation approach significantly reduces performance when using predictions based on the modelling approach in the previous section.

Table 2 - These results show the performance of adapting using the various prediction methods. The table shows the testing combination, the mean goals scored per game scored by both teams and the result of the 5% t-test. The table also shows the number of wins for both teams. Note that these may not necessarily add to fifty games due to tied games.

Testing Combination	$\bar{x}_{ADAPT}$	$\bar{x}_{NONE}$	t-test	Wins Adaptive	Wins None
No Adaptation					
BPS, DBP	5.36	5.34	0	23	23
BPS, GS	5.10	5.26	0	20	20
CS, DBP	11.34	12.20	0	18	25
CS, GS	14.74	15.08	0	20	24
First Order					
BPS, DBP	4.06	5.56	-	14	28
BPS, GS	5.30	5.10	0	21	18
CS, DBP	8.16	12.96	-	7	40
CS, GS	9.56	15.42	-	6	44
Modelling (Stationary)					
BPS, DBP	3.88	6.58	-	10	37
BPS, GS	5.60	5.92	0	19	25
CS, DBP	9.26	13.12	-	9	39
CS, GS	8.88	16.58	-	3	46
Modelling (First Order)					
BPS, DBP	3.66	6.20	-	5	39
BPS, GS	4.70	5.76	-	16	27
CS, DBP	6.42	14.00	-	0	48
CS, GS	8.88	16.58	-	3	46
Modelled a priori					
BPS, DBP	2.70	6.66	-	5	43
BPS, GS	4.08	5.48	-	16	27
CS, DBP	8.66	15.48	-	4	46
CS, GS	7.14	13.24	-	3	44

4.5 Tuning of MAPS Parameters

MAPS has a variety of weights and field width parameters that are tuned by observation of the team to maximise performance. For example, creating the highly successful dribbling play was achieved by setting appropriate balances between the Basefield, Clear Path, Distance From, Opponent Regions, and Previous Dribble-To Regions. In the adaptive experiments, the weighting of the

Clear Path and Opponent Regions were increased to account the weaker strength of the grid cell values in the distributed representations.

Further tuning of the Clear Path, Opponent Regions and Previous Dribble-To Region was performed in an attempt to eliminate the vacillation illustrated in the case studies. The tuning was based on a static prediction model generated a priori from long periods of play. The observations during this tuning process were that the underlying predictive distributions would always lead to large changes in the point of attack, and that no amount of linear weighting against other fields, or increased hysteretic effect by altering the Dribble-To weight could alleviate the problem. The level of hysteresis required to prevent the oscillation stagnates the target location so much that the attacker would attempt to dribble straight into a changing defence. The problem appears to be rooted in the predictive data, and the way that a potential field based planner such as MAPS can use the predictive data.

5 Discussion

The results show that this approach to adaptation by prediction of agent motion negatively impacts the number of goals scored and the number of games won in all but two of the adaptation experiments. This includes the critical experiments where the modelling and adaptation is occurring simultaneously, as it would in a competition game. It is also relevant to compare the mean goals in the adaptation experiments with those without. This also shows a negative impact on the mean goals scored by the team that was adapting.

The possibility that the lack of accuracy in early predictions caused poor planning performance was thoroughly investigated. The last set of experiments where the model learned *a priori* for a significant amount of time was included to test this possible problem. However, it shows that even with a longer training time there is still poor performance. The results even indicate that a longer learning time gives worse performance. This could show that as the predictions become more accurate the adaptation performance decreases, or it could be that, as learning did not occur while adaptation was taking place, the predictions were no longer accurate.

The case study illustrates the cause of poor performance repeatedly observed during the experiments. Planner indecision rooted in the predictive fields for defender locations caused oscillations in the desired dribble-to location. As the agent with the ball reaches a certain point the predictions show that the opponent defence will have moved into a position covering the goal. When MAPS is building the dribble field component representing clear paths for shooting at the goal the dribble-to target location will have moved. The system is predicting (correctly) that there is a low likelihood of a gap there by the time the agent dribbles the ball to the location and the other side of the goal seems more appropriate now. This behaviour is mirrored on the other side of the goal and causes oscillation. The result of this oscillating behaviour is that there is no positive or exploratory action taken and eventually the opponents tackle the agent that has the ball. By contrast, an agent that is unable to predict the likely future performs a high risk strategy that occasionally pays off.

This is a problem with using short term local predictions for planning with a potential field based planner. Attempts to tune this behaviour out of the system by modifying the MAPS parameters were not effective, as the underlying predictive fields bias the attack in the wrong direction no matter the weightings. To alleviate some issues with oscillation, MAPS biases the previous action location (in this case the dribble to location) but this does not solve the problem for a predictive agent. Global long term planning could potentially help address these issues and ensure exploration and positive action. However, this is challenging in such a highly dynamic environment.

An observed long term effect is that the model builds up a broad range of predictions of the opponent agent's motion, from one particular grid cell. This is because of the following repeated process:

- the model learns a new agent motion, or the transition probabilities change,
- the model changes which causes the resultant MAPS potential field to change,
- the selected role location changes,
- the agent's move differently, and
- the opponent agent's may change their motion.

This repeats each time the agent visits a grid cell. Further, as soon as MAPS makes a change in response to a prediction, the prediction may no longer be valid for a period longer than the opponent's observation time. This is an issue inherent in this approach where the predictions are not learned relative to the choice of action. This potentially decreases the model's usefulness for planning.

6 Conclusion

This paper investigated adaptation by prediction of agent motion. The results show that the most literal interpretation of "reading the play", by accurately modelling and predicting the future state of dynamic agents, is not necessarily the best resource for creating successful plans. The trend in the results indicates that more accurate predictions of agent occupancy lead to worse goal scoring performance. Is there a fundamental problem with the basis of this approach, is it a problem with the implementation, or is it how the predictions are integrated into the planner? While at first it seems obvious that a planner should be able to exploit knowledge of the future occupancy of the opponent agents, this paper has raised doubts.

It seems counter intuitive that planning becomes worse with more accurate predictions but experimental results demonstrate otherwise. Therefore, future work should investigate other methodologies for integrating predictions of agent occupancy into a planning system.

Douglas Adams and John Lloyd (1983) have named a familiar type of navigation failure:

"Droitwich (n): A street dance. The two partners approach from opposite directions and try politely to get

out of each other's way. They step to the left, step to the right, apologise. Step to the left again, apologise again, bump into each other and repeat as often as unnecessary."

It is exactly this type of behaviour that was observed repeatedly when predictive fields were introduced in the place of broad potential fields for planning, regardless of the parameter values. The nature of the underlying predictive data does not suit a potential field based planner.

References

[Ball and Wyeth, 2004] David Ball and Gordon Wyeth. Multi-Robot Control in Highly Dynamic, Competitive Environments. Springer-Verlag, 2004.

[Ball and Wyeth, 2005] David Ball and Gordon Wyeth. Predicting the State of Agents Using Motion and Occupancy Grids. 2005.

[Ball, *et al.*, 2003] David Ball, Gordon Wyeth and David Cusack. Design of a Team of Soccer Playing Robots. 2003.

[Bowling, 2003] Michael Bowling. Multiagent Learning in the Presence of Agents with Limitations. PhD dissertation. School of Computer Science, Carnegie Mellon University, 2003.

[Bowling, *et al.*, 2004] Michael Bowling, Brett Browning and Manuela Veloso. Plays as Effective Multiagent Plans Enabling Opponent-Adaptive Play Selection. 2004.

[Carmel and Markovitch, 1996] David Carmel and Shaul Markovitch. Incorporating Opponent Models into Adversary Search. AAAI Press, 1996.

[Chen, *et al.*, 2001] Mao Chen, Klaus Dorer, Ehsan Foroughi, Fredrik Heintz, ZhanXiang Huang, Spiros Kapetanakis, Kostas Kostiadis, Johan Kummeneje, Jan Murray, Itsuki Noda, Oliver Obst, Patrick Riley, Timo Steffens, Yi Wang and Xiang Yin. *Soccerserver Manual v7*, RoboCup Federation, 2001.

[Elfes, 1990] A.E. Elfes. Occupancy grids: A Stochastic Spatial Representation for Active Robot Perception. 1990.

[Habibi, *et al.*, 2002] Jafar Habibi, Ehsan Chiniforooshan, A. HeydarNoori, M. Mirzazadeh, MohammadAli Safari and HamidReza Younesi. Coaching a Soccer Simulation Team in RoboCup Environment. Springer-Verlag, 2002.

[Kitano, *et al.*, 1995] Hiroaki Kitano, Minoru Asada, Yasuo Kuniyoshi, Itsuki Noda and Eiichi Osawa. RoboCup: The Robot World Cup Initiative. 1995.

[Kuhlmann, *et al.*, 2005] Gregory Kuhlmann, Peter Stone and Justin Lallinger. "The UT Austin Ailla 2003 Champion Simulator Coach: A Machine Learning Approach" RoboCup 2004: Robot Soccer. World Cup VIII, Lecture Notes in Artificial Intelligence 3276, 2005.

[Maes, 1994] Pattie Maes. "Modeling Adaptive Autonomous Agents" Artificial Life, 1, 135-162, 1994.

[Moravec and Elfes, 1985] Hans P. Moravec and Alberto Elfes. High Resolution Maps from Wide Angle Sonar. 1985.

[Neumann and Morgenstern, 1944] John Von Neumann and Oskar Morgenstern. *Theory of Games and Economic Behavior*, Princeton, Princeton University Press (Reprinted in 2004), 1944.

[Riley and Veloso, 2001] Patrick Riley and Manuela Veloso. "Planning for Distributed Execution Through Use of Probabilistic Opponent Models" IJCAI-2001 Workshop PRO-2: Planning under Uncertainty and Incomplete Information, 2001.

[Riley and Veloso, 2004] Patrick Riley and Manuela Veloso. Advice Generation from Observed Execution: Abstract Markov Decision Process Learning. 2004.

[Steffens, 2004] Timo Steffens. "Feature-based Declarative Opponent-Modelling" RoboCup-2003: Robot Soccer World Cup VI, 2004.

[Takahashi, 1999] Tomoichi Takahashi. "Kasugabito III" RoboCup-99: Robot Soccer World Cup III, Lecture Notes in Computer Science, 1999.

[Tews, 2002] Ashley Desmond Tews. Achieving Multi-Robot Cooperation in Highly Dynamic Environments. PhD dissertation. School of Computer Science and Electrical Engineering, University of Queensland, 2002.

[Thrun, 2002] Sebastian Thrun Robotic Mapping: A Survey. In G. Lakemeyer and B. Nebel, Ed., Exploring Artificial Intelligence in the New Millenium, Morgan Kaufmann, 2002.

[Visser, *et al.*, 2001] Ubbo Visser, Christian Drücker, Sebastian Hübner, Esko Schmidt and Hans-Georg Weland. "Recognizing Formations in Opponent Teams" Lecture Notes in Computer Science, 2019, 2001.

[Wyeth, *et al.*, 2002] Gordon Wyeth, David Ball, David Cusack and Adrian Ratnapala. UQ RoboRoos: Achieving Power and Agility in a Small Size Robot. Springer-Verlag, 2002.