Best-Response Learning of Team Behaviour in Quake III

Sander Bakkes, Pieter Spronck and **Eric Postma**
Universiteit Maastricht
Institute for Knowledge and Agent Technology (IKAT)
P.O. Box 616, NL-6200 MD Maastricht, The Netherlands
{s.bakkes, p.spronck, postma}@cs.unimaas.nl

Abstract

This paper proposes a mechanism for learning a best-response strategy to improve opponent intelligence in team-oriented commercial computer games. The mechanism, called TEAM2, is an extension of the TEAM mechanism for team-oriented adaptive behaviour explored in [Bakkes *et al.*, 2004] and focusses on the exploitation of relevant gameplay experience. We compare the performance of the TEAM2 mechanism with that of the original TEAM mechanism in simulation studies. The results show the TEAM2 mechanism to be better able to learn team behaviour. We argue that the application as an online learning mechanism is hampered by occasional very long learning times due to an improper balance between exploitation and exploration. We conclude that TEAM2 improves opponent behaviour in team-oriented games and that for online learning the balance between exploitation and exploration is of main importance.

1 Introduction

In recent years, commercial computer game developers have emphasised the importance of high-quality game opponent behaviour. *Online learning* techniques may be used to significantly improve the quality of game opponents by endowing them with the capability of adaptive behaviour (i.e., artificial creativity and self-correction). However, to our knowledge online learning has never been used in an actual commercial computer game (henceforth called 'game'). In earlier work [Bakkes *et al.*, 2004], we have proposed a mechanism named TEAM (Team-oriented Evolutionary Adaptability Mechanism) for team-oriented learning in games. Our experiments revealed TEAM to be applicable to commercial computer games (such as Quake-like team-games). Unfortunately, the applicability is limited due to the large variation in the time needed to learn the appropriate tactics.

This paper describes our attempts to improve the efficiency of the TEAM mechanism using *implicit opponent models* [van den Herik *et al.*, 2005]. We propose an extension of TEAM called TEAM2. The TEAM2 mechanism employs a data store of a limited history of results of tactical team behaviour, which constitutes an implicit opponent model, on which a best-response strategy [Carmel and Markovitch, 1997] is formulated. We will ague that *best-response learning* of team-oriented behaviour can be applied in games. We investigate to what extent it is suitable for online learning.

The outline of this paper is as follows. Section 2 discusses team-oriented behaviour (team AI) in general, and the application of adaptive team AI in games in particular. The TEAM2 best-response learning mechanism is discussed in section 3. In section 4, an experiment to test the performance of the mechanism is discussed. Section 5 reports our findings, and section 6 concludes and indicates future work.

2 Adaptive Team AI in Commercial Computer Games

We defined adaptive team AI as the behaviour of a team of adaptive agents that competes with other teams within a game environment [Bakkes *et al.*, 2004]. Adaptive team AI consists of four components: (1) the individual agent AI, (2) a means of communication, (3) team organisation, and (4) an adaptive mechanism.

Figure 1: Screenshot of the game QUAKE III. An agent fires at a game opponent.

The first three components are required for agents to establish team cohesion, and for team-oriented behaviour to emerge. The fourth component is crucial for improving the quality of the team during gameplay. The next sub-sections discuss a mechanism for adaptive team AI, and its performance.

2.1 The Team-oriented Evolutionary Adaptability Mechanism (TEAM)

The observation that humans players prefer to play against other humans over players against artificial opponents [van Rijswijck, 2003], led us to design the Team-oriented Evolutionary Adaptability Mechanism (TEAM). TEAM is an online evolutionary learning technique designed to adapt the team AI of Quake-like games. TEAM assumes that the behaviour of a team in a game is defined by a small number of parameters, specified per game state. A specific instance of team behaviour is defined by values for each of the parameters, for each of the states. TEAM is defined as having the following six properties: 1) state-based evolution, 2) state-based chromosome encoding, 3) state-transition-based fitness function, 4) fitness propagation, 5) elitist selection, and 6) manually-designed initialisation [Bakkes *et al.*, 2004].

For evolving successful behaviour, typical evolutionary learning techniques need thousands of trials (or more). Therefore, at first glance such techniques seem unsuitable for the task of online learning. Laird [2000] is skeptical about the possibilities offered by online evolutionary learning in games. He states that, while evolutionary algorithms may be applied to tune parameters, they are "grossly inadequate when it comes to creating synthetic characters with complex behaviours automatically from scratch". In contrast, the results achieved with the TEAM mechanism in the game QUAKE III show that it is certainly possible to use online evolutionary learning in games.

2.2 Enhancing the Performance of TEAM

Spronck [2005] defines four requirements for qualitatively acceptable performance were defined: speed, robustness, effectiveness, and efficiency. For the present study, the requirement of efficiency is of main relevance. Efficiency is defined as the learning time of the mechanism. In adaptive team AI, efficiency depends on the number of learning trials needed to adopt effective behaviour. Applied to the QUAKE III capture-the-flag (CTF) team game, the TEAM mechanism requires about 2 hours of real-time play to significantly outperform the opponent. Since QUAKE III matches take on average half an hour, the TEAM mechanism lacks efficiency to enable successful online learning in games such as QUAKE III.

When one aims for efficient adaptation of opponent behaviour in games, the practical use of evolutionary online learning is doubtful [Spronck, 2005]. Therefore, the design of TEAM needs to be enhanced with a different approach to learning team-oriented behaviour. The enhanced design, named TEAM2, is discussed next.

3 Best-Response Learning of Team-oriented Behaviour

The design of TEAM2, aimed at efficiently adapting opponent behaviour, is based on a best-response learning approach (instead of evolutionary learning)[1]. This section discusses the properties of the enhanced design: (1) a symbiotic learning concept, (2) learning a best-response team strategy, (3) a state-transition-based fitness function, and (4) a scaled roulette-wheel selection. The popular QUAKE III CTF game [van Waveren and Rothkrantz, 2001], is used for illustrative purposes.

3.1 Symbiotic Learning

Symbiotic learning is a concept for learning adaptive behaviour for *a team as a whole* (rather than learning adaptive behaviour for each individual). The TEAM mechanism successfully applied the concept for the purpose of adapting opponent behaviour in team-oriented games. The onset of the design of TEAM was the observation that the game state of team-oriented games can typically be represented as a finite state machine (FSM). By applying an instance of an adaptive mechanism to each state of the FSM, one is able to learn relatively uncomplicated team-oriented behaviour for the specific state. Cooperatively, from all instances of the applied adaptive mechanism, relatively complex team-oriented behaviour emerges in a computationally fast fashion. The concept of symbiotic learning is illustrated in figure 2. The figure exemplifies how instances of an adaptive mechanism cooperatively learn team-oriented behaviour, which is defined as the combination of the local optima for the states (in this example there are four states).

An instance of the adaptive mechanism automatically generates and selects the best team-configuration for the specific state. A team-configuration is defined by a small number of parameters which represent team behaviour (e.g. one team-configuration can represent an offensive tactic, whereas another team-configuration can represent a defensive tactic).

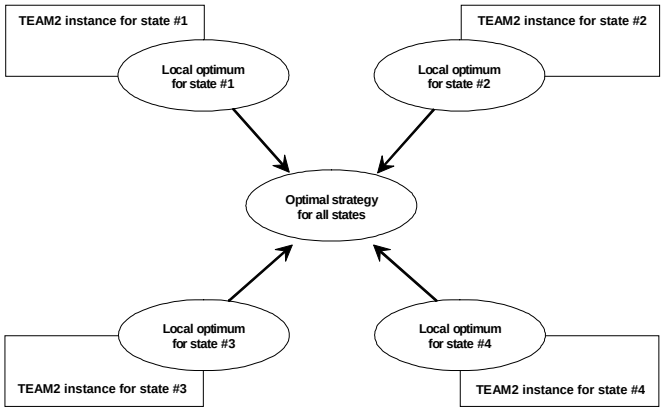

Figure 2: Symbiotic learning.

[1]Since TEAM2 is not inspired by evolutionary algorithms, we let the reader imagine that the letter 'E' is an abbreviation for 'Exploitative' (instead of 'Evolutionary').

3.2 Learning a Best-Response Team Strategy

Adaptation to the opponent takes place via an implicit opponent model, which is built and updated when the team game is in progress. Per state of the game, the sampled data merely concerns the specific state and represents all possible team-configurations for the state. The implicit opponent model consists of historic data of results per team-configuration per state. An example of the structure of an implicit opponent model is given in table 1. In the example, the team-configuration represents the role division of a team with four members. Each of which has either an offensive, a defensive or an roaming role. The history can anything from a store of fitness values, to a complex data-structure.

Team configuration	*History*	*Fitness*
(0,0,4)	[0.1,0.6,...,0.5]	0.546
(0,1,3)	[0.3,0.1,...,0.2]	0.189
⋮	⋮	⋮
(4,0,0)	[0.8,0.6,...,0.9]	0.853

Table 1: Example of an implicit opponent model for a specific state of the QUAKE III capture-the-flag game.

On this basis, a best-response strategy is formulated when the game transits from one state to another. For reasons of efficiency and relevance, only recent historic data are used for the learning process.

3.3 State-transition-based Fitness Function

The TEAM2 mechanism uses a fitness function based on state transitions. Beneficial state transitions reward the tactic that caused the state transition, while detrimental state transitions penalise it. To state transitions that directly lead to scoring (or losing) a point, the fitness function gives a reward (or penalty) of 4. Whereas to the other state transitions, the fitness function gives a reward (or penalty) of 1. This ratio is empirically decided by the experimenters. In figure 3, an example of annotations on the FSM of the QUAKE III CTF game is given.

Usually, judgement whether a state transition is beneficial or detrimental cannot be given immediately after the transition; it must be delayed until sufficient game-observations are gathered. For instance, if a state transition happens from a state that is neutral for the team to a state that is good for the team, the transition seems beneficial. However, if this is immediately followed by a second transition to a state that is bad for the team, the first transition cannot be considered beneficial, since it may have been the primary cause for the second transition.

3.4 Scaled Roulette-Wheel Selection

The best-response learning mechanism selects the preferred team-configuration by implementing a roulette wheel method [Nolfi and Floreano, 2000], where each slot of the roulette wheel corresponds to a team-configuration in the state-specific solution space, and the size of the slot is proportional to the obtained fitness-value of the team-configuration. The selection mechanism quadratically scales the fitness values to select the higher-ranking team-configurations more often, acknowledging that game opponent behaviour must be non-degrading. In acknowledgement of the inherent randomness of a game environment, the selection mechanism protects against selecting inferior top-ranking team-configurations.

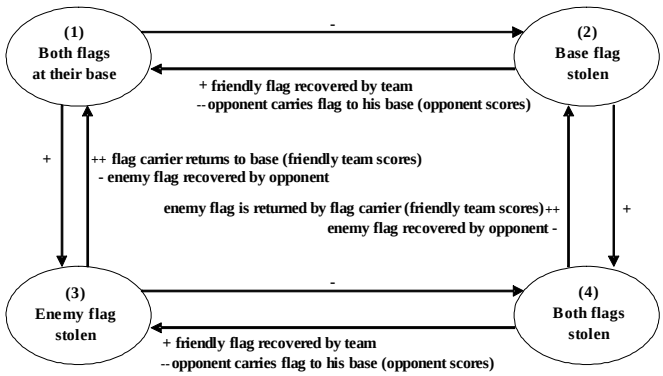

Figure 3: Annotated finite state machine of QUAKE III CTF. Highly beneficial and beneficial transitions are denoted with "$++$" and "$+$" respectively, whereas detrimental and highly detrimental state transitions are denoted with "$-$" and "$--$" respectively.

4 Experimental Study of the TEAM2 Mechanism

To assess the efficiency of the TEAM2 mechanism, we incorporated it in the QUAKE III CTF game. We performed an experiment in which an adaptive team (controlled by TEAM2) is pitted against a non-adaptive team (controlled by the QUAKE III team AI). In the experiment, the TEAM2 mechanism adapts the tactical behaviour of a team to the opponent. A tactic consists of a small number of parameters which represent the offensive and defensive division of roles of agents that operate in the game.

The inherent randomness in the QUAKE III environment requires the learning mechanism to be able to successfully adapt to significant behavioural changes of the opponent. Both teams consist of four agents with identical individual agent AI, identical means of communication and an identical team organisation. They only differ in the control mechanism employed (adaptive or non-adaptive).

4.1 Experimental Setup

An experimental run consists of two teams playing QUAKE III CTF until the game is interrupted by the experimenter. On average, the game is interrupted after two hours of gameplay, since the original TEAM mechanism typically requires two hours to learn successful behaviour, whereas the TEAM2 mechanism should perform more efficiently. We performed 20 experimental runs with the TEAM2 mechanism. The results obtained will be compared to those obtained with the TEAM mechanism (15 runs, see [Bakkes *et al.*, 2004]).

4.2 Performance Evaluation

To quantify the performance of the TEAM2 mechanism, we determine the so-called turning point for each experimental run. The turning point is defined as the time step at which the adaptive team takes the lead without being surpassed by the non-adaptive team during the remaining time steps.

We defined two performance indicators to evaluate the efficiency of TEAM2: the median turning point and the mean turning point. Both indicators are compared to those obtained with the TEAM mechanism. The choice for two indicators is motivated by the observation that the amount of variance influences the performance of the mechanism [Bakkes *et al.*, 2004].

To investigate the variance of the experimental results, we defined an outlier as an experimental run which needed more than 91 time steps to acquire the turning point (the equivalent of two hours).

4.3 Results

In table 2 an overview of the experimental results of the TEAM2 experiment is given. It should be noted that in two tests, the run was prematurely interrupted without a turning point being reached. We incorporated these two test as having a turning as high as the highest outlier, which is 358. Interim results indicate that, should the runs be not prematurely interrupted, their turning points would have been no more than half of this value.

The median turning point acquired is 38, which is significantly lower that the median turning point of the TEAM mechanism, which is 54. The mean turning point acquired with TEAM2, however, is significantly higher than the mean turning point acquired with the TEAM mechanism (102 and 71, respectively). The percentage of outliers in the total number of tests is about equal. However, the range of the outliers has significantly increased for TEAM2.

To illustrate the course of an experimental run, we plotted the performance for a typical run in figure 4. The performance is expressed in terms of the lead of the adaptive team, which is defined as the score of the adaptive team minus the score of the non-adaptive team. The graph shows that, initially, the adaptive team attains a lead of approximately zero. At the turning point (labeled 38 in figure 4), the adaptive team takes the lead over the non-adaptive team. Additionally, the graph reveals that the adaptive team outperforms the non-adaptive team without any significant degradation in its performance.

		TEAM	TEAM2
# Experiments	*Total*	15	20
	Outliers	4	6
	Outliers in %	27%	30%
Mean		71.33	102.20
Std. Deviation		44.78	125.29
Std. Error of Mean		11.56	28.02
Median		54	38
Range		138	356
Minimum		20	2
Maximum		158	358

Table 2: Summary of experimental results. With TEAM2 the median turning point is significantly lower, yet, outliers have a negative effect on the mean turning point.

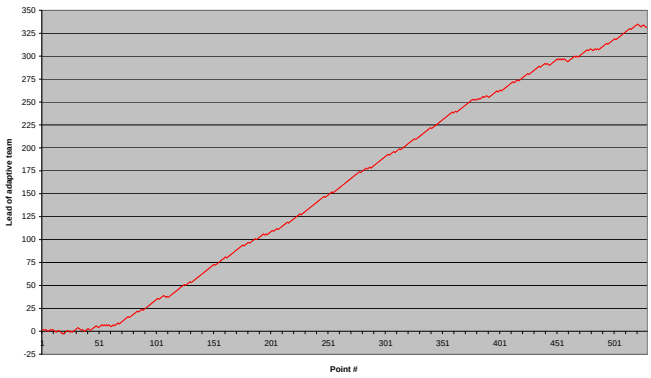

Figure 4: Illustration of typical experimental results obtained with the TEAM2 mechanism. The graph shows the lead of the adaptive team over the non-adaptive team as a function of the number of scored points.

4.4 Evaluation of the Results

The experimental results show that TEAM2 is able to successfully adapt game opponent behaviour in an highly non-deterministic environment, as it challenged and defeated the fine-tuned QUAKE III team AI.

The results listed in table 1 show that the TEAM2 mechanism outperforms the TEAM mechanism in terms of the median turning point. However, the mean turning point is larger for TEAM2 than for TEAM, which is explained by the increased range of the outliers. The median turning point indicates that the TEAM2 best-response learning mechanism is more efficient than the TEAM online evolutionary learning mechanism, as the adaptation to successful behaviour progresses more swiftly than before; expressed in time only 48 minutes are required (as compared to 69 minutes).

Therefore, we may draw the conclusion that the TEAM2 mechanism exceeds the applicability of the TEAM mechanism for the purpose of learning in games. The qualitative acceptability of the performance is discussed next.

5 Discussion

Our experimental results show that the TEAM2 mechanism succeeded in enhancing the learning performance of the TEAM mechanism with regard to its median, but not mean, efficiency. In sub-section 5.1 we give a comparison of the learned behaviour of both mechanisms. Sub-section 5.2 discusses the task of online learning in a commercial computer game environment with regard to the observed outliers.

5.1 Comparison of the Behaviour Learned by TEAM and TEAM2

In the original TEAM experiment we observed that the adaptive team would learn so-called "rush" tactics. Rush tactics aim at quickly obtaining offensive field supremacy. We noted that the QUAKE III team AI, as is was designed by the QUAKE III developers, uses only moderate tactics in all states, and therefore, it is not able to counter *any* field supremacy.

The TEAM2 mechanism is inclined to learn rush tactics as well. Notably, the experiment showed that if the adaptive team uses tactics that are slightly more offensive than the non-adaptive team, it is already able to significantly outperform the opponent. Besides the fact that the QUAKE III team AI cannot adapt to superior player tactics (whereas an adaptive mechanism can), it is not sufficiently fine-tuned; for it implements an obvious and easily detectable local-optimum.

5.2 Exploitation versus Exploration

In our experimental results we noticed that the exploitative TEAM2 mechanism obtained a significant difference between the relatively low median and relatively high mean performance, whereas the original, less exploitative, TEAM mechanism obtained a moderate difference between the median and mean performance. This difference is illustrated in figure 5. It reveals that the exploitative TEAM2 mechanism obtained a significant difference between the relatively low median and relatively high mean performance, whereas the original, less exploitative, TEAM mechanism obtained a moderate difference between the median and mean performance.

An analysis of the phenomenon revealed that it is due to a well-known dilemma in machine learning [Carmel and Markovitch, 1997]: the exploitation versus exploration dilemma. This dilemma entails that a learning mechanism requires the exploration of derived results to yield successful behaviour in the future, whereas at the same time the mechanism needs to directly exploit the derived results to yield successful behaviour in the present. Acknowledging the need for an enhanced efficiency, the emphasis of the TEAM2 mechanism lies on exploiting the data represented in a small amount of samples.

In the highly non-deterministic QUAKE III environment, a long run of fitness values may occur that, due to chance, is not representative for the quality of the tactic employed. Obviously, this problem results from the emphasis on exploiting the small samples taken from the distribution of all states. To increase the number of samples, an exploration mechanism can be added. The TEAM online evolutionary learning mechanism employed such an exploration mechanism with a fitness propagation technique, which led to loss of efficiency. We tested several exploration mechanisms in TEAM2, which we found also led to loss of efficiency. However, since it is impossible to rule out chance runs completely, an online learning mechanism must be balanced between an exploitative and explorative emphasis.

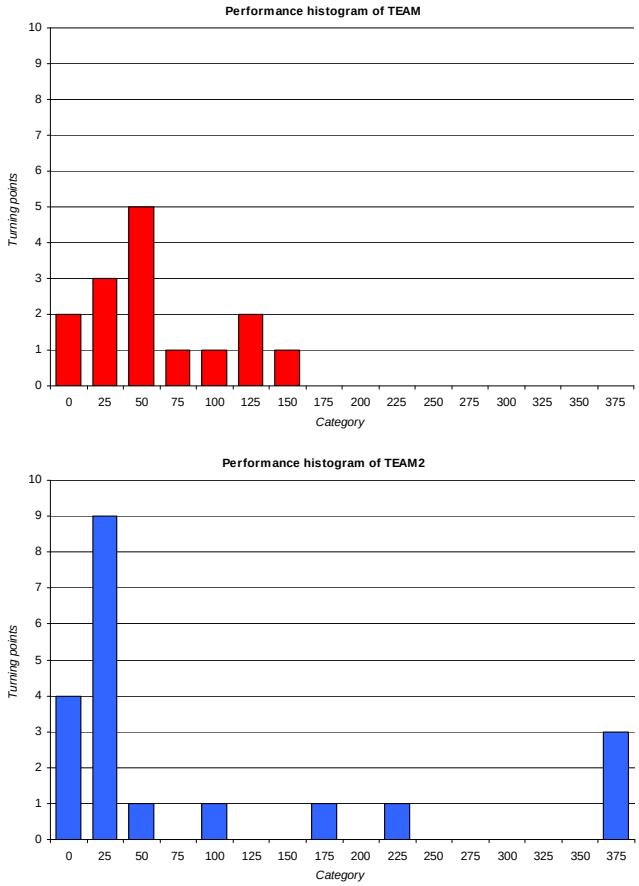

Figure 5: Histograms of the results of both the TEAM2 and TEAM experiment. The graphs show the number of turning points as a function of the value of the turning point, grouped by a category value of 25.

6 Conclusions and Future Work

The TEAM2 mechanism was proposed as an enhancement to the novel Tactics Evolutionary Adaptability Mechanism (TEAM), designed to impose adaptive behaviour on opponents in team-oriented games. The original TEAM mechanism is capable of unsupervised and intelligent adaptation to the environment, yet, its efficiency is modest. From the experimental results of the best-response learning experiment, we drew the conclusion that the TEAM2 best-response learning mechanism succeeded in enhancing the median, but not mean, learning performance. This reveals that in the current experimental setup the exploitation and exploration are not sufficiently well balanced to allow efficient and effective online learning in an actual game. As the TEAM2 mechanism is easily able to defeat a non-adaptive opponent, we may therefore conclude that the mechanism is suitable for online learning in an actual game if, and only if, a balance between exploitation and exploration is found for that specific game. Moreover, the TEAM2 mechanism can be used during game development practice to automatically validate and produce

AI that is not limited by a designer's vision.

Future research should investigate how an effective balance between exploitation of historic data and exploration of alternatives can be achieved. We propose to create a data store of gameplay experiences relevant to decision making processes, and use it to build an opponent model. Thereupon, game AI can either predict the effect of actions it is about to execute, or explore a more creative course of action.

Acknowledgements

This research was funded by a grant from the Netherlands Organization for Scientific Research (NWO grant No 612.066.406).

References

[Bakkes *et al.*, 2004] Sander Bakkes, Pieter Spronck, and Eric Postma. TEAM: The Team-oriented Evolutionary Adaptability Mechanism. In Matthias Rauterberg, editor, *Entertainment Computing - ICEC 2004*, volume 3166 of *Lecture Notes in Computer Science*, pages 273–282. Springer-Verlag, September 2004.

[Carmel and Markovitch, 1997] David Carmel and Shaul Markovitch. Exploration and adaptation in multiagent systems: A model-based approach. In *Proceedings of The Fifteenth International Joint Conference for Artificial Intelligence*, pages 606–611, Nagoya, Japan, 1997.

[Laird, 2000] John E. Laird. Bridging the gap between developers & researchers. *Game Developers Magazine*, Vol 8, August 2000.

[Nolfi and Floreano, 2000] Stefano Nolfi and Dario Floreano. *Evolutionary Robotics*. MIT Press, 2000. ISBN 0-262-14070-5.

[Spronck, 2005] Pieter Spronck. *Adaptive Game AI*. PhD thesis, SIKS Dissertation Series No. 2005-06, Universiteit Maastricht (IKAT), The Netherlands, 2005.

[van den Herik *et al.*, 2005] Jaap van den Herik, Jeroen Donkers, and Pieter Spronck. Opponent modelling and commercial games. Universiteit Maastricht (IKAT), The Netherlands, 2005.

[van Rijswijck, 2003] Jack van Rijswijck. Learning goals in sports games. Department of Computing Science, University of Alberta, Canada, 2003.

[van Waveren and Rothkrantz, 2001] Jean-Paul van Waveren and Leon Rothkrantz. Artificial player for Quake III Arena. In Norman Gough Quasim Mehdi and David Al-Dabass, editors, *Proceedings of the 2nd International Conference on Intelligent Games and Simulation GAME-ON 2001*, pages 48–55. SCS Europe Bvba., 2001.