SYMBIOTIC LEARNING IN COMMERCIAL COMPUTER GAMES

Sander Bakkes and Pieter Spronck
Institute for Knowledge and Agent Technology (IKAT)
Universiteit Maastricht
The Netherlands
E-mail: s.bakkes, p.spronck@cs.unimaas.nl

KEYWORDS
Adaptive behaviour, team-play, machine learning, commercial computer games.

ABSTRACT

This paper proposes an approach for learning team behaviour to improve agent intelligence in team-oriented commercial computer games. The approach, named 'symbiotic learning', focuses on the exploitation of relevant gameplay experiences. The results of an experiment in the game QUAKE III show the symbiotic learning approach to be able to successfully learn effective agent behaviour. We conclude that symbiotic learning can be used during game development practice to automatically validate and produce AI, and, provided a good balance is found between exploitation and exploration, the approach can be applied in practice for the purpose of online learning in commercial computer games.

INTRODUCTION

One of the fundamental goals of artificial intelligence (AI) research is the development and understanding of human-level intelligence. Laird (2001) argues there is little active research directly pursuing that goal, and suggests to research AI in a testbed that offers complex and realistic environments for emergent agent behaviour. Laird says that current commercial computer games are so realistic, that agents operating in these environments are required to behave realistically. An agent's behaviour in a game is determined by its game AI. Currently, in state-of-the-art games, the game AI still lacks the characteristic human-level capability of adaptive behaviour (i.e., artificial creativity and self-correction).

In our research, we focus on the game AI of agents that function as a team. Specifically, we investigate a team's ability to automatically learn to improve its performance in a so-called 'Capture the Flag' (CTF) team game. In such team games, every agent must be capable of making (or following) a long-term plan, while reacting to and taking short-term actions. Due to the nature of a team game, game agents are only successful if they cooperate.

It has been observed that in CTF games, human players prefer to play against other humans instead of against game agents (Schaeffer 2001, Rijswijck van 2003), which is often attributed to unrealistic, ineffective and non-entertaining agent behaviour. With adaptive behaviour, a team of computer-controlled agents might become a more interesting challenge for human players. Therefore, our research goal is to endow game agents with effective adaptive team behaviour. This paper discusses our approach to achieve this goal, which we named 'symbiotic learning'. It describes the concept of symbiotic learning, and practical experiments with symbiotic learning in the action game QUAKE III.

The outline of the paper is as follows. The concept of symbiotic learning is discussed in section 2. Section 3 discusses an adaptive mechanism for team behaviour. In section 4, an experiment to test the performance of the mechanism is discussed. Section 5 reports our findings, and section 6 concludes with predictions of the state of AI in future games.

SYMBIOTIC LEARNING

Symbiotic learning is a concept based on the simple idea that for team-oriented AI, a learning mechanism should learn behaviour for *a team as a* *whole* (rather than learning behaviour for each individual). The concept's design goal is to allow effective adaptation of agent behaviour in team-oriented games. The onset of the symbiotic learning concept was the observation that the game state of team-oriented games usually is relatively easy to represent. In typical team-oriented action games, team behaviour is represented as a small number of parameters which define global agent behaviour, and a finite state machine (FSM) with only a few states. The concept of symbiotic learning is illustrated in figure 1.

Inspired by representing behaviour on a per-state basis, the symbiotic learning approach comprises multiple instances of an adaptive mechanism (one instance per state) that cooperatively learn team-oriented behaviour. The team behaviour is defined as the combination of the local optima of each instance. Cooperatively, from all instances of the applied adaptive mechanism, relatively complex team-oriented behaviour emerges in a computationally fast fashion. An instance can incorporate any adaptive mechanism that is capable of learning partial team behaviour, and is not necessarily restricted to one particular machine learning technique.

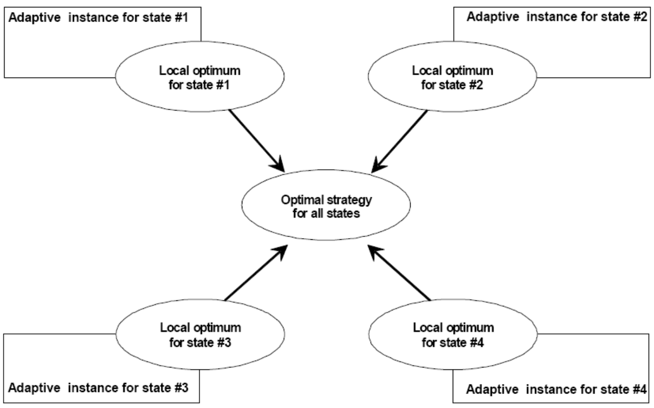

Figure 1: Example of symbiotic learning applied to a team game with four states.

BEST-RESPONSE LEARNING OF TEAM BEHAVIOUR

We applied the concept of symbiotic learning to team behaviour in the capture-the-flag team-based mode of the game QUAKE III. For the present research, we chose to implement the adaptive mechanism for team behaviour as best-response learning. The adaptive mechanism is particularly aimed at the efficient exploitation of relevant gameplay experiences. Each 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 specific team configuration can represent an offensive tactic, whereas another team configuration can represent a defensive tactic).

Adaptation takes place via an *implicit opponent model* (Van den Herik et al., 2005), which is built and updated when the team game is in progress. Per state of the game, the sampled data represents all possible team-configurations for the state. The implicit opponent model consists of historic data of results per team-configuration per state. In our research, we implemented the implicit opponent model as a simple history for storing fitness values (see table 1 for an example), though a more complex data-structure can be used if this is needed. In the example, the team configuration represents the role division of a team with four members. Each team member has either an offensive, a defensive or a roaming role. On the basis of history results, 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 is used for the learning process.

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.

We evaluate the fitness of a team configuration by a fitness function that monitors which team configuration leads to which state-transition. The basis of the fitness function is demarcating between ‘beneficial’ and ‘detrimental’ state-transitions (and possible nuances of these). 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. In figure 2, an example of annotations on the FSM of the QUAKE III CTF game is given.

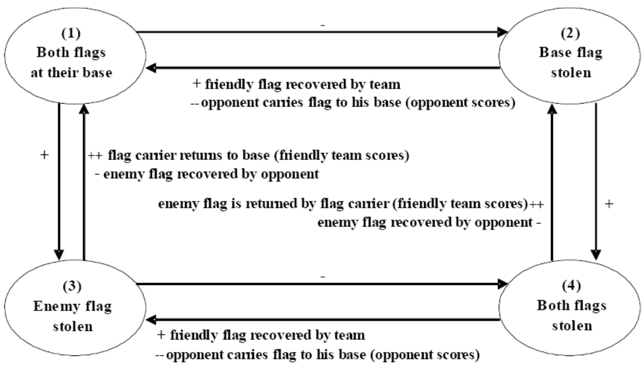

Figure 2: Annotated finite state machine of QUAKE III CTF. Highly beneficial and beneficial transitions are denoted with "++" and "+" respectively, whereas detrimental and highly detrimental state transitions are denoted with "-" and "--" respectively.

The adaptive mechanism selects the preferred team-configuration by implementing a roulette wheel selection method (Nolfi 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 scales the fitness values to select the higher-ranking team-configurations more often, acknowledging that game agent 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.

EXPERIMENTAL STUDY OF THE LEARNING APPROACH

We performed an experiment in order to assess whether the symbiotic learning approach endows agents with effective team behaviour. For effective learning, the inherent randomness in the QUAKE III environment requires the learning mechanism to be able to adapt successfully to significant behavioural changes of the opponent.

Experimental Setup and Performance Evaluation

We performed an experiment in which an adaptive team (controlled by the learning mechanism) is pitted against a non-adaptive team (controlled by the QUAKE III team AI). An experimental run consists of two teams playing QUAKE III CTF until the game is interrupted by the experimenter. In the experiment, the learning mechanism adapts the configuration of a team to the opponent. Both teams consist of four agents with identical individual agent AI. They only differ in the control mechanism employed (adaptive or non-adaptive).

To quantify the performance of the learning 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 the learning mechanism, namely 1) the median turning point, and 2) the mean turning point.

In an actual commercial computer game, the learning mechanism should learn efficiently, thus we decided to allow the learning mechanism only a relatively short time to learn successful behaviour. For application in commercial computer games, a learning time of longer than two hours of real-time gameplay is not acceptable. Therefore, we defined an outlier as an experimental run with a turning point of 91 or above (the equivalent of two hours of real-time gameplay, on average).

Results

In table 2 an overview of the experimental results is given. The median turning point acquired is 45. The mean turning point acquired is 67. The percentage of outliers in the total number of tests is relatively high, viz 23%. To illustrate the course of an experimental run, we plotted the performance for a typical run in figure QUAKE III. 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 (time step 38 in figure 3), 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.

The experimental results show that in all runs the learning mechanism is able to successfully adapt game agent behaviour in an highly non-deterministic environment, as it challenged and defeated the fine-tuned QUAKE III team AI. Therefore, we may draw the conclusion that the symbiotic learning approach can be successfully applied for the purpose of learning agent behaviour in team games. The qualitative acceptability of the performance is discussed next.

# Experimental runs	*Total*	30
	Outliers	7
	Outliers in %	23.3%
Median		45.0
Mean		67.4
Std. Deviation		80.4
Std. Error of Mean		14.7

Table 2: Summary of experimental results.

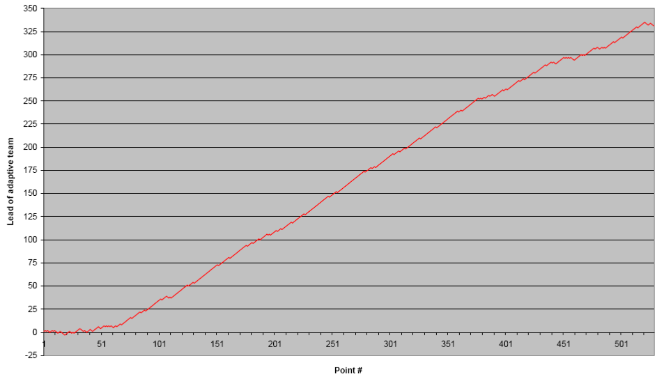

Figure 3: Illustration of typical experimental results obtained with the learning mechanism. The graph shows the lead of the adaptive team over the non-adaptive team as a function of the number of scored points.

DISCUSSION

In the experiment we observed that the adaptive team is inclined to 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 field supremacy. 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.

Additionally, in our experimental results we noticed that the learning mechanism obtained a notable difference between the median and mean performance. This difference is illustrated by a histogram in figure 4. An analysis of the phenomenon points to a well-known dilemma in machine learning: the exploitation versus exploration dilemma (Carmel and Markovitch, 1997). 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 a significant efficiency, the emphasis of the adaptive mechanism lies on exploiting the data represented in a small number of samples. However, in a highly non-deterministic environment (such as QUAKE III), a long run of fitness values may occur that, due to chance, are not representative for the quality of the behaviour employed, which results in a relatively high variance in the learning performance. Exploration mechanisms can be incorporated to enforce more consistent learning, however, this in turn would have a negative effect on the efficiency of the mechanism. Therefore, a learning mechanism only can be used in practice if, for a particular game, it is sufficiently balanced between an exploitative and explorative emphasis.

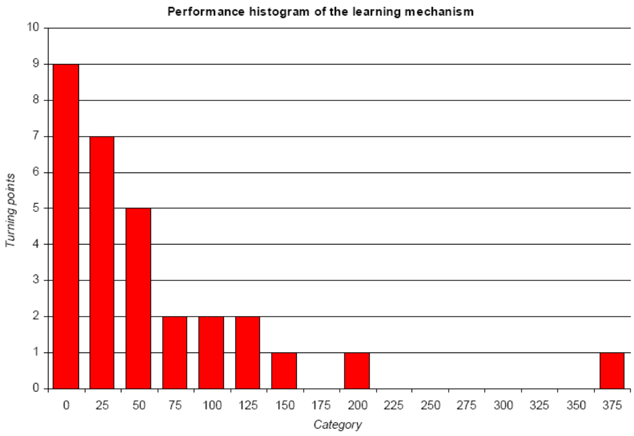

Figure 4: Histogram of the results of the adaptive mechanism experiment. The graphs show the number of turning points as a function of the value of the turning point, grouped by a category value of 25.

CONCLUSIONS AND FUTURE WORK

The symbiotic learning approach was proposed as a design to impose adaptive behaviour on opponents in team-oriented games. From the experimental results of our QUAKE III capture-the-flag experiment, we drew the conclusion that, since successful agent behaviour was discovered in all experimental runs, symbiotic learning can be applied in actual games. For instance, symbiotic learning can be used during game development practice to automatically validate and produce AI that is not limited by a designer's vision. From our analysis of the experimental results, we may draw the conclusion that the symbiotic learning approach can be applied for online learning in a game if a good balance between exploitation and exploration is found for that specific game.

We predict that, in order to cope with the exploitation versus exploration dilemma, online learning techniques will increasingly utilize large data-stores of experiences. These experiences can be used by decision making processes to either predict the effect of actions it is considers to execute, or explore a more creative course of action. With massive multiplayer online games, storing and using experiences of thousands of players will be feasible in the near future.

ACKNOWLEDGEMENTS

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

REFERENCES

Carmel, D. and Markovitch, S., 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.

Herik, J. van den, Donkers J. and Spronck P., Opponent modelling and commercial games. Universiteit Maastricht (IKAT), The Netherlands, 2005.

Laird, J.E. and Lent, M. van, Human-level ai's killer application: Interactive computer games. *AI Magazine*, 22(2):15-25, 2001.

Nolfi, S. and Floreano, D., *Evolutionary Robotics*. MIT Press, 2000. ISBN 0-262-14070-5.

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

Schaeffer, J., A gamut of games. *AI Magazine*, 22(3):29-46, 2001.