Adaptive Agent-Based Simulation for Individualized Training

Doctoral Consortium

Johan Källström
Linköping University
Linköping, Sweden
johan.kallstrom@liu.se

ABSTRACT

Agent-based simulation can be used for efficient and effective training of human operators and decision-makers. However, constructing realistic behavior models for the agents is challenging and time-consuming, especially for subject matter experts, who may not have expertise in artificial intelligence. In this work, we investigate how machine learning can be used to adapt simulation contents to the current needs of individual trainees. Our initial results demonstrate that multi-objective multi-agent reinforcement learning is a promising approach for creating agents with diverse and adaptive characteristics, which can stimulate humans in training.

KEYWORDS

Modelling for agent based simulation; Agents competing and collaborating with humans; Agents for improving human cooperative activities; Reinforcement learning; Multi-agent learning

ACM Reference Format:

Johan Källström. 2020. Adaptive Agent-Based Simulation for Individualized Training. In *Proc. of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2020), Auckland, New Zealand, May 9–13, 2020,* IFAAMAS, 3 pages.

1 INTRODUCTION

Agent-Based Simulation (ABS) can be used for study of complex systems of interacting agents. The purpose of the simulation can be, e.g., prediction, verification, training and analysis [4]. An important component of the simulation is the behavior models of the agents. Constructing realistic behavior models is challenging and time-consuming [6, 31], especially for subject matter experts, who may not have expertise in artificial intelligence. In this work, we investigate how machine learning can be used to construct these models, and to adapt the contents of agent-based simulation to end-user needs. As a case study we use a simulation-based air combat training system. Fighter aircraft are becoming increasingly complex, and there is a growing need for efficient and effective pilot training solutions. By using simulations to a greater degree, higher training value can be achieved at lower cost [19]. Ideally, human participants in a training session would all be receiving training, i.e., we would like to minimize the dependence on human support personnel. For instance, synthetic agents could replace human role-players and real aircraft in training sessions. This would improve the availability of training and make it possible to realize more complex training scenarios.

Proc. of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2020), B. An, N. Yorke-Smith, A. El Fallah Seghrouchni, G. Sukthankar (eds.), May 9–13, 2020, Auckland, New Zealand.

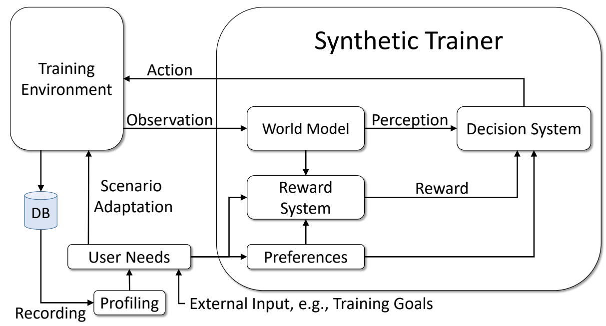

Figure 1: Architecture of an adaptive training system [10].

2 PROPOSED APPROACH

Our approach for constructing an adaptive training system, with a high level of autonomy, is illustrated by the system architecture in Figure 1. For our case study, a Synthetic Trainer should be able to act as ally or adversary in air combat training scenarios. While doing so, this agent should consider objectives related to the simulated scenario, as well as the learning objectives of human trainees. During training sessions the agent forms a high-level perception of its environment, a World Model, based on observations through its low-level sensors. The objectives of the agent and their relative importance are represented by a Reward System and a set of Preferences. A Decision System, with learning and planning capabilities, is used to select actions based on the current perception and prioritized objectives. A Profiling function is used to determine User Training Needs, based on trainees' historical performance. User needs may also be partially supplied by external input, for instance, training goals provided by a human instructor. Based on the inferred training needs, scenario contents and synthetic agent characteristics are then adapted to provide effective training.

To move towards the identified goals, we intend to first study basic principles in simple simulation environments, e.g., gridworlds. Promising concepts will then be developed further and evaluated in high-fidelity simulations. Finally, human-agent interaction will be studied in the target system. We will try to answer the following overarching research questions:

- **RQ1**: How can agents learn to act as synthetic trainers for human trainees?
- **RQ2**: How can simulation contents be automatically adapted to fit the training needs of an individual trainee?
- **RQ3**: What is required for human and synthetic agents to interact effectively in a simulation-based training environment?

3 PRELIMINARY RESULTS

Our case study of ABS for air combat training provides several challenges for learning agents (**RQ1**):

- Many interacting human and synthetic agents in mixed cooperative and competitive scenarios
- Multiple conflicting objectives that must be considered by teams of agents, e.g., tactical mission goals, resource consumption and safety, as well as the learning objectives of human trainees
- Partial observability of the environment due to limitations in sensors and data links, as well as effects of electronic warfare
- Decision-making over long time horizons, corresponding to hundreds or thousands of time steps
- High-fidelity simulation models that are computationally heavy, resulting in long simulation times during training of synthetic agents
- Agent behavior needs to be explainable to humans, so that debriefing of training sessions can be performed effectively

In the early stages of the project we have identified techniques that could be useful to address these challenges [11]. For the Decision System in Figure 1 we intend to study multi-objective multi-agent learning and planning [16, 20]. To tackle the complexity of the application domain, and to achieve efficient learning, we will combine these techniques with, e.g., reward shaping [5, 15], curriculum learning [2, 7], and hierarchical learning [1, 23, 32]. Multi-objective methods will allow us to adjust the agents' priorities among objectives at runtime, making it possible to adapt agent characteristics to trainees' needs, which may vary among training sessions.

In an initial concept study we have investigated how multi-objective deep reinforcement learning could be used to build agent-based simulations with tunable dynamics [12]. As illustrated in Figure 1, we created an agent with a reward system affected by a set of preferences, which specify the relative importance of the agent's objectives. We then conditioned the agent's policy on these preferences, so that the agent's behavior could be adjusted at runtime. We evaluated this approach in gridworld environments, and showed that the competitiveness (in a Gathering Environment) and risk-taking (in a Traffic Environment) of the agent could be significantly affected after training. Such properties are of interest when designing air combat training scenarios.

To better understand the limitations of current state-of-the-art algorithms, we have also evaluated the performance of multi-agent and multi-objective reinforcement learning in the target system [10]. We studied scenarios that required agents to coordinate their actions to efficiently solve tasks, and to take risk into account when selecting actions. We noted that seemingly simple scenarios can still be difficult to tackle, and that the geographical extension of the scenario, in combination with the level of abstraction of the chosen action space design, had a significant impact on the performance. The long sequences of actions that are typically required to solve tasks in the air combat domain makes it challenging for the agents to explore and find efficient tactics.

In recent work, we have conducted interviews with experienced pilots to identify important aspects of air combat training (**RQ3**) [9]. We have used this information to define scenarios that will be used for future development and evaluation of learning agents.

4 RELATED WORK

Learning approaches for agent-based modelling have great potential, and have been investigated within many application domains. Some approaches that have been studied for building behavior models for air combat simulation are evolutionary algorithms [3, 14, 35], neural networks [8, 18, 24, 25] and dynamic scripting [28–30]. These attempts have produced some interesting results, but the techniques have not been used much in operational systems [27], due to limitations in performance. Recent advances in machine learning for game playing agents, e.g., AlphaGo [21, 22], has sparked interest in using deep reinforcement learning [13, 17, 26, 34]. However, the scenarios studied are still quite simple, with only a few interacting agents. We intend to study more complex scenarios, that more closely resemble those used in actual training.

Recently, it has become possible to train agents to reach human-level performance in real-time strategy games, such as StarCraft II [33]. These games have some elements in common with air combat scenarios, e.g., multiple competing agents, partial observability, and decision-making over long time horizons. However, the approaches currently used require massive computation resources, which a typical training facility can not be expected to have access to. Improving the sample efficiency of algorithms is an important direction for future research.

5 FUTURE WORK

Based on our initial results, we are currently investigating ways of improving algorithms for learning and planning in our application domain. Some directions that we are interested in pursuing are:

- Improving the efficiency of algorithms in multi-objective multi-agent scenarios through, e.g., reward shaping, curriculum learning, and hierarchical learning (**RQ1**)
- Learning from demonstration as a way of user preference elicitation (**RQ2**)
- Agent modelling to support the decision-making of synthetic agents (World Model in Figure 1; **RQ1**), and for inferring human users' training needs (Profiling in Figure 1; **RQ2**)

We note that there is an interesting overlap in performance evaluation and generation of training curricula for human and synthetic agents respectively, which we would like to explore in future work (Profiling, Reward System, and Scenario Adaptation in Figure 1; **RQ1** and **RQ2**).

We are also currently in the process of planning for evaluations of our initial findings in experiments with manned simulators. The intention is to train synthetic agents in 2-vs-2 air combat scenarios, using multi-objective multi-agent reinforcement learning, and then conduct experiments where some of the synthetic agents are replaced by human pilots. With these experiments we would like to find out how humans can cooperate with the synthetic agents, and how robust their behavior is (**RQ3**).

ACKNOWLEDGMENTS

This work was partially supported by the Swedish Governmental Agency for Innovation Systems (NFFP7/2017-04885), and the Wallenberg Artificial Intelligence, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation.

REFERENCES

[1] Pierre-Luc Bacon, Jean Harb, and Doina Precup. 2017. The option-critic architecture. In *Thirty-First AAAI Conference on Artificial Intelligence*.

[2] Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. 2009. Curriculum learning. In *Proceedings of the 26th annual international conference on machine learning*. 41–48.

[3] Magdalena D Bugajska, Alan C Schultz, J Gregory Trafton, Shaun Gittens, and Farilee Mintz. 2001. *Building Adaptive Computer-Generated Forces: The Effect of Increasing Task Reactivity on Human and Machine Control Abilities.* Technical Report. NAVAL RESEARCH LAB WASHINGTON DC CENTER FOR APPLIED RESEARCH IN ARTIFICIAL INTELLIGENCE.

[4] Paul Davidsson, Johan Holmgren, Hans Kyhlbäck, Dawit Mengistu, and Marie Persson. 2006. Applications of agent based simulation. In *International Workshop on Multi-Agent Systems and Agent-Based Simulation*. Springer, 15–27.

[5] Sam Devlin and Daniel Kudenko. 2011. Theoretical considerations of potential-based reward shaping for multi-agent systems. In *The 10th International Conference on Autonomous Agents and Multiagent Systems-Volume 1*. International Foundation for Autonomous Agents and Multiagent Systems, 225–232.

[6] Nigel Gilbert. 2019. *Agent-based models*. Vol. 153. Sage Publications, Incorporated.

[7] Jayesh K Gupta, Maxim Egorov, and Mykel Kochenderfer. 2017. Cooperative multi-agent control using deep reinforcement learning. In *International Conference on Autonomous Agents and Multiagent Systems*. Springer, 66–83.

[8] Amy E Henninger, Avelino J Gonzalez, Michael Georgiopoulos, and Ronald F DeMara. 2000. Modeling Semi-automated forces with neural networks: Performance improvement through a modular approach. In *The Ninth Conference on Computer Generated Forces and Behavioral Representation Proceedings*.

[9] Johan Källström, Rego Granlund, and Fredrik Heintz. 2020. Design of Simulation-Based Pilot Training Systems using Machine Learning Agents. *Accepted for presentation at ICAS 2020* (2020).

[10] Johan Källström and Fredrik Heintz. 2019. Multi-Agent Multi-Objective Deep Reinforcement Learning for Efficient and Effective Pilot Training. *FT2019* (2019).

[11] Johan Källström and Fredrik Heintz. 2019. Reinforcement Learning for Computer Generated Forces using Open-Source Software. In *Interservice/Industry Training, Simulation, and Education Conference, December 2-6, 2019, Orlando, USA*. 1–11.

[12] Johan Källström and Fredrik Heintz. 2019. Tunable dynamics in agent-based simulation using multi-objective reinforcement learning. In *Adaptive and Learning Agents Workshop (ALA-19) at AAMAS, Montreal, Canada, May 13-14, 2019*. 1–7.

[13] Budi Kurniawan, Peter Vamplew, Michael Papasimeon, Richard Dazeley, and Cameron Foale. 2019. An Empirical Study of Reward Structures for Actor-Critic Reinforcement Learning in Air Combat Manoeuvring Simulation. In *Australasian Joint Conference on Artificial Intelligence*. Springer, 54–65.

[14] Sandeep Mulgund, Karen Harper, Kalmanje Krishnakumar, and Greg Zacharias. 1998. Air combat tactics optimization using stochastic genetic algorithms. In *SMC'98 Conference Proceedings. 1998 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No. 98CH36218)*, Vol. 4. IEEE, 3136–3141.

[15] Andrew Y Ng, Daishi Harada, and Stuart Russell. 1999. Policy invariance under reward transformations: Theory and application to reward shaping. In *ICML*, Vol. 99. 278–287.

[16] Roxana Rădulescu, Patrick Mannion, Diederik M Roijers, and Ann Nowé. 2020. Multi-objective multi-agent decision making: a utility-based analysis and survey. *Autonomous Agents and Multi-Agent Systems* 34, 1 (2020), 10.

[17] Roel Rijken and Armon Toubman. 2016. The future of autonomous air combat behavior. In *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 3089–3094.

[18] Ervin Y Rodin and S Massoud Amin. 1992. Maneuver prediction in air combat via artificial neural networks. *Computers & mathematics with applications* 24, 3 (1992), 95–112.

[19] JJ Roessingh and GG Verhaaf. 2009. *Training Effectiveness of Embedded Training in a (Multi-) Fighter Environment.* Technical Report. NATIONAL AEROSPACE LAB AMSTERDAM (NETHERLANDS).

[20] Diederik M Roijers, Peter Vamplew, Shimon Whiteson, and Richard Dazeley. 2013. A survey of multi-objective sequential decision-making. *Journal of Artificial Intelligence Research* 48 (2013), 67–113.

[21] David Silver, Aja Huang, Chris J Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, et al. 2016. Mastering the game of Go with deep neural networks and tree search. *nature* 529, 7587 (2016), 484.

[22] David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, et al. 2017. Mastering the game of go without human knowledge. *Nature* 550, 7676 (2017), 354.

[23] Richard S Sutton, Doina Precup, and Satinder Singh. 1999. Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning. *Artificial intelligence* 112, 1-2 (1999), 181–211.

[24] Teck-Hou Teng, Ah-Hwee Tan, Yuan-Sin Tan, and Adrian Yeo. 2012. Self-organizing neural networks for learning air combat maneuvers. In *The 2012 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 1–8.

[25] Teck-Hou Teng, Ah-Hwee Tan, and Loo-Nin Teow. 2013. Adaptive computer-generated forces for simulator-based training. *Expert Systems with Applications* 40, 18 (2013), 7341–7353.

[26] Babak Toghiani-Rizi, Farzad Kamrani, Linus J Luotsinen, and Linus Gisslén. 2017. Evaluating deep reinforcement learning for computer generated forces in ground combat simulation. In *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 3433–3438.

[27] Armon Toubman, Gerald Poppinga, Jan Joris Roessingh, Ming Hou, Linus Luotsinen, Rikke Amilde Løvlid, Christophe Meyer, Roel Rijken, and M Turcanık. 2015. Modeling CGF behavior with machine learning techniques: Requirements and future directions. In *Proceedings of the 2015 Interservice/Industry Training, Simulation, and Education Conference*. 2637–2647.

[28] Armon Toubman, Jan Joris Roessingh, Pieter Spronck, Aske Plaat, and Jaap Van Den Herik. 2014. Dynamic scripting with team coordination in air combat simulation. In *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*. Springer, 440–449.

[29] Armon Toubman, Jan Joris Roessingh, Pieter Spronck, Aske Plaat, and Jaap van den Herik. 2015. Transfer learning of air combat behavior. In *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 226–231.

[30] Armon Toubman, Jan Joris Roessingh, Pieter Spronck, Aske Plaat, and Jaap van den Herik. 2016. Rapid adaptation of air combat behaviour. In *Proceedings of the Twenty-second European Conference on Artificial Intelligence*. IOS Press, 1791–1796.

[31] TW van den Berg, NM de Reus, and JM Voogd. 2011. *LVC Architecture study*. Simulation Interoperability Standards Organization (SISO).

[32] Alexander Sasha Vezhnevets, Simon Osindero, Tom Schaul, Nicolas Heess, Max Jaderberg, David Silver, and Koray Kavukcuoglu. 2017. Feudal networks for hierarchical reinforcement learning. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*. JMLR. org, 3540–3549.

[33] Oriol Vinyals, Igor Babuschkin, Wojciech M Czarnecki, Michaël Mathieu, Andrew Dudzik, Junyoung Chung, David H Choi, Richard Powell, Timo Ewalds, Petko Georgiev, et al. 2019. Grandmaster level in StarCraft II using multi-agent reinforcement learning. *Nature* 575, 7782 (2019), 350–354.

[34] Bogdan Vlahov, Eric Squires, Laura Strickland, and Charles Pippin. 2018. On Developing a UAV Pursuit-Evasion Policy Using Reinforcement Learning. In *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 859–864.

[35] Jian Yao, Qiwang Huang, and Weiping Wang. 2015. Adaptive human behavior modeling for air combat simulation. In *2015 IEEE/ACM 19th International Symposium on Distributed Simulation and Real Time Applications (DS-RT)*. IEEE, 100–103.