Behaviour Analysis of Mixed Game-Theoretic Learning Algorithms

(Extended Abstract)

Michalis Smyrnakis
Department of Automatic Control and Systems Engineering
University of Sheffield
Sheffield, United Kingdom
m.smyrnakis@sheffield.ac.uk

Hongyang Qu
Department of Automatic Control and Systems Engineering
University of Sheffield
Sheffield, United Kingdom
h.qu@sheffield.ac.uk

Sandor M. Veres
Department of Automatic Control and Systems Engineering
University of Sheffield
Sheffield, United Kingdom
s.veres@sheffield.ac.uk

ABSTRACT

Distributed optimisation becomes ever more important to boost the operational efficiency of autonomous systems and a number of decision algorithms have been proposed in recent years. A common assumption is usually made that individual agents use the same type of learning algorithm. There are however applications, such as reconfigurable robotics and coordination within robot teams, where this assumption is not always valid. In this paper we propose a methodology that allows the study of agents' joint behaviour when they use different game-theoretic learning algorithms. Our methodology is based on probabilistic model checking, and we use a new a behaviour-similarity-relation to build compact state spaces. Our theory and computational procedures for formal verification provide a framework to study the properties of various algorithms. The proposed methodology is demonstrated on four learning algorithms that are used to provide decisions in distributed optimisation tasks formulated as multi-player games.

Categories and Subject Descriptors

I.2.11 [**Distributed Artificial Intelligence**]: Multiagent systems

Keywords

Multi-agent systems, verification, game-theoretic learning, learning agents

1. INTRODUCTION

Distributed optimisation is a crucial part of many multi-agent systems applications, such as robotics [8] and wireless sensor networks [5], where agents need to coordinate to achieve a common goal. Many learning algorithms have been proposed as coordination mechanism between agents in order to reach consensus and accomplish their joint task.

A common assumption in the analysis of learning algorithms is that all agents use the same learning algorithms. In this work we consider cases where each agent can use a different algorithm to coordinate in order to achieve their common goal. This can occur when the agents of a multi-agent system have different capabilities and also when a part of the multi-agent system faces computational or communication restrictions. Thus, it is reasonable to assume that each agent is able to use algorithms that are only suitable to them and that a single learning algorithm might not be appropriate for every agent.

We focus on distributed optimisation problems which can be cast as games. To analyse the behaviour of agents when different learning algorithms are used, we propose a methodology that is based on probabilistic model checking for discrete-time Markov chains (DTMCs). Our methodology examines all possible joint actions of the agents from all states of the world during a finite number of iterations of game playing. The alternative approach of simulations would suffer from a large number of runs and there would be no rule to specify the number of runs that were sufficient to reveal all possible joint behaviours of the agents.

2. GAMES & LEARNING ALGORITHMS

A game has a set of players where each player (here also called 'agent') has a set of actions and gains a reward for each of its actions. A Nash equilibrium [6] is a joint strategy of actions among all agents in a way that no agent can increase its reward by unilaterally changing its own strategy. A Pareto efficient equilibrium is a Nash equilibrium where all agents' rewards are maximised. In this paper, we study four iterative game-theoretic learning algorithms namely Fictitious play (FP) [1], Geometric fictitious play (GFP) [4], Regret matching (RM) [2] and Spatial adaptive play (SAP)[2]. Players that adopt either FP or GFP take into account other players' actions in all previous iterations when choosing actions in order to maximise their expected rewards. RM and SAP are myopic algorithms that use only the actions of other players in the previous iteration of the game to select actions that minimise the regret that they will have after choosing a specific action.

3. ANALYSIS METHODOLOGY

Our methodology was inspired by probabilistic model checking, in particular, verification of DTMCs [3]. It explores every possible execution trace of a system to verify the prob-

Appears in: *Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2015), Bordini, Elkind, Weiss, Yolum (eds.), May 4–8, 2015, Istanbul, Turkey.*

abilistic behavioural properties the system has. A model of a game, under a combination of learning algorithms of the agents, is composed of a set of states and a set of transitions between states. A state contains all the information that an agent's algorithm needs to determine its next action. A transition is then the result of the joint actions of the agents and each transition is associated with a probability determined by the agents' algorithms, expressing the likelihood of choosing the corresponding joint action in the next iteration.

One of the key issues in the development of this framework is how to restrain the size of a model to keep computational complexity acceptable without losing accuracy of coverage. In principle, such a model can be unbounded because a game can be played infinitely long, even if a Nash equilibrium is reached. To avoid this, we propose a *behaviour similarity relation* between states. If two states are behaviour similar, then the agents will demonstrate similar behaviour from these two states, hence the name of the similarity relation. Thus, there is no need to explore both states. Expanding one of them is sufficient. Using this relation, a compact model can be obtained in which a Nash equilibrium is captured by a bottom-strongly- connect-component (BSCC) in the model. Consequently, the probabilistic model checking technique for computing reachability probabilities on DTMCs [3] can be adopted to find out the probability of reaching a given Nash equilibrium.

4. EXPERIMENTS

We demonstrate the outcome of the proposed methodology when the FP, GFP, RM and SAP are combined to represent a game of cooperation as distributed optimisation. We use a symmetric game, prisoners' dilemma, hawk and dove game, a coordination game, and a leader-follower game as case studies. The definition of these games can be found in [7]. The probability of convergence to the Pareto efficient equilibrium is used as a performance measure between the various combinations of the algorithms. For the Hawk and Dove game, we measure the probability that the agents choose one of the two non Pareto efficient equilibria because no Pareto efficient equilibria exist. We report the average results of 100 replications for each combination with random initial conditions.

Figure 1 is a stack diagram that depicts the performance of each combination when the probability to converge to (non)Pareto efficient equilibria is concerned as a comparison measure. In particular, each bar in the figure represents one combination of a learning algorithms, and is divided into 5 sections as the number of the games that we used (each with different reward structure). The height of each section illustrates the probability with which an algorithm can reach (non)Pareto efficient equilibria in the corresponding game. Hence the maximum performance, i.e., height of a bar for a combination can be at most 5. When GFP is combined with RM, we can observe the best overall performance between all the possible combinations. This apporach can ultimately indicate that combinations of different algorithms, in some specific distributed optimisation tasks, may provide better solutions than using the same algorithm by every agent.

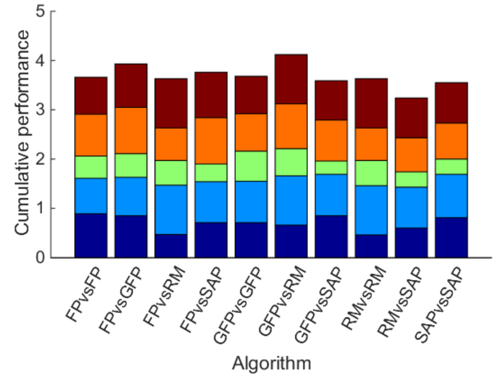

Figure 1: Performance of combinations of various game-theoretic learning algorithms.

5. CONCLUSIONS

In this paper we propose a verification-based methodology to study the behaviour of agents when they solve distributed optimisation tasks representable as game-theoretic learning algorithms. We use well studied learning algorithms to validate the outcome of our methodology. The experiments studied here show that a combination of two algorithms, GFP and RM, have the best overall performance even when compared with cases where all agents use the same algorithm. This indicates that it is not always desirable to force all agents to adopt a single learning algorithm. Future research may examine more learning algorithms and more games using our methodology and can apply the results to real-world scenarios.

REFERENCES

[1] G. W. Brown. Iterative solutions of games by fictitious play. In T. C. Koopmans, editor, *Activity Analysis of Production and Allocation*, pages 374–376. Wiley, 1951.

[2] A. C. Chapman, A. Rogers, and N. R. Jennings. Benchmarking hybrid algorithms for distributed constraint optimisation games. *Autonomous Agents and Multi-Agent Systems*, 22(3):385–414, 2011.

[3] C. Courcoubetis and M. Yannakakis. Verifying temporal properties of finite state probabilistic programs. In *Proc. FOCS'88*, pages 338–345. IEEE Computer Society Press, 1988.

[4] D. Fudenberg and D. Levine. *The theory of Learning in Games*. The MIT Press, 1998.

[5] J. Kho, A. Rogers, and N. R. Jennings. Decentralized control of adaptive sampling in wireless sensor networks. *ACM Transactions on Sensor Networks (TOSN)*, 5(3):19:1–19:35, 2009.

[6] J. Nash. Equilibrium points in n-person games. In *Proc. the National Academy of Science, USA*, volume 36, pages 48–49, 1950.

[7] E. Rasmusen. *Games and Information: An Introduction to Game Theory*. Wiley, 2006.

[8] M. Smyrnakis and S. Veres. Coordination of control in robot teams using game-theoretic learning. In *Proceedings of IFAC'14*, pages 1194–1202, 2014.