Applied Soft Computing 18 (2014) 12–24

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

A robust approach for multi-agent natural resource allocation based on stochastic optimization algorithms

N. Barbalios *, P. Tzionas

Intelligent Control Lab, Department of Automation Engineering, Alexander Technological Educational Institute of Thessaloniki, GR 54101, Greece

ARTICLE INFO

Article history:
Received 26 October 2012
Accepted 4 January 2014
Available online 24 January 2014

Keywords:
Multi-agent learning
Natural resource allocation
Stochastic optimization
Evolutionary methods
Genetic algorithms
Covariance Matrix Adaptation

ABSTRACT

Natural resource allocation is a complex problem that entails difficulties related to the nature of real world problems and to the constraints related to the socio-economical aspects of the problem. In more detail, as the resource becomes scarce relations of trust or communication channels that may exist between the users of a resource become unreliable and should be ignored. In this sense, it is argued that in multi-agent natural resource allocation settings agents are not considered to observe or communicate with each other. The aim of this paper is to study multi-agent learning within this constrained framework. Two novel learning methods are introduced that operate in conjunction with any decentralized multi-agent learning algorithm to provide efficient resource allocations. The proposed methods were applied on a multi-agent simulation model that replicates a natural resource allocation procedure, and extensive experiments were conducted using popular decentralized multi-agent learning algorithms. Experimental results employed statistical figures of merit for assessing the performance of the algorithms with respect to the preservation of the resource and to the utilities of the users. It was revealed that the proposed learning methods improved the performance of all policies under study and provided allocation schemes that both preserved the resource and ensured the survival of the agents, simultaneously. It is thus demonstrated that the proposed learning methods are a substantial improvement, when compared to the direct application of typical learning algorithms to natural resource sharing, and are a viable means of achieving efficient resource allocations.

1. Introduction

Multi-agent learning is an area of study that deals with the theoretical challenge of understanding how agents learn and adapt their behaviour in the presence of other agents that are also learning. It has drawn much attention over the past decades, as the expansion of multi-agent modelling approaches across interdisciplinary domains (i.e. economics, robotics and others) created a need for robust multi-agent learning algorithms. Learning in multi-agent settings however is a hard task, since the actions of one agent affects the environment of the other and vice versa [29]. In such dynamic environments, the convergence properties of single agent learning algorithms do not necessarily hold [29,54], thus learning becomes difficult even for simple two-agent settings [10,44,29].

Many approaches have been proposed to deal with multi-agent learning, with Fictitious Play, Bayesian Learning, No-Regret Learning, Targeted Learning, Evolutionary Learning being the most representative ones [42,27]. Although these approaches have been successfully applied to game theory and converge to near optimal policies for specific types of games, most of them assume a priori knowledge of transition probabilities from one stage of the game to the other [42,43,7]. A popular approach that does not make such assumptions is *Reinforcement Learning*, where agents gradually acquire knowledge through scalar rewards that result from the actions they perform [45].

The most typical algorithm of reinforcement learning is *Q*-learning [49], that iteratively updates a utility-estimation function (i.e. *Q*-function) for state-action pairs and associates each action with a value (i.e. *Q*-value). Given that all actions have been sufficiently explored, *Q* learning is known to converge to optimal solutions for single-agent settings [27,7]. Early approaches of multi-agent reinforcement learning entail the generalization of *Q*-learning on multi-agent environments, called Decentralized-*Q* learning, where each agent ignores all the other agents. This approach however fails in many cases, as it ignores the nature of the multi-agent settings [43,42]. To that extent, variants of *Q*-learning have been proposed to address multi-agent learning.

One of the first extensions of *Q*-learning towards multi-agent applications was presented in [31], where the *Q*-function was extended to encapsulate the value of joint actions. The same concept was adopted in [24], where each agent observes the other

* Corresponding author. Tel.: +1 (412) 273 0846.
E-mail addresses: nbarba@autom.teithe.gr, nbarmpalios@gmail.com (N. Barbalios), ptzionas@autom.teithe.gr (P. Tzionas).

1568-4946/$ – see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.asoc.2014.01.004

agents and keeps track of their Q-values [7]. Observation or communication is used in most multi-agent reinforcement learning (MARL) algorithms in order to assist learning of joint actions [4,46,31,24,32,14,35,25,2]. In more detail, in [4] agents select their actions taking under consideration the opinion of other agents, whereas in [32,14] Q-values are updated using a value function, estimated through observations. In [25] an agent's 'attraction' to a specific strategy is updated according to a mixture of reinforcement learning and belief learning techniques, using inter-agent signalling. Similarly, in [35] the pheromone concept of swarm intelligence is used to allow agents to communicate with each other and to jointly learn how to solve a problem. In the same context, agents communicate in order to adapt their actions using Bayesian learning in [1], while [53] uses a bargain scheme where an agent 'announces' his action to other agents.

Several MARL algorithms exist that bias the Q-values by modifying them (or purposefully failing to update them) according to some predefined criteria. In [30], agents ignore penalties related to non-coordination when updating their Q-function, thus evaluating actions using the maximum reward received. A similar approach is used in [34], where two distinct learning rates for updating the Q-function are used, depending on the reward the agent receives, whereas in [38] Q-values are updated only if an action has been accessed for several times. In [26], the probability of choosing an action is biased by modifying the Q-function according to a heuristic value, that is estimated based on how frequently the maximum reward for that action was received.

Although multi-agent learning algorithms provide near optimal solutions for several types of games (i.e. zero sum or common pay-off games [43]), they fail when they are scaled in real world problems [33]. This is because the complexity of learning is dramatically increased due to a variety of reasons related not only to the non-stationarity of the environment, but to the nature of the problem under study as well. In more detail, even though reinforcement learning algorithms are well defined and extensively studied for discrete action and state spaces, real world applications entail variables of continuous nature [45]. Although this problem can be alleviated using function approximation (e.g. [50]), there is no proof of convergence to optimal solutions for continuous state/action space problems [16]. Additionally, due to the stochasticity and the non-deterministic nature of real world phenomena, measurement uncertainties and noise are almost certain to occur if the problem entails a large number of agents or ambiguous signals [29,33]. Moreover, many real world problems also face the problem of aliasing, i.e. different environmental states that are similar and cannot be easily distinguished. Learning is even made more difficult due to physical or social constraints that may restrict the information available to the agents. For example, unlike game theory where agents are expected to have access to other agent's utilities or to an overall utility welfare, in real world scenarios this is hardly the case [33,29]. In natural resource sharing for example, as the resource becomes scarce relations of trust or communication channels that may exist between the users of the resource become unreliable and should be ignored [3]. In addition, the importance of an equilibrium solution (e.g. Nash) requires a different justification for real world applications, where it should be investigated if indeed a Nash equilibrium reflects a desirable solution to the problem under study [33]. For example, it may be impossible to describe a desirable solution by such a unique equilibrium and multiple equilibria may exist.

A typical approach that has been successfully applied to alleviate these issues and scale multi-agent learning beyond illustrative academic problems is the enhancement of learning algorithms with domain knowledge about the specific problem under study [33]. Recently, we have proposed a multi-agent simulation model that deals with natural resource allocation, together with a simple multi-agent learning approach that encapsulates domain knowledge in the form of empirical rules [3]. In this paper, we provide a more elaborate formulation of the learning approach and extend it using stochastic derivative-free optimization methods. The proposed approach, called *Adaptive Rectification – AR* employs a parametrized rectification function that is trained using a problem specific metric and reveals the way agents should modify their actions in order to achieve efficient resource allocation. In contrast to [3], domain knowledge is not hard coded in terms of heuristic rules but are rather revealed using a supervised learning scheme. Two *AR* methods are introduced in this study, *AR-RCGA* and *AR-CMA* that employ different methods for the estimation of the parameters of the rectification function, namely the *Covariance Matrix Adaptation (CMA)* evolutionary strategy and *Real Coded Genetic Algorithms (RCGA)*. The main contributions of the proposed methods are that: (a) they achieve the coordination of the actions of the agents towards the solution of the problem using exclusively local information, without employing inter-agent communication or observation. This is an advancement compared to most typical multi-agent learning algorithms and multi-agent resource allocation solutions, that rely mostly on inter-agent communication or observation (i.e. combinatorial auctions); (b) they can be used to scale any general purpose multi-agent learning algorithm that does not violate the imposed constraints (i.e. lack of inter-agent communication and observation), to the real world problem of natural resource allocation. This is achieved through the parametrized function that embeds domain knowledge about the problem under study.

The proposed methods were applied to the multi-agent simulation (MAS) model presented in [3], that accurately simulates a real-world natural resource allocation phenomena (i.e. a water resource exploited by a community of users). Extensive experiments were conducted following a Monte Carlo procedure, that evaluated the robustness of the proposed methods in providing efficient resource allocation schemes for agent populations with varying water needs. For the decision making process of the agents, agent policies employing popular decentralized multi-agent learning algorithms were implemented, whereas experiments entailed simulations with and without the use of the proposed *AR* methods, in order to demonstrate the complexity of learning and the superiority of the proposed methods. The impact of the policies under study on the sustainability of the resource and on the survival of the users was evaluated using qualitative and quantitative figures of merit. In addition, the occurrence probabilities for characteristic events such as the depletion of the resource and the survival of the users were estimated, in order to evaluate the overall performance of the policies and investigate the robustness of the proposed methods. As it will be demonstrated, policies entailing decentralized learning algorithms fail to ensure the simultaneous survival of the agents and preservation of the resource. However, the use of the proposed methods proves to be adequate for producing efficient resource allocations schemes under any given scenario (i.e. for any water need of the population), despite the imposed constraints (i.e. lack of communication/observation). Thus the proposed methods provide a substantial improvement, when compared to the direct application of typical learning algorithms to natural resource sharing, and are a viable means of simultaneously preserving the resource and ensuring the survival of the agents.

2. Problem description and challenges

Multi-agent resource allocation (MARA) is a problem met across various interdisciplinary domains, and refers to the process of distributing a resource to a number of agents [11]. In this study