Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 00 (2010) 000–000

Procedia Computer Science

www.elsevier.com/locate/procedia

ICCS 2010 – Workshop on Cognitive Agents: Theory and Practice

A Unified Framework for Reinforcement Learning, Co-Learning and Meta-Learning How to Coordinate in Collaborative Multi-Agent Systems

Predrag T. Tošić and Ricardo Vilalta

ptosic@uh.edu, vilalta@cs.uh.edu

Department of Computer Science, University of Houston,
Houston, Texas, USA

Abstract

Coordination among multiple autonomous, distributed cognitive agents is one of the most challenging and ubiquitous problems in Distributed AI and its applications in general, and in collaborative multi-agent systems in particular. A particularly prominent problem in multi-agent coordination is that of group, team or coalition formation. A considerable majority of the approaches to this problem found in the literature assume fixed interactions among autonomous agents involved in the coalition formation process. Moreover, most of the prior research where agents are actually able to learn and adapt based on their past interactions mainly focuses on *reinforcement learning* techniques at the individual agent level. We argue that, in many important applications and contexts, complex large-scale collaborative multi-agent systems need to be able to learn and adapt at multiple organization, hierarchical and logical levels. In particular, the agents need to be able to learn both at the level of individual agents and at the system or agent ensemble levels, and then to integrate these different sources of learned knowledge and behavior, in order to be effective at solving complex tasks in typical dynamic, partially observable and noisy multi-agent environments. In this paper, we describe a conceptual framework for addressing the problem of learning how to coordinate effectively at three qualitatively distinct levels – those of (i) individual agents, (ii) small groups of agents, and (iii) very large agent ensembles (or alternatively, depending on the nature of a multi-agent system, at the system or central control level). We briefly illustrate the applicability and usefulness of the proposed conceptual framework with an example of how it would apply to an important practical coordination problem, namely that of distributed coordination of a large ensemble of unmanned vehicles on a complex multi-task mission.

Keywords: autonomous agents; multi-agent systems; distributed problem solving; agent coordination; coalition formation; distributed task allocation; reinforcement learning; co-learning; meta-learning; multi-level learning

1. Introduction and Motivation

Multi-agent coordination is one of the central problems in distributed AI research. According to an authoritative source [21], the coordination problem can be defined as managing inter-dependencies among the activities of different agents. These interdependencies of agent activities can be of various types, as discussed in [21] (pp. 200 – 202), and the type of interdependencies is one of the factors determining the appropriate multi-agent coordination paradigm. Moreover, various interdependencies and hence the need to coordinate may arise both among self-interested, competitive agents, and the cooperative, distributed problem solving agents who typically share the same goals or objectives and are not competing with each other.

There are several major varieties of coordination problems. Among them, distributed consensus problems are particularly pervasive and prominent. An important distributed consensus problem is *coalition formation* (e.g., [8, 9, 10, 11, 12, 16, 17]). In this paper, we will mainly focus on the coordination distributed consensus problem coalition formation among multiple collaborative autonomous agents that are *cooperating* with each other in order to better perform on their tasks [11, 12, 16, 17]. While the problem setting we consider is admittedly somewhat simpler than the one where the autonomous agents are *self-interested* and *competitive* (and hence, in general, may engage in both competition and cooperation depending on the situation), as we shall see this setting still provides an abundance of research challenges – especially when it comes to addressing the appropriate ways of enabling these cooperative agents to learn how to coordinate more effectively based on their past interactions.

Distributed coalition formation is a classical coordination problem that has been extensively studied in the Distributed AI literature. In many important collaborative *multi-agent system* (MAS) applications, autonomous agents need to dynamically form groups or coalitions in an efficient, reliable, fault-tolerant and partly or entirely distributed manner. In most of the literature on distributed coalition formation, agents' negotiation and coalition formation strategies are *static*, i.e., the agents do not learn from past experience nor adapt their strategies to become more effective in their future coalition formation interactions. Moreover, most of the prior research that does consider learning and adaptation to improve coordination in general and the effectiveness of coalition formation in particular, mainly or exclusively focuses on various models of *reinforcement learning* (RL) at the level of *individual agents.* As far as we know, co-learning at the level of small groups of agents has, thus far, received some (limited) attention in the research community (e.g., [14]), whereas the system-level meta-learning in the context of improving multi-agent coordination has been hardly addressed at all.

We argue that a more promising and more systematic way to address agents' adaptability and ability to improve at forming coalitions in typical complex, noisy and dynamic environments, is to combine RL of individual agents with *co-learning* at the level of small groups of agents (such as pairs or triples), as well as *meta-learning* at higher organizational levels (such as large agent ensembles, the entire MAS, or the MAS designer). We further argue that the interaction and synergy between reinforcement learning of individual agents, co-learning of small groups of agents and meta-learning at the large ensembles' and/or system level would enable agents to considerably improve in their coordination capabilities, even when there's only a modest amount of past experience, limited computational resources of each individual agent (hence imposing constraints on how much can an agent spend on RL), and where there are possibly considerable changes in the agents' dynamic environments in which subsequent coalition formations take place. Needless to say, the more experience (i.e., recorded history) about the past multi-agent interactions is stored and then the subsequent inferences based upon, the more effectiveness we can expect from the multi-level learning to coordinate approach, and especially its meta-learning component (as typically the most computationally and data intensive of all).

In the sections that follow, we will elaborate on our conceptual framework on learning how to coordinate effectively at the three different qualitative levels outlined above, and illustrate how this multi-level, synergetic learning would improve coordination and coalition formation capabilities of a particular collaborative MAS application that has been of a considerable interest among the MAS applications community.

2. Collaborative Multi-Agent Coalition Formation: Existing Approaches and Some Challenges

Distributed coalition formation in multi-agent domains is an important coordination problem extensively studied by the *Multi-Agent Systems* (MAS) research community, especially among those interested in *collaborative* MAS (e.g., [8, 9, 10, 12, 16, 18]). There are many important collaborative MAS applications where autonomous agents need to form groups, teams or coalitions, in many instances repeatedly. The agents may need to form teams or coalitions in order to share resources, complete tasks that exceed the abilities of individual agents, and/or improve some system-wide performance metric such as the speed of task completion (see, e.g., [12, 7]).

One well-studied general problem domain is a collaborative MAS environment populated with distinct tasks, where each task requires a tuple of resources on the agents' part in order for the agents to be able to complete that task [10, 11, 12, 17]. In this distributed task allocation context, agents need to form coalitions such that each coalition has sufficient cumulative resources or capabilities across the coalition members in order to be able to complete the assigned task – given some knowledge about that task's resource requirements [10-12].

Most of prior-art mathematical models and algorithmic solutions for the coalition formation problem are characterized by a property where the individual agents' coalition forming strategies are fixed; that is, agents don't learn and subsequently improve upon how they engage in future coalition formation interactions, based on the success attained by past strategies. However, most realistic collaborative multi-agent environments, including those where coalition formation naturally arises as a way of solving the distributed task and/or resource assignment problem, could benefit from the agents' ability to learn from past experience and hence be able to adapt their coalition formation strategies in future interactions.

There are at least two fundamental properties shared by many practical MAS applications that require the agents to be adaptable and capable of learning how to improve their coalition formation strategies. First, the same larger group of agents may need to engage in coalition formation interactions with each other repeatedly, and for the purpose of effectively coordinating in the same or similar kind of environment, they need to become effective at completing the same or similar set of repetitive tasks. Clearly, being able to learn from past experience and then improve in future coordination interactions would be very beneficial.

Second, most realistic MAS environments, including those where coalition formation naturally arises as a way of solving a distributed task or resource assignment problem as outlined above, are characterized by a number of possible sources of uncertainty and noise [3, 4, 7, 17]. Some of the sources of uncertainty and noise include (i) inaccuracies and inconsistencies in different agents' estimates of the tasks' utility values and/or resource requirements, (ii) a possibility of an agent's failure while working on a task as a part of one's current coalition [17], and (iii) inaccurate, incomplete and/or inconsistent estimates of individual agents' abilities and their potential contribution as members of various coalitions, that is, *imperfect beliefs* about other agents [4]. Once these sources of uncertainty are taken into account, and assuming agents would need to form coalitions repeatedly, clearly each agent should be able to learn how to better identify which candidate coalitions with other agents have a high chance of success, i.e., are most likely to succeed at completing future tasks.

We argue that a need for learning arises naturally in this kind of noisy, imperfect information collaborative MAS environments with repeated coalition formations and coalition-to-task assignments at multiple (in particular, at least two) qualitatively distinct levels. At one level, we find learning to identify individual agents that among their peers are better (more reliable and effective) coalition partners than others. In most scenarios that we have considered or found in the existing literature, this individual agent learning is of the *reinforcement* nature: an agent learns based on the past track record of rewards from various completed tasks, which were accomplished while the agent was a member of various coalitions. At a different level, we encounter a kind of learning that takes place at the 'system level' or agent ensemble level. For example, how would the MAS designer (e.g., in team robotics applications) or the central command-and-control (e.g., in emergency response, military or law enforcement applications) go about re-defining or modifying the agents' coalition formation strategies, the incentives given to the agents to form various different coalitions, and how to reconcile inconsistencies of different agents' views of the world, in order to make the future autonomous coalition formation process among its agents as effective as possible?

We make the case that what is required is to combine the reinforcement learning models and techniques with those of meta-learning. In our view, only such a multi-tiered approach to learning and adaptation in multi-agent coordination in general, and distributed coalition formation in particular, holds a true promise for making a breakthrough on *learning how to coordinate effectively*. That is, we argue that the problem of learning how to coordinate fundamentally needs to be tackled at different organizational and logical levels. Some form of reinforcement learning at the level of individual agents is necessary, but often not sufficient, for enabling a collaborative MAS to be able to considerably improve in its coordination abilities and effectiveness over time. Therefore, in our view, the first challenge is to understand the need for learning and adaptation at different levels of granularity; we try to make the case supporting this view. The second challenge is to identify and design the effective learning mechanisms at each of these different levels of granularity; while the state of the art is quite impressive already insofar as the RL of individual agents (as shall be seen in the next section), we argue that understanding how either co-learning or meta-learning can help improve coordination is, in both cases, in its infancy. The third challenge is to provide relevant formalisms and practical mechanisms for integrating RL of individual agents with co-learning of small, local groups of agents and meta-learning at the global, system level. Our work on addressing this third challenge, admittedly, is still at a very initial stage. Hence, the present paper will primarily focus on the first two challenges identified above. We will define more rigorously what we mean by co-learning and meta-learning in the sections that follow.

3. Reinforcement Learning for Better Coalition Formation

In the discussion that follows, we assume a collaborative multi-agent environment, and a general setting of coalition formation for task allocation purposes, as in [10, 11, 16, 17]. During the coalition formation process in such a distributed task allocation setting, each agent may potentially encounter various sources of uncertainty and noise that affect its effectiveness as well as its preferences over possible candidate coalitions with other agents. Uncertainty and noise may affect some or all of the following:

a. Agent's perception of various tasks, and in particular tasks' (i) utility values (to the agent and/or to the entire system) and (ii) resource requirements (i.e., how difficult is it going to be to complete those tasks, and, in particular, how many other agents, and which ones, would be required to jointly tackle any given task?);
b. Agent's perception of other agents, and in particular those other agents' capabilities and reliability as coalition members;
c. Inconsistencies in task preferences (e.g., in terms of different utility evaluations of a given task by different agents, or different estimates of that task's resource requirements) by different members of a potential coalition of agents.

In many applications, the same ensemble of agents will need to perform multiple stages of coalition formation and coalition-to-task mapping. Each member of such agent ensemble, therefore, could benefit from being able to learn which coalition partners are more reliable or useful than others, based on the past experience. In most situations, learning is of the *reinforcement* nature: rather than being provided clues by an outside teacher, an agent receives, in general, different *payoffs* for different choices of coalition partners and of tasks that the formed coalitions are mapped to. These differences in payoff outcomes are the result of varying effectiveness of different coalitions that this agent forms at different stages of the MAS deployment.

Several reinforcement learning models in the context of multi-agent coalition formation have been studied. In [4], Bayesian models of RL for coalition formation in the presence of noise are proposed. Each agent maintains its explicit beliefs about properties of other agents. These beliefs are then refined and updated based on experience, i.e., on prior outcomes resulting from repeated multi-agent interactions. In [4], an agent is learning to control a stochastic environment which is modeled as a Markov Decision Process (MDP). Direction in [1] assumes an underlying organizational structure of the multi-agent system, and a distributed coalition formation process that is guided by that organizational structure. The proposed approach uses RL techniques to improve upon local agent decisions within the larger organizational structure. That is, the learning is guided by the assumed organization structure of a MAS, but it does not take place at different organization levels – rather, it still mainly takes place at the level of individual agents. We will say more about this aspect when we outline how we believe meta-learning at the system level can improve multi-agent coordination. Another interesting approach is found in [22], where *genetic algorithms* (GA) are used to train agents to get better in how they form coalitions based on past experience. One observation about that approach is that this GA based training is, in general, *non-local*, in a sense that multi-agent interactions via e.g. crossovers may apply to agents that, in case of large, geographically dispersed MAS, may not actually be able to directly communicate with each other. Thus, in a sense, this approach contains elements of meta-learning (even though there is no explicit mention of meta-learning in that paper), or at least a form of global-level (as contrasted to local) learning; we will discuss meta-learning as a global, 'strategic' approach to improving distributed multi-agent coordination below.

Several other approaches based on RL at the level of an *individual agent* in order to improve effectiveness of coalition formation have been studied. More detailed surveys of the state-of-the-art of RL in the context of coalition formation can be found in [6] and [7]. In summary, RL-based approaches to improving MAS coordination via learning at the level of individual agents is by far the most investigated approach to learning how to better coordinate and how to improve upon agents' coalition formation effectiveness. While indisputably very useful, we argue that, in many important practical situations, RL of individual agents *alone* does not suffice.

4. Agent Co-Learning for Better Coordination and Coalition Formation

By (multi) agent co-learning, we will mean mutual observation of each other's behavior, and inferences that an agent makes based on observing how some of the other agents it interacts with have been acting during the course of those interactions. To contrast co-learning as informally defined here with the "traditional" reinforcement learning of a single agent, it's important to emphasize that the inferences and predictions about future interactions are based on various cues obtained from observing past specific behaviors of other agents, as well as various cues an agent may come up with on its own to direct (or manipulate) other agents into actions or behavior more desirable by that agent – and not just on the received payoffs from the past interactions (i.e., co-learning typically goes beyond relatively simplistic, "If I do action *a* and agent Y does action *a'*, I can expect payoff amount *$p*"). Typically, co-learning involves small groups of agents that repeatedly closely interact with each other. The main reason for this restriction to small groups is the computational cost associated with maintaining and updating ones model of other agents' beliefs, intentions etc. As we shall discuss shortly, co-learning in general can be applied to both strictly cooperative agents and self-interested agents.

We find some interesting prior art on co-learning in the context of multi-agent coordination (and, in one instance, coalition formation) in references [13] and [14]. We will discuss the conceptual framework and main ideas found in [13], and how they relate to our integrating multi-level learning in our framework, in the next section. In the present section, we focus on the main ideas found in [14]. In multi-agent domains where each agent is *self-interested* (so that, in general, agents may both compete and cooperate with each other), an agent may try to manipulate or induce other agents to change their behavior by choosing actions that, in the short run, may be suboptimal for the manipulating agent; namely, this agent hopes that, in the longer run, such an attempt at manipulating or "nudging" other agents will eventually result in other agents' change of behavior that would benefit the manipulator, that is, lead to a higher payoff to the manipulating agent in the long run [14]. Some examples of various types of (typically in this setting, *communicationless*) attempts of coordination via co-learning studied in [14] include manipulation by preemptive actions, manipulation by nudging, and manipulation into a mutual compromise. While more detailed discussion of coordination via such manipulative behavior is beyond our present scope, we will just mention *iterated prisoner's dilemma* (see, e.g., [21]) as a classical example where coordination may arise among self-interested agents via manipulation, nudging and co-learning mechanisms (especially when the agents in question happen to be self-interested humans!).

We agree with [14] that co-learning mechanisms can be very useful in enabling or enhancing multi-agent coordination. However, [14] primarily has in mind competitive, self-interested agents, whereas we study collaborative MAS in which the agents are unselfish distributed problem solvers. The nature of underlying agents certainly has considerable implications insofar as the appropriate coordination paradigm, including what are the appropriate models of learning how to coordinate. For instance, in our context, an agent need not worry about being manipulated by another agent into sub-optimal behavior, or about other agents' counter-action to its own attempts to manipulate or direct the behavior of others. Rather, agents have to 'sync up' as efficiently as possible in the presence of resource limitations, uncertainty and noise coming from the outside environment, the tasks that they are trying to solve jointly, and possibly the imperfect communication links between the agents (see, e.g., discussion in [17] for more details). For these reasons, elaborate (competitive) game-theoretic considerations and analyses in [14] aren't directly applicable to our setting. In particular, while the strictly collaborative MAS can still take advantage of the co-learning paradigm, the exact co-learning model cannot be directly taken from [14]. As far as we know, no prior work has addressed how to enable agents to co-learn in a collaborative, resource constrained distributed problem solving setting such as in references [10, 11, 12, 16, 17].

Lastly, we observe that the co-learning mechanisms discussed in [14] (which, interestingly enough, are being referred to as *meta-learning* by the author) are still being studied within the reinforcement learning context. In contrast, we'd like to emphasize the classical RL of a single agent from the mutual co-learning of (typically, small groups of) multiple agents since the latter requires more sophisticated and detailed model that an agent needs to have of other agents, and more computationally intensive manipulations of such models in order to be able to infer other agents' beliefs, intentions (for example, how would they respond to various forms of 'nudging'), etc. In collaborative MAS domains, the main issue with these requirements of co-learning is that of individual agents' resources, as well as (when it comes to physical agents such as robots, unmanned vehicles or smart sensors) the locality of agents' perceptions and communications, and the practical constraints that such locality implies (for more detailed discussion, see, e.g., [16, 17]).

Based on the discussion in this and the previous section, one may wonder, where exactly does the division line lie between "ordinary" single agent learning (such as RL models briefly reviewed in Section 3) on the one hand, and co-learning as discussed in this Section, on the other? One distinct characteristic of co-learning is that it entails higher order reasoning of an agent about other agents (and usually, such higher order reasoning is reciprocal, i.e. applies to all or most agents in the system). Examples of higher order reasoning include an agent having a model of other agents' beliefs, intentions, preferences, beliefs about those other agents' beliefs, etc. So, for instance, the RL mechanism in [4] (that we discussed in Section 3), an agent maintains and evolves explicit beliefs about other agents' preferences. While that paper makes no references to co-learning, we would argue that, in this instance, the distinction between pure individual agent reinforcement learning and multi-agent co-learning appears somewhat blurred. However, the third type of learning to coordinate that we address in this paper, namely, meta-learning, is in our view clearly distinguishable conceptually, logically and architecturally from RL of individual agents as well as co-learning of agent pairs or triples or other smaller (informal) groups. We discuss how meta-learning can enhance a collaborative MAS' ability to coordinate effectively next.

5. The Case for Meta-Learning

Various forms of reinforcement learning, as briefly discussed in the previous section, pertain to how an *individual agent* can adapt and improve its coalition formation strategy and selection of coalition partners. However, learning from past experience can take place at higher organizational levels than that of individual agents (or small groups of agents, in case of co-learning). In particular, it can take place at the system level, as well. Depending on the nature of MAS, this system level learning in general could refer to, e.g., self-organizing adaptability of agent ensembles or to meta-learning of the MAS system designer or other central authority. For example, in case of a collaborative MAS application of a system of autonomous *micro unmanned aerial vehicles* (micro-UAVs) on a complex, multi-stage, multi-task mission (see e.g. [15]), this higher-level learning could take place at the central command-and-control. We will discuss this application example in more detail in the next section.

In contrast to the relatively rich literature on individual agent's RL in the context of various MAS coordination problems in general, and the problem of coalition formation in particular, prior art on meta-learning [19, 20] applied to improving coordination among collaborative agents is very modest. We note that [14] studies meta-learning processes in MAS among self-interested agents that are *competing* with each other, as opposed to engaging in cooperative distributed problem-solving. This work focuses on algorithmic game-theoretic aspects of multi-agent interactions. In that context, a number of assumptions are made that are not suitable for out context, from competitive nature of inter-agent interactions to small, *a priori* known finite sets of available actions to each agent at each "move" of the "game". Furthermore, what [14] refers to as meta-learning is more properly described as agent co-learning, as the described learning mechanisms take place at the level of small groups of agents (in [14], typically two), and this learning does not involve a knowledge base or meta-datasets that capture the past experience and interaction patterns of all agents in the system, and then making inferences based on such a rich knowledge base.

The closest in spirit to our proposed approach to unifying traditional learning, co-learning and meta-learning for more effective dynamic coalition formation and task allocation in collaborative MAS is found in [13]. That paper addresses learning how to improve coalition formation at different organizational levels for general MAS that need not be strictly collaborative. The paper studies learning at what the authors refer to as 'tactical' and 'strategic' levels. At the tactical level (in this case, of an individual agent's decision-making), *reinforcement learning* is used to identify most viable candidates for coalition partners, whereas *case-based learning* (CBR) is used to refine specific negotiation strategies used by an individual agent. Tactical level takes place in an online manner; for applications involving teams of robots or ensembles of micro-UAVs, this tactical level reinforcement learning would basically need to happen in real time. At the strategic level, a distributed, cooperative CBR is used to improve the overall negotiation capabilities, thereby hopefully leading to a more effective coalition formation. The authors point out that strategic level learning would take place offline, not in real-time.

In our model, meta-learning, which is conceptually constituted of (i) creating the knowledge base with meta-data and then (ii) making inferences from that knowledge base, also takes place offline. The main reason for that is that meta-learning tends to be highly resource-demanding and to require integration of knowledge that cannot be captured either "internally" within an agent, or (in general) locally within small groups of agents. We note that [13] does not refer to, or employ, any specifically meta-learning techniques.

In our perspective, it is precisely this, strategic level where the offline, computationally and data intensive meta-learning techniques can be expected to be most powerful and of greatest practical use for those who design and/or deploy large-scale collaborative MAS. As long as the agent ensemble can be expected to engage in the same or similar type of interactions repeatedly, clearly these collaborative agents can hugely benefit from such offline analysis and useful inferences, coordination "hints" and "incentives" that would make subsequent interactions in general, and coalition formation interactions in particular, more successful *in the long run*. To ground this high-level discussion in reality, we discuss how this conceptual framework would apply to improving coordination (more specifically, coalition formation for the purposes of distributed task allocation) of an ensemble of micro unmanned aerial vehicles in the next section.

In a nutshell, this is how we propose that meta-learning would enable a better multi-agent coordination and more effective, adaptable and efficient coalition formation at the system or strategic level. Performance and past coordination strategies (incl. choice of coalition partners, tie-breaking mechanisms, and how successful various resulting coalitions were in performing the tasks that they were formed to address) of collaborative MAS can be stored in a meta-dataset, in a central *knowledge base* (KB). Such meta-dataset would contain various parameters that are used by the agents during the coalition formation process, where selected values of these parameters, in general, are associated with different levels of coalition formation efficiency and/or subsequent coalition successfulness. A meta-learning system can exploit this meta-knowledge to learn to associate various parameters with successfulness. This meta-learning system would use the large system-level KB with complex data sets in order to make complex, typically probabilistic/statistical inferences about the future effectiveness of various coalition formation strategies and choices, based on past history (that is, cumulative experience of individual agents in the system). Such cumulative experience and inferences based on that experience can be exploited to adjust how agents select future coalition partners, as well as to dynamically adapt coalitions and their overall capabilities to tasks and their resource requirements. In the fairly common MAS scenario where an agent ensemble repeatedly engages in coalition formation and coalition-to-task mapping interactions over a considerable time, the accumulated experience can reveal statistically relevant patterns to suggest the best coalition formation strategy for particular tasks [2].

We notice that, in most practical scenarios of our interest, accumulating and storing all this experience across the entire agent ensembles, as well as making non-trivial statistical inferences of the knowledge base created from that stored experience, would likely be beyond the computational resources of any single agent [17]. Moreover, assuming that the inter-agents interactions and communication are primarily local, both creation and subsequent usage of such a system- or ensemble-level knowledge base would also likely exceed the joint resources or abilities of smaller groups of agents, as well; as such, inferences at this level, and presumably the improved coordination abilities based on those inferences, would therefore also be beyond what is achievable via agent co-learning at the level of smaller groups of agents (such as agent pairs or triples – see [13] for more on co-learning mechanisms among small groups of agents). Therefore, complex statistical inferences at the level of entire (presumably large) agent ensemble aren't feasible to achieve at the lower organization levels, nor via either the classical reinforcement learning or co-learning mechanisms. We argue that a meta-learning approach is the right learning paradigm in this setting – in fact, in many applications, meta-learning is indeed necessary if the system designer hopes to take maximal advantage of the historical records of his or her MAS system performance. Moreover, in a number of important practical MAS applications, such meta-learning approach, and resources necessary for successfully undertaking it, are readily available – at least insofar as offline learning and inference are concerned. The results of such offline learning would be made available to the agents as those agents repeatedly engage in the same, or similar, type(s) of interactions.

What are the main differences between individual agent learning and co-learning on one hand, and meta-learning on the other? Beside the organizational "granularity level" we have already emphasized, some additional fundamental conceptual and architectural differences include the following:

- Reinforcement learning of individual agents and co-learning of (typically, small) groups of agents take place in a dynamic, online and (in many practical applications, such as micro-UAVs discussed in Section 6), real-time manner; in contrast, meta-learning takes place in an offline manner.
- RL and co-learning take place logically internally to each agent, and is hence limited by the time, memory, processing power and other resource limitations of those agents; in contrast, meta-learning takes place externally to the agents, and can be expected to have access to much larger data sets and knowledge bases, as well as to more raw computational power, than typical individual software, robotic or smart sensor agents.
- Individual learning and co-learning tent to be *locally constrained* by the sensing and communication limitations of individual agents, and these locality constraints carry over to the kind of data and knowledge available to the agents to learn from -- whereas meta-learning is not subject to such locality constraints.
- Meta-learning and meta-reasoning inference engines can be expected to have an access to much richer data sets, knowledge representations, and type of knowledge than the individual agents' inference mechanisms that are trying to learn in an online manner, within those agents' locality and resource constraints.

- The time scales at which reinforcement learning vs. meta-learning would enable considerable improvement in coordination effectiveness can be expected, in general, to be considerably different; that is, the benefits of meta-learning should be expected in the long(er) run.

In the sequel, we summarize out multi-tiered, multi-level learning to coordinate framework, which is based on integrating reinforcement learning, co-learning and meta-learning. The main challenges and steps in our programmatic approach to learning how to effectively coordinate in collaborative MAS domains are as follows:

- recognizing the need for meta-learning;
- defining appropriate conceptual and theoretical meta-learning formalisms and frameworks suitable for collaborative multi-agent coordination purposes;
- defining appropriate conceptual and theoretical co-learning formalisms and frameworks suitable for collaborative multi-agent coordination purposes, and for the given types of agents' resource limitations;
- designing practical meta-learning systems (where by meta-learning systems we mean, both the knowledge bases and the inference engines that extract, or learn, meta-knowledge out of meta-datasets collected and stored in those knowledge bases), and lastly
- successfully integrating the meta-learning approach and techniques with the learning techniques (such as, but not necessarily limited to, the 'traditional' individual agent reinforcement learning as well as co-learning among small groups of agents) that take place at lower organizational and logical levels of a collaborative MAS.

To illustrate how the proposed program could be applied to "real", practical collaborative multi-agent systems, in the next section we briefly discuss an example of a well-known and important collaborative MAS application, and how would our conceptual framework of integrating reinforcement learning, co-learning and meta-learning work in the context of that application.

6. An Example: Coordinating a Large Ensemble of Micro-UAVs on a Complex, Multi-Task Mission

A collection of *micro unmanned aerial vehicles* (micro-UAVs) that are autonomous (in particular, not remotely controlled by either a human operator or a computer program) and that need to coordinate with each other in order to accomplish a complex, multi-task mission in a highly dynamic and unpredictable, partially observable environment provide an ideal tested for modeling, designing and analyzing large-scale collaborative MAS operating in "the real world". Such ensembles of micro-UAVs can be used for various surveillance, reconnaissance, search-and-rescue and other similar tasks, including longer-term missions made of a variety of such tasks (e.g., [15]). Such micro-UAVs are, in general, equipped with sophisticated sensors (radars, infra-red cameras etc.), actuators or "payloads", and communication links (typically, radios). Their communication, in general, may include peer-to-peer (i.e., single UAV-to-UAV) message exchanges, local broadcasts or multicasts, global broadcasts/multicasts, and message exchanges with centralized command-and-control. The coordination problems encountered by a team or ensemble of micro-UAVs can range from conceptually simple (but quite challenging in practice) collision avoidance to distributed divide-and-conquer "single shot" task allocation to complex fully or partially distributed planning [15, 17]. Some consensus problems that naturally arise in UAV deployments that are fully distributed (for example, when no communication with command-and-control is possible or feasible) include coalition formation and leader election (e.g., [16]).

A large micro-UAV ensemble on a complex, multi-stage mission comprised of diverse tasks with varying time and other resource requirements provide an excellent context for multi-tiered learning of how to better coordinate. A variety of tasks and their resource demands, complexity of the overall environment, a variety of coordination problems that the UAVs may encounter in the course of their mission, multiple time scales at which the overall system can use learning and adaptation in order to perform better in the future, and multiple logical and organizational levels at which large such micro-UAV ensembles can be analyzed and optimized, all suggest the need for a multi-tiered approach to learning. At the level of an individual UAV, standard reinforcement learning paradigm is suitable. Due to space constraints, we won't discuss it further; we will focus, instead, on co-learning and meta-learning in the outlined setting.

Co-learning among small groups of UAVs could take place along similar lines to the approaches in [4] or [14]. One caveat is that the need for one agent to model some of the other agents explicitly would not be motivated by differing, possibly conflicting, interests of different agents. Instead, it would be due to any combination of the following: (i) imperfections of communication links, (ii) inaccuracies in how agents evaluate tasks and, in particular, the suitability of their own capabilities or resources to perform those tasks (either on their own, or as a part of a coalition with other agents), (iii) inconsistencies in perceived value and resource requirements of a task as seen by different agents, (iv) different capabilities of agents, and (v) differing, possible inconsistent, beliefs about each other's capabilities). Consider a simple specific example: an agent, A, identifies some task *T* that A estimates would require two agents of A's capabilities to complete. Among near-by UAVs, A can pick UAV B or UAV C to form a two-member coalition that would be assigned to task *T*. Ability of A (and other agents, including B and C in this case) to co-learn would enable agent A to (i) solicit feedback from B and C on how they view task *T*'s value and resource requirements, and to compare those with its own view of the task, (ii) based on past interactions with B and C, to have a preference for one over the other as a coalition partner, (iii) to learn from B and C if they happen to have identified other tasks worth completing, (iv) to have a degree of trust or confidence in B's and C's evaluations of their own abilities, as well as (v) of the values, and resource requirements, of other tasks (if any) that B and C may be interested in. Based on (i) – (v), agent A may be able to make a more informed decision on matters such as (a) whether to still pursue task T or opt for some T' that it learns about from B or C, (b) which of the alternative tasks (if more than one such T' exists) to choose, and (c) which coalition partner, B or C, to choose as preferred coalition partner for the task of choice (assuming the chosen task still requires two agents).

Co-learning as outlined above, however, could hardly be expected to scale up; that is, a micro-UAV can perhaps maintain explicit models of a handful of other micro-UAVs, but in case of a very large ensemble (made of hundreds or thousands of such micro-UAVs), trying to model most or all of other agents would simply exceed the memory and processing power of an individual agent. Moreover, such swarm micro-UAV deployments would likely entail each UAV being able to see and/or directly communicate with only a handful of others; and flooding this ad hoc network of micro-UAVs with the global information (say, sent from the central command-and-control) would likely not work well either, both from the communication cost standpoint and from the processing stand-point (i.e., even if each micro-UAV got all the information there is about all other UAVs, would it have the resources for effectively using all that information?) We argue that, to take advantage of accumulated global knowledge and meta-knowledge about all the agents in the system, their past interactions, various tasks and their properties, and successfulness of different previously used coordination strategies, a meta-learning approach is required. This meta-learning would take place offline, presumably at a centralized command-and-control. The kinds of meta-data that the KB at command-and-control center would store could include properties of all tasks encountered by any of the micro-UAVs to date (incl. the payoffs or value associated with those tasks, the combined capabilities/resources of agents or agent coalitions that were involved in completing the tasks, etc.), individual experiences of agents across the entire large ensemble and across epochs, and meta-knowledge derived from the raw data (for example, the appropriate agent coalitions and expected resources and time to complete particular type of tasks, such as search-and-rescue, in a particular type of environment, such as mountainous terrain, about 100 square miles land area across which the search took place, particular weather conditions, etc.). Such meta-knowledge could then be used in the second stage of meta-learning inference engine's operation to provide the agents with summaries of all the past experience in forms of task and candidate coalition rankings, "bonus" incentives to build coalition with one subset of agents instead of another, revised estimates of the values (expected payoffs) and resource requirements of new tasks that are essentially similar to some of the previously encountered tasks, and so on.

Once again, the main point here is that inferring useful meta-knowledge and then using that meta-knowledge to help agents revise their beliefs and intentions and ultimately to coordinate more efficiently and effectively, would in most situations be beyond the computational, communication and storage resources of individual agents, as well as the joint capabilities of small groups of agents. Therefore, large-scale micro-UAV deployments, and many collaborative large-scale MAS conceptually similar to micro-UAVs, could uniquely benefit from meta-learning and meta-reasoning techniques, that would necessarily need to take place offline and externally to the agents themselves, in contrast to reinforcement learning and co-learning mechanisms reviewed earlier in this paper.

7. Summary

The central thesis of this paper is that, in order to be able to design distributed collaborative intelligent systems that are capable of learning how to effectively coordinate, it is necessary to simultaneously address learning and adaptation at several qualitatively distinct levels: that of the individual agents, that of small groups of agents, and that of the entire distributed system or the MAS designer. The first level has been extensively studied in the literature, and many, mostly reinforcement learning based approaches to the individual agent learning have been proposed. At the level of small, local groups of robotic or other agents that operate in the physical space, there have been some investigations of the co-learning mechanisms and how they can help improve coordination; however, this area is still very immature. Lastly, at the level of the entire system, or very large, possibly geographically dispersed agent ensembles, we propose a meta-learning approach – something that, as far as we know, has not been systematically or satisfactorily addressed by the research community.

The particular multi-agent coordination problem where the interaction of reinforcement learning, co-learning and meta-learning is of a considerable interest is the problem of dynamic coalition formation for distributed task and/or resource allocation [10, 12, 16, 17]. However, as far as we know, this interplay of these different learning modes in the coalition formation setting has not yet been adequately addressed in the research literature.

We posit that our approach is both novel and solidly grounded in properties of many collaborative MAS applications. To provide some application-level support for that thesis, we apply our conceptual framework of multi-level learning how to better coordinate to a well-known application, that of distributed coalition formation among a large ensemble of micro-UAVs that are deployed on a complex, non-episodic multi-task mission. In our ongoing research, we are combining reinforcement learning, co-learning and meta-learning models and techniques to the important problems in distributed multi-agent coordination, in order to systematically address a very fundamental problem in collaborative MAS – the problem of learning how to coordinate effectively.

References

[1] Abdallah, S., Lesser, V., *Organization-Based Cooperative Coalition Formation, in* Proc. IEEE / WIC / ACM Int'l Conf. Intel. Agent Technology, IAT (2004)

[2] P.Brazdil, C. Giraud-Carrier, C. Soares, R. Vilalta. *Metalearning: Applications to Data Mining.* Springer (2009)

[3] G Chalkiadakis, C Boutilier. *Coordination in Multiagent Reinforcement Learning: A Bayesian Approach*, in Proc. Autonomous Agents & Multi-Agent Systems AAMAS'03 (2003)

[4] Chalkiadakis, G., Boutilier, C., *Bayesian reinforcement learning for coalition formation under uncertainty*, in Proc. AAMAS'04 (2004)

[5] Jang, M., Reddy, S., Tosic, P., Chen, L., Agha, G. *An Actor-based Simulation for Studying UAV Coordination*, in 16th Euro. Simulation Symposium (ESS '03), pp. 593-601, Delft (2003)

[6] Li, X., Soh, L.K. *Investigating reinforcement learning in multiagent coalition formation*, TR WS-04-06, AAAI Workshop Forming and Maintaining Coalitions & Teams in Adaptive MAS (2004)

[7] de Oliveira, D. *Towards Joint Learning in Multiagent Systems Through Opportunistic Coordination*, PhD Thesis Proposal, Univ. Federal Do Rio Grande Do Sul, Brazil (2007)

[8] T.W. Sandholm, V.R. Lesser. *Coalitions among computationally bounded agents*, in *Artificial Intelligence* vol. 94, pp. 99-137 (1997)

[9] T. W. Sandholm, K. Larson, M. Andersson, O. Shehory, F. Tohme. *Coalition structure generation with worst case guarantees*, in *AI Journal* vol.111 (1-2), pp. 210-238 (1999)

[10] O. Shehory, S. Kraus. *Task allocation via coalition formation among autonomous agents*, in Proc. IJCAI-95, Montréal, pp. 655–661 (1995)

[11] O. Shehory, K. Sycara, S. Jha, *Multi-agent coordination through coalition formation*, in *Intelligent Agents IV: Agent Theories, Architectures and Languages*, LNAI, vol. 1365, pp. 153-164, Springer (1997)

[12] Shehory, O., Kraus, S. *Methods for task allocation via agent coalition formation,* in *AI Journal* vol. 101 (1998)

[13] Soh L. K., Li, X. *An integrated multilevel learning approach to multiagent coalition formation*, in Proc. Int'l Joint Conf. on Artif. Intel. IJCAI'03 (2003)

[14] R. Sun. *Meta-Learning Processes in Multi-Agent Systems*, in *Intelligent agent technology: research and development*, N. Zhong, J. Liu (eds.), pp. 210-219, World Scientific, Hong Kong (2001)

[15] Tosic, P., Agha, G. *Understanding and Modelling Agent Autonomy in Dynamic Multi-Agent, Multi-Task Environments*, in Proc. EUMAS '03, Oxford, England (2003). See also Tosic et al., *Modeling a System of UAVs on a Mission,* invited session on agent-based computing, in Proc. 7th World Multiconference on Systemics, Cybernetics, and Informatics (SCI'03), pp. 508-514, 2003

[16] Tosic, P., Agha, G. *Maximal Clique Based Distributed Coalition Formation for Task Allocation in Large-Scale Multi-Agent Systems*, in LNAI vol. 3446, pp. 104-121, Springer (2005)

[17] Tosic, P. *Distributed Coalition Formation for Collaborative Multi-Agent Systems,* MS thesis, Univ. of Illinois at Urbana-Champaign (UIUC), Urbana, Illinois, USA (2006)

[18] Vig, L., Adams, J.A. *Issues in multi-robot coalition formation*, in Proc. Multi-Robot Systems: From Swarms to Intel. Automata, #3 (2005)

[19] Vilalta, R. *Research Directions in Meta-Learning: Building Self-Adaptive Learners*, in Int'l Conference on AI, Las Vegas, Nevada (2001)

[20] Vilalta, R., Giraud-Carrier, C., Brazdil, P. *Meta Learning Concepts and Techniques* (2005)

[21] Wooldridge, M. *An Introduction to Multi-Agent Systems*, Wiley (2002)

[22] Yang, J. , Zhenghu, L. *Coalition formation mechanism in multi-agent systems based on genetic algorithms,* in *Applied Soft Computing*, vol. 7 no. 2 (2007)