Team Formation with Learning Agents that Improve Coordination

(Extended Abstract)

Somchaya Liemhetcharat
Institute for Infocomm Research
Agency for Science, Technology and Research
Singapore 138632, Singapore
liemhet-s@i2r.a-star.edu.sg

Manuela Veloso
Computer Science Department
Carnegie Mellon University
Pittsburgh, PA 15213, USA
veloso@cs.cmu.edu

ABSTRACT

Learning agents increase their team's performance by learning to coordinate better with their teammates, and we are interested in forming teams that contain such learning agents. In particular, we consider finite training instances for learning agents to improve their coordination before the final team is formed. We formally define the learning agents team formation problem, and focus on learning agent pairs that improve their coordination. Learning agent pairs have heterogeneous rates of improving coordination, and hence the allocation of training instances has a large impact on the performance of the final team.

Categories and Subject Descriptors

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

General Terms

Algorithms, Experimentation

Keywords

Team formation; ad hoc agent; learning agent

1. INTRODUCTION

In multi-agent team formation, the capabilities of agents are typically assumed to be fixed, and the performance of a team is the sum of single-agent capabilities. We recently introduced the Synergy Graph model, where team performance is a function of single-agent capabilities and the coordination among pairs of agents [3].

What if some agents can learn to improve their coordination with teammates? For instance, some learning agents model their teammates in order to improve the overall team performance (e.g., [1, 2]). We consider *learning agent pairs*: pairs of agents that simultaneously learn, and pairs consisting of a learning and a regular agent. In particular, our formulation is general in that the rate at which agent pairs improve coordination is modeled, and not *how* they improve.

Appears in: *Alessio Lomuscio, Paul Scerri, Ana Bazzan, and Michael Huhns (eds.), Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2014), May 5-9, 2014, Paris, France.*

We formally define the learning agents team formation problem, where there are a fixed number of training instances for learning agent pairs to improve coordination. The goal is to form a multi-agent team with high performance after all the training instances. Sports provides a motivating example, where there are limited opportunities for a coach to train his team before the actual game. The coach allocates training instances to pairs, and after all the training, the coach selects which members form the team.

Thus, the performance of the formed team is closely linked to the allocation of training instances. For example, consider a team with low performance but with learning agent pairs that improve quickly. After training, such a team may outperform other teams that do not have learning agent pairs, or pairs that improve slowly. These heterogeneous rates of learning are initially unknown, so the allocation of training instances has to balance between exploring and exploiting, i.e., to improve the estimates of the learning rates and to improve the performance of the final formed team.

2. FORMAL PROBLEM DEFINITION

In this section, we formally define the learning agents team formation problem. We first begin with the set of agents and the definition of a team.

Let $\mathcal{A} = \{a_1, \ldots, a_N\}$ be the set of agents, where each $a_i \in \mathcal{A}$ is an agent. Let any $A \subseteq \mathcal{A}$ be a team. We allow teams to be any subset of $\mathcal{A}$ to signify that teams can be of different sizes in general, following our formalism in the Synergy Graph model [3].

In the Synergy Graph model, agents are vertices in a connected graph, and each agent a_i's capability is represented as a Normally-distributed variable $C_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$. We use a modified version of the Synergy function of our Synergy Graph model to compute the performance of a team $A \subseteq \mathcal{A}$:

$$P(A) = \frac{1}{\binom{|A|}{2}} \sum_{\{a_i,a_j\} \in A} P_2(a_i, a_j), \text{ such that} \tag{1}$$

$$P_2(a_i, a_j) = \phi_{i,j} \cdot (C_i \ + \ C_j)$$

$\phi_{i,j} \in \mathbb{R}^+$ is the level of coordination of a_i and a_j, that is a function of the distance between a_i and a_j in the Synergy Graph model. In this paper, we ignore the distance between a_i and a_j in the graph, and focus solely on $\phi_{i,j}$.

Eqn. 1 shows that the performance of a team is a function of the agents' capabilities and their coordination. In our previous work [3], we assumed that the capabilities and coordination were fixed and unknown, and we contributed an algorithm that learned them from observations. In this work, we are interested in learning agents, and hence team performance improves over time. Learning agents can be represented in two ways: an improvement in the agent capability C_i, or an improvement in the coordination $\phi_{i,j}$. Improvement in an agent's capability signifies that the agent learns more about the *task* through experience; improvement in the coordination signifies that the agent is learning to *coordinate* better with its teammates. We are interested in the latter case.

In particular, we are interested in *learning agent pairs* $\{a_i, a_j\} \in \mathcal{A}^2$. Let $\mathcal{L} \subset \mathcal{A}^2$ be the set of all learning agent pairs. There are K training instances, where each training instance is allocated to a learning agent pair $\{a_i, a_j\} \in \mathcal{L}$.

Let $\phi_{i,j}^{(0)}$ be the initial coordination of a learning agent pair $\{a_i, a_j\} \in \mathcal{L}$, and let $\phi_{i,j}^{(k)}$ be the coordination after the k^{th} training instance. Since each training instance is allocated to a single learning agent pair, $\phi_{i,j}^{(k)} > \phi_{i,k}^{(k-1)}$ if and only if $\{a_i, a_j\}$ was allocated the k^{th} training instance.

After every training instance is allocated, an observation $o_{i,j} \sim P_2(a_i, a_j)$ is obtained. Note that $P_2(a_i, a_j)$ improves on expectation as training instances are allocated to $\{a_i, a_j\}$, since $\phi_{i,j}^{(k)}$ increases.

The goal of the learning agent team formation problem is to form the optimal team after the K training instances. We assume that the size n^* of the optimal team is given, and define the optimal team:

$$A_K^* = \text{argmax}_{A \subseteq \mathcal{A} \text{ s.t.} |A| = n^*} E(P(A)) \qquad (2)$$

The performance of a team $A \subseteq \mathcal{A}$ after K training instances depends on the number of learning agent pairs in A, and the number of training instances each pair was allocated. Hence, the allocation of training instances has a large impact on the performance of the formed team.

3. OUR APPROACH

While we use the modified Synergy function to model team performance [3], our approach is general and other multi-agent team models are applicable. We consider two models of coordination improvement in learning agent pairs:

$$\phi_{i,j}^{(K)} = \phi_{i,j}^{(0)} + k_{i,j} * l_{i,j} \qquad (3)$$

$$\phi_{i,j}^{(K)} = \phi_{i,j}^{(0)} + \sum_{k=1}^{k_{i,j}} l_{i,j} \cdot \gamma_{i,j}^{k-1} \qquad (4)$$

where $k_{i,j} \leq K$ is the number of training instances allocated to $\{a_i, a_j\}$, $\sum_{\{a_i,a_j\} \in \mathcal{L}} k_{i,j} = K$, and $l_{i,j}$ is the *learning rate* of $\{a_i, a_j\}$.

Eqn. 3 shows a linear model, where a learning agent pair improves its coordination by $l_{i,j}$ after every training instance. Eqn 4 shows a geometric model, where the marginal coordination improvement decreases by a factor of $\gamma_{i,j}$ after each training instance allocated to $\{a_i, a_j\}$.

The learning rates $l_{i,j}$ are initially unknown. Training instances are allocated iteratively, and we use the observations $o_{i,j}$ to improve our estimate of $l_{i,j}$. We assume that the only unknowns are $l_{i,j}$ and we use a Kalman filter to estimate it with the observations $o_{i,j}$. We contribute algorithms that balance exploring (improving $l_{i,j}$'s estimate) and exploiting (training a pair that improves team performance).

4. COMPARISON WITH MULTI-ARMED BANDIT PROBLEM

The learning agent team formation problem and multi-armed bandit problem have similarities, by considering a learning agent pair as an arm, and a training instance as a pulling an arm. Each learning agent pair has an initially unknown learning rate $l_{i,j}$, similar to the probability of reward in an arm. Every training instance provides an observation that improves the estimate of $l_{i,j}$, and there exists an optimal allocation of training instances if all $l_{i,j}$ were known.

The main difference between the two problems is their goals. The goal of the learning agents team formation problem is to maximize the mean performance of a team after the K trials, while the goal of the multi-armed bandit problem is to maximize the cumulative sum of rewards.

In the learning agents team formation problem, allocating a training instance to a learning agent pair improves its coordination. However, the performance of the final team may not be affected unless the pair is in the team. For example, if $a_i, a_j \notin A$ then the improvements from $k_{i,j} \leq K$ training instances did not affect A's team performance. In the bandit problem, pulling an arm increases the cumulative reward.

5. CONCLUSION

We formally define the learning agent team formation problem, where team performance is a function of the agents' capabilities and their pairwise coordination, and the goal is to form a team with maximum team performance after all training instances are allocated. The coordination of a learning agent pair increases when training instances are allocated to it, and we consider linear and geometric learning rates. These learning rates are initially unknown, and our approach iteratively allocates training instances and updates the estimates of learning rates.

Acknowledgments

This work was partially supported by the Air Force Research Laboratory under grant number FA87501020165, by the Office of Naval Research under grant number N00014-09-1-1031, and the Agency for Science, Technology, and Research (A*STAR), Singapore. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of any sponsoring institution, the U.S. government or any other entity.

6. REFERENCES

[1] N. Agmon and P. Stone. Leading Ad Hoc Agents in Joint Action Settings with Multiple Teammates. In *Proc. Int. Conf. Autonomous Agents and Multiagent Systems*, pages 341–348, 2012.

[2] S. Barrett, P. Stone, and S. Kraus. Empirical Evaluation of Ad Hoc Teamwork in the Pursuit Domain. In *Proc. Int. Conf. Autonomous Agents and Multiagent Systems*, pages 567–574, 2011.

[3] S. Liemhetcharat and M. Veloso. Modeling and Learning Synergy for Team Formation with Heterogeneous Agents. In *Proc. Int. Conf. Autonomous Agents and Multiagent Systems*, pages 365–375, 2012.