MAS: If multi-agent learning is the answer, what is the question?

Yoav Shoham, Rob Powers, Trond Grender

Presented by DJ Carpet:

Derek Weitzel, Jinfu Leng,

Christopher Niemann

University of Nebraska Lincoln

Paper Reviewed

Y. Shoham, R. Powers, and T. Grenager, “If multi-agent learning is the answer, what is the question?” *Artificial Intelligence,* vol. 171, no. 7, pp. 365–377, 2007.

Outline

- Introduction
- Characteristics of Multi-agent Learning
- Sampling of Multi-agent Learning
- Agendas of Multi-agent Learning
- Summary

Introduction

- Multi-agent learning (MAL) has seen explosive growth in AI research.

- For all the research, there has been large number of questions that have been raised that haven’t been answered.

Introduction - Questions about MAL

- What exact question or questions is MAL addressing?

- How do we measure the answers to these questions?

Introduction – Adapting AI to Multi-agent

- Significant MAL research has been adapting AI learning methods to multi-agent setting.
  - For example, Q learning

- AI Learning is not always applicable, AI focuses on 1 vs. 1, multi-agent is many vs. many.

Scope of Paper

- This paper will focus on Stochastic Games
  - Multiple agents that have strategies, rewards, and a probability function on transitioning to a next state.
  - Our learning day was a Stochastic Game.

Outline

- Introduction
- Characteristics of Multi-agent Learning
- Sampling of Multi-agent Learning
- Agendas of Multi-agent Learning
- Summary

Learning form Stochastic Games

- Agents can learn from a Stochastic Game in two methods
  - Model-based learning – Create a model for opponents, play best response. Update model after observing opponent's actions.
  - Model-free learning – Do not create a model for opponents, instead over time learn how one's own actions do in the environment.

Characteristics of Multi-agent Learning

- Example that shows one of these characteristics.
- 2 Players – Row and Column

	$Left$	$Right$
Up	$1, 0$	$3, 2$
$Down$	$2, 1$	$4, 0$

Characteristics of Multi-agent Learning

- Row's turn – Picks down

	$Left$	$Right$
Up	$1, 0$	$3, 2$
$Down$	$2, 1$	$4, 0$

Characteristics of Multi-agent Learning

- Column's turn – Picks left

	$Left$	$Right$
Up	$1, 0$	$3, 2$
$Down$	$2, 1$	$4, 0$

Characteristics of Multi-agent Learning

- Not an optimal strategy for either row or column.
- They can do better…

	$Left$	$Right$
Up	$1, 0$	$3, 2$
$Down$	$2, 1$	$4, 0$

Characteristics of Multi-agent Learning

- Row learns column's strategy
- Row's turn – Picks up

	$Left$	$Right$
Up	$1, 0$	$3, 2$
$Down$	$2, 1$	$4, 0$

Characteristics of Multi-agent Learning

- Column keeps strategy, best outcome
- Column's turn – Picks right

	$Left$	$Right$
Up	$1, 0$	$3, 2$
$Down$	$2, 1$	$4, 0$

Characteristics of Multi-agent Learning

- Simple but profound example:
  - In multi-agent systems, one cannot separate *learning* from *teaching*.

	$Left$	$Right$
Up	$1, 0$	$3, 2$
$Down$	$2, 1$	$4, 0$

Characteristics of Multiagent Learning

- Simple but profound example:
  - In multi-agent systems, one cannot separate *learning* from *teaching*.

- Quick conclusion:

  There is no reason to expect that machine learning techniques for AI in single-agent settings to prove relevant in multi-agent settings.

Characteristics of Multiagent Learning

- Consider Rock-Paper-Scissors
- Unique Nash-Equilibrium uniformly distributed across all three choices.
- Can you win by randomly picking options?

	Rock	*Paper*	*Scissors*
Rock	$0, 0$	$1, 1$	$1, 1$
Paper	$1, 1$	$0, 0$	$1, 1$
Scissors	$1, 1$	$1, 1$	$0, 0$

Characteristics of Multiagent Learning

- No, can’t win by randomly picking, especially if the opponent is not.
- In such complex games it is not reasonable to expect that players contemplate the entire strategy space.
- Therefore equilibrium doesn’t play here a great predictive or prescriptive role.

Outline

- Introduction
- Characteristics of Multi-agent Learning
- Sampling of Multi-agent Learning
- Agendas of Multi-agent Learning
- Summary

A (very partial) sample of MAL work

- *Some MAL techniques*
- *Some typical results*
- *Some observations and questions*

Some MAL techniques

- *Model-based approaches*
- *Model-free approaches*
- *Regret minimization approaches*

Model-based approaches

General scheme:

1. Start with some model of the opponent's strategy.
2. Compute and play the best response.
3. Observe the opponent's play and update your model of her strategy.
4. Go to step 2.

Model-based approaches (example)

http://www.telegraph.co.uk/news/worldnews/europe/france/8404781/Chess-world-rocked-by-French-cheating-scandal.html

Model-free approaches

- An entirely different approach which avoids building an explicit model of the opponent's strategy.
- Instead, over time one learns how well one's own various possible actions fare.

Regret minimization approaches

- A learning rule is universally consistent or (equivalently) exhibits no regret if, loosely speaking, against any set of opponents it yields a payoff that is no less than the payoff the agent could have obtained by playing any one of his pure strategies throughout.

Regret minimization approaches (example)

regret $r_i^t(a_j, s_i)$: agent i for playing the sequence of actions s_i instead of playing action a_j, given that the opponents played the sequence s_{-i}.

$$r_i^t(a_j, s_i | s_{-i}) = \sum_{k=1}^{t} R(a_j, s_{-i}^k) - R(s_i^k, s_{-i}^k)$$

The agent then selects each of its actions with probability proportional to max $(r_i^t(a_j, s_i), 0)$ at each time step $t+1$.

Some typical results

- Convergence of the strategy profile to an (e.g., Nash) equilibrium of the stage game in self play (that is, when all agents adopt the learning procedure under consideration).
- Successful learning of an opponent's strategy (or opponents' strategies).
- Obtaining payoffs that exceed a specified threshold.

Some observations and questions (1)

- While the learning procedures apply broadly, the results for the most part focus on self play. They also tend to focus on games with only two agents. Why does most of the work have this particular focus?

Some observations and questions (2)

- With the exception of no-regret learning, the work focuses on the play to which the agents converge, not on the payoffs they obtain. Which is the right focus?

Some observations and questions (3)

- No-regret learning is distinguished by its starting with criteria for successful learning, rather than a learning procedure. Are the particular criteria adequate?

Outline

- Introduction
- Characteristics of Multi-agent Learning
- Sampling of Multi-agent Learning
- Agendas of Multi-agent Learning
- Summary

Five distinct agendas in multi-agent learning (MAL)

- Prerequisite in the field of MAL is to be very explicit about the problem being addressed
- Five distinct goals
  1. Computational
  2. Descriptive
  3. Normative
  4. Prescriptive, cooperative
  5. Prescriptive, non-cooperative

Computational

- Views learning algorithms as an iterative way to compute properties of the game

Computational

- Example:
  - Fictitious play was proposed to compute a sample Nash equilibrium for zero-sum games.

- Advantages: quick-and-dirty → easily understood and implemented

Descriptive

- Asks how natural agents learn in the context of other learners

Descriptive

- Goal is to investigate formal models for learning that that agree with people's behavior or other agents (from Lab results)
- This can be applied to large-population models
- Applications in social science and economics

Normative

- Focuses on which sets of learning rules are in equilibrium with each other, and which repeated-game strategies are in equilibrium

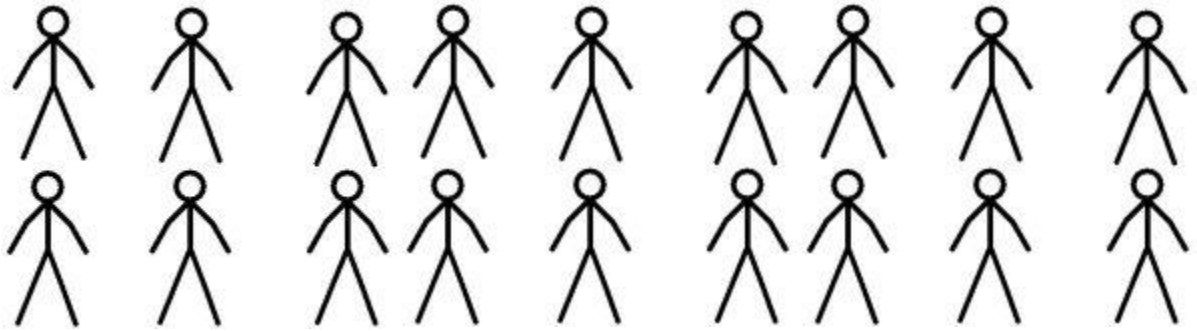
	$Left$	$Right$
Up	$1, 0$	$3, 2$
$Down$	$2, 1$	$4, 0$

Normative

- Example:
  - Can ask if fictitious play and Q-Learning (initialized appropriately) are in equilibrium with each other in the Prisoner's Dilemma game
- Problematic rule of equilibria

Prescriptive, Cooperative

- How agents should learn; involving distributed control in the dynamic system



Prescriptive, Cooperative

- Best model: repeated or stochastic common-payoff (‘team’) game
- Approaches can be evaluated based on the value achieved by the joint policy and the resources required (computation, communication, learning time)

Prescriptive, non-cooperative

- How should an agent act to obtain reward in the repeated game
- What would be an effective agent for a given system with other agents
- Equilibrium is not a goal here

Outline

- Introduction
- Characteristics of Multi-agent Learning
- Sampling of Multi-agent Learning
- Agendas of Multi-agent Learning
- Summary

Summary

1. Learning in MAS is conceptually, not only technically challenging.
2. One needs to be crystal clear about the problem being addressed and the associated evaluation criteria.

Summary

3. For the Field to advance one cannot simply define arbitrary learning strategies, and analyze weather the resulting dynamics converge in certain cases to a Nash equilibrium or some other solution concept of the stage game.

Summary

4. Five coherent agendas.
5. Not all work in the field falls into one of these buckets. This means that either we need more buckets, or some work needs to be revisited or reconstructed so as to be well grounded.

Group Summary

- Praises:
  - Paper took a step back, asked “Why are we using Multi-agent learning?”
    - Asked questions about current MAL research.
  - Compared AI learning with Multi-agent Learning, concluding that techniques used in AI are not always applicable in MAL.
  - Described current MAL research

Group Summary

- Critiques
  - Described scenarios where MAL or AI learning doesn't help, but offered no useful hints for solutions other than cryptic conclusion:

Our point has only been that in the context of complex games, so-called "bounded rationality", or the deviation from the ideal behavior of omniscient agents, is not an esoteric phenomenon to be brushed aside.

  - Survey paper – very light on details

Group Summary – Relation to Project

- If our project had learning, it would be descriptive.
  - Simulating a real life scenario.
  - Agents are in an environment full of learners.
  - Agents would learn from past trades that their current method isn't working, and update their view of the environment to be more risk averse
- Reinforcement learning is easy when we have explicit reward – Profit!

Questions?