Mechanism Design for Multiple AIs

1 Introduction

The project will focus on mechanism design for human and artificially and intelligent agents. The aim is to develop mechanisms that enables their co-operation. The PhD program will be principally taking place in the University of Lille (3), embedded in the team SequeL at INRIA-Lille. The PhD student will also have the opportunity to be a visiting scholar at Harvard University, and at Chalmers University, with a possibility for a dual-department thesis.

The main supervisor will be Christos Dimitrakakis at the university of Lille. If necessary, Alessandro Lazaric may act as a co-supervisor at INRIA, and David Parkes at Harvard.

2 Research Overview

The PhD will be a 3-year program of fundamental research, funded by the Future of Life Institute, to understand the role of mechanism design, multi-agent dynamical models, and privacy preserving algorithms, in promoting the emergence of beneficial AI, for example, social-welfare maximizing AI, in multi-agent systems, and especially in multi-agent systems in which the AIs are built through reinforcement learning.

In making progress, we propose to study two specific multi-agent learning or planning problems, both situated within the formalism of Markov decision processes. The first is *experiment design*, typically formalized as a *multi-armed bandit process* [Che59, Rob85], which we intend to study in a multi-agent, privacy-preserving setting. The second is the more general problem of learning to act in Markovian *dynamical systems*.

Experiment design. In regard to experiment design, an illustrative example of the challenges we face is the following. Drug companies want to design a drug with certain properties, and each company has an AI to plan experiments into the efficacy of drugs. There is a plethora of compounds that may be useful, and not all can be tested. However, there exist large scale databases of drug toxicity and activity. Each AI is able to use data from previous clients to do better planning at a lower cost. The AIs can post drug descriptions, *in vitro* results, simulations for *in silico* experiments, and the results of clinical trials.

An illustrative question in the context of this mulit-agent experiment design is how to align incentives so that the joint plan is beneficial to society (in terms of access to useful therapies), while simultaneously balancing the computational and human cost associated with designing, performing and analyzing experiments. An additional challenge relates to privacy– not only in regard to individual's concern about their own data, but in regard to AIs, for example looking to minimize information revealed in order to avoid a "ratchet effect" where other AIs can take advantage of this in the future, in the context of this competitive, market-based mechanism.

A starting point for our research will be [CPS06], who proposed an incentive-aligned mechanism, where self-interested agents report *Gittins indices* on different bandit arms. This needs to be extended to the case where indices cannot be computed exactly. In a setting without incentive design, [LZ10] report a multi-user bandit algorithm which implements the *Lai-Robbins algorithm*, and [KNJ14] describe an *upper confidence bound* (UCB) algorithm where each agent has different preferences over the arms. Finally [MT14] examine the single-agent bandit problem under privacy constraints. This has been extended to a multi-agent framework recently by Tossou and Dimitrakakis [TD15]. As a first step, we plan to investigate whether either the Lai-Robbins or UCB algorithms could replace Gittins indices, and be combined with mechanisms to align incentives.

Multi-agent dynamical systems. In regard to the more general dynamic setting, an illustrative example of the challenges come from AIs for the smart grid, where each AI acts on behalf of a household, and decide when to consumer power, how to allocate power to different devices, and how to control the set-points of devices. The AIs interact with people through preference elicitation (e.g., temperature of house, importance of charging an electric vehicle, importance of dry clothes) and with other agents because of competition for a scare resource with a variable price.

Human-agent collectives hold great promise [JMN$^+$14], but many challenges remain. For example, different human and AI agents make decisions based on a specific belief system and their beliefs will be unlikely to be perfectly aligned. This is firstly because they do not have the same information. Secondly, and perhaps more importantly, it is because they do not have a common prior belief or belief system, perhaps due to differing computational capabilities. How should information about beliefs be priced and communicated? Is it possible to design a general mechanism that near-optimally allocates effort among the different agents, that takes privacy into account?

On the one hand, there exists a significant amount of work in the options and semi-Markov decision process (SMDP) [SPS99] framework. Although this mainly focused in the hierarchical single agent case, [Gha05] extended the framework to co-operative multi-agent problems. However, at the moment the topic of mechanism design has not received much attention. [SCP08] proposed a mechanism to align incentives in agents acting within a decentralized MDP framework, but not with more general SMDP frameworks. In other work, [ASS15] cast multi-agent shortest path problems as combinatorial auctions, and provide a limited analysis of incentives.

In research related to the interaction between AI and people, [VR13] describe an agent that uses a Bayesian approach to shapes the behavior of another agent so that a desirable joint state is reached. The research proposed in this proposal will use the SMDP framework within a non-cooperative multi-agent context, while simultaneously handling incentive and privacy concerns.

3 Student background

The ideal candidate would have a solid grasp of probability theory, and the basics of machine learning, statistics or game theory. In particular, experience with reinforcement learning and familiarity with learning theory are highly desirable.

4 How to apply

Send an email to `christos.dimitrakakis@inria.fr` with the subject: "Lille-3 PhD: Mechanism Design", with the following information:

- A cover letter describing why you are interested in this specific field of research, and why you think you are suitable.
- A curriculum vitae, detailing your education, skills and experience, and at least one reference.
- One or more examples of research work.

References

[ASS15] Ofra Amir, Guni Sharon, and Roni Stern. Multi-agent pathfinding as a combinatorial auction. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015.

[Che59] Herman Chernoff. Sequential design of experiments. *The Annals of Mathematical Statistics*, pages 755–770, 1959.

[CPS06] Ruggiero Cavallo, David C. Parkes, and Satinder Singh. Optimal coordinated planning amongst self-interested agents with private state. In *Proceedings of the 22nd Conference on Uncertainty in Artificial Intelligence (UAI'06)*, pages 55–62, Cambridge, MA, 2006.

[Gha05] Mohammad Ghavamzadeh. *Hierarchical reinforcement learning in continuous state and multi-agent environments.* PhD thesis, University of Massachusetts Amherst, 2005.

[JMN+14] Nicholas R Jennings, Luc Moreau, D Nicholson, S Ramchurn, S Roberts, T Rodden, and Alex Rogers. Human-agent collectives. *Communications of the ACM*, 57(12):80–88, 2014.

[KNJ14] Dileep Kalathil, Naumaan Nayyar, and Rahul Jain. Decentralized learning for multiplayer multiarmed bandits. *Information Theory, IEEE Transactions on*, 60(4):2331–2345, 2014.

[LZ10] Keqin Liu and Qing Zhao. Distributed learning in multi-armed bandit with multiple players. *Signal Processing, IEEE Transactions on*, 58(11):5667–5681, Nov 2010.

[MT14] Nikita Mishra and Abhradeep Thakurta. Private stochastic multi-armed bandits: From theory to practice. In *ICML Workshop on Learning, Security, and Privacy*, 2014.

[Rob85] Herbert Robbins. Some aspects of the sequential design of experiments. In *Herbert Robbins Selected Papers*, pages 169–177. Springer, 1985.

[SCP08] Sven Seuken, Ruggiero Cavallo, and David C. Parkes. Partially synchronized DEC-MDPs in dynamic mechanism design. In *Proc. 23rd AAAI Conference on Artificial Intelligence(AAAI'08)*, pages 162–169, 2008.

[SPS99] Richard S Sutton, Doina Precup, and Satinder Singh. Between mdps and semi-mdps: A framework for temporal abstraction in reinforcement learning. *Artificial intelligence*, 112(1):181–211, 1999.

[TD15] Aristide C. Y. Tossou and Christos Dimitrakakis. Differentially private multi-agent multi-armed bandits. In *EWRL 2015*, 2015.

[VR13] Aris Valtazanos and Subramanian Ramamoorthy. Bayesian interaction shaping: Learning to influence strategic interactions in mixed robotic domains. In *Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems*, pages 63–70. International Foundation for Autonomous Agents and Multiagent Systems, 2013.