Fully funded PhD position - sponsored VISA

Learning Algorithms for the Design of Undeceivable Policies

About the Institut Polytechnique de Paris

The Institut Polytechnique de Paris is a world-class Institute of science and technology, which has contributed to major industrial and technological breakthroughs over the last two centuries. Their alumni include Nobel prize-winners and prominent figures in the worlds of politics, business and research. It is ranked 36th worldwide in Engineering and Computer Science. [IPP]

Team and international collaborators

The PhD student will work in a team composed of:
Assoc. Prof. Andrea Araldo (araldo@telecom-sudparis.eu / Télécom SudParis)
Prof. Moshe Ben-Akiva (Massachusetts Institute of Technology (MIT))
Assoc. Prof. Ravi Seshadri (Technical University of Denmark (DTU))

The possibility to conduct the PhD either in DTU or IP-Paris can be discussed. This is facilitated by the fact that both DTU and IPP are within the Eurotech consortium that facilitates these kinds of settings. In any case, long visiting stays (e.g., 6 months) at either DTU, MIT or IPP are envisaged.

Practical Information

When: starting date is flexible. Duration: 3 years
Requirements: Excellent mathematical modeling and analytical skills, good programming skills (no preference on the language)
To apply: Please send your CV, an explanation of 5 lines explaining why you are the best fit for this position (with factual non-vague or generic elements), all the marks of your BSc and MSc level courses; Sending your ranking is not mandatory (but it is a big plus)

Abstract

We consider a regulator willing to encourage the sustainable behavior of agents, i.e., individuals and businesses. Individuals may be confronted with different options when choosing goods, foods, services and mobility modes. Businesses' choices can be related to production modes, internal organizations, etc. Unfortunately, the choices that maximize the agent's selfish utility do not generally correspond with sustainable choices.

To encourage sustainable choices, the regulator can apply appropriate "signals" to agents, which may be positive (incentives, subsidies) or negative (prices, taxes, bans). Personalized policies adapt the signal to the needs and preferences of each agent.

To implement a personalized policy, the regulator has to learn the preferences of agents by observing their previous choices. Up to now, personalized policies that have been proposed rely on the classic hypothesis of discrete choice modeling, i.e., agents are assumed to be rational and honest, making each choice in order to maximize their utility. This assumption does not hold in the case of personalized policies, where agents may adopt a deceptive behavior, acting in order to hide their true preferences, with the aim to manipulate the regulator and get a more favorable signal than they would deserve.

Fully funded PhD position - sponsored VISA

Our overarching aim is to analyze and enhance the robustness of personalized policies to deceptive agents. We will investigate a possible fundamental trade-off between depth of personalization and robustness to deceptive behavior. We will base our method on Discrete Choice Modeling, Game Theory and Sequential Decision Making. We will validate our findings in stated preference experiments and simulations.

MOTIVATION

To achieve sustainable development, technological advances alone are insufficient. Demand-side mitigation, which consists of favoring changes in the behavior of agents toward sustainable choices [IE21], is also necessary. The greenhouse gasses reduction potential of these initiatives is estimated to exceed 40% by 2050 compared to baselines [IP23], [IP22]. Appropriate signals applied by the regulator are an important example of demand-side mitigation. However, negative signals (e.g., prices, taxes, bans) generally lack public acceptability (to the point of violent protests [IM23]), while positive signals (e.g., incentives, subsidies) are expensive for the regulator and may not be sufficiently strong to induce relevant behavior shifts [Gn00].

Personalized policies can help overcome the above issues [Ar18, Ar23, As21] by adapting signals to the condition and needs of each agent. Signals can target agents where relevant behavioral shift is possible, and at a reasonably low cost for them and the regulator.

However deceptive agents may “trick” personalized policies by making choices aimed to maliciously manipulate the regulator’s signals in their favor. For instance, if the regulator provides incentives to convince frequent car drivers to take public transport, some agents may decide to drive a car for a while, even if it is not their best option, in order to get the incentives.

OBJECTIVES

We aim to understand under which conditions personalized policies are robust to deceptive behavior and how to improve such robustness. Our objectives are to: (i) analyze the extent to which agents can benefit from deceptive behavior, (ii) assess the “cost of deception” suffered by agents to mislead the regulator, and (iii) devise mechanisms to deter deceptive behavior.

NOVELTY

Under classic Random Utility Theory (RUT), agents are assumed to make each choice in order to maximize their utility [Sec.3.1 of Be19]. But RUT cannot model deceptive agents who may make a sequence of suboptimal choices in order to mislead the regulator and get a more favorable signal. To overcome this barrier, we will extend RUT to model agents as sequential decision makers, aiming to maximize their cumulative utility.

Recent work in Game Theory [Gan20,Xu21] studies deceptive agents but assumes they can mislead the regulator “instantaneously”. The novelty of our formulation is that we account for the learning process of the regulator within the game theoretical framework, thereby inferring agent preferences [Be19] based on previous choices. Therefore, we capture the fact that agents may need to make many suboptimal choices in order to manipulate the learning process of the regulator, thus suffering from a loss, which we call “Cost of Deception” (CoD). While deceptive capabilities have been shown to be extremely powerful in the state-of-the-art [Gan20, Xu21], we believe that by considering CoD we can show that personalized policies are much more robust than what has been believed so far.

METHODOLOGY

The project is organized into the following tasks:

Task1: Game Theory formulation

We will formulate a Stackelberg game as in [Gan20,Xu21] with a leader (regulator) and a follower (agent). Instead of following their true utility, a deceptive agent behaves according to a "fake" utility in order to get from the regulator a more favorable signal. We will extend this framework to the case of multiple followers (agents) interacting with each other, resulting in an equilibrium.

Equilibrium among multiple followers has been studied recently in [Ro19, Ta22], but without considering the presence of deceptive agents. We conjecture that the deceptive capability of an agent at equilibrium with others is more limited than when the agent is alone. Indeed, equilibrium constraints the set of fake utilities that are beneficial for the agent. We will verify this conjecture with our formulation.

Task2: Assessment of Cost of Deception

In Task1 we assume that the regulator knows the utility function followed by agents (either the true or the fake one). In reality, the regulator must learn it using statistical methods such as Hierarchical Bayesian Estimators [Be19]. Such methods may be slow and therefore considered a barrier for the regulator. However, we conjecture that this learning process adds robustness to deceptive behavior, since it induces a cost of deception (CoD) for an agent.

We will model the agent as a sequential decision maker willing to maximize their cumulative true utility. Via a Markov Decision Process formulation, we will compute the theoretically optimal agent policy, i.e., the sequence of choice options to select and, accordingly, a sequence of fake utility functions to adopt, in order to maximize the cumulative true utility. At each choice, the CoD will be computed as the difference between the true utility of the best option and the true utility of the option chosen, under the theoretically optimal agent policy. We will verify under which conditions CoD is sufficiently high to deter any deception.

Thanks to "data borrowing" [Ba00] the regulator can base its estimation of the utility of each agent not only on the observed choices of that agent, but also on the other agents with similar socio-economic profiles. Our conjecture is that data borrowing limits the deceptive capabilities of a single agent. We will verify it with our model.

For automatic online advertisements, agents are modeled as sequential decision-makers. The regulator sets prices via bandit-like algorithms [Mo15, Ro13] or based on perfect knowledge of the agent utility [Dr19], having observed millions of transactions. In our case the regulator instead needs to learn agents' preferences before choosing its policy, based on fewer transactions, which completely changes the problem.

Task3: Stated Preference Experiment (SPE)

We will conduct a SPE [Be91] to provide data on attitudes and perceptions towards personalized policies and to determine how people value different attributes and combinations of personalized policies and product or service choices.

Task4: Setup of the simulator.

We will use data and calibration parameters from our previous work [Ar23] to model heterogeneous agents with utility functions obeying the classic random utility framework (Sec.3.1 of [Be19]) as well as a set of options with different characteristics. A hypothetical convex cost function will represent congestion due to different agents selecting a certain option. The choices of all agents will result in a certain sustainability outcome, e.g., CO2 emissions as in [Ar23].

Task5: Mechanisms preventing deceptive behavior

Simple policies have been shown to be more robust than complex policies, as they do not "overfit" to the fake utilities deceptive agents may adopt [Xu21]. We will thus first limit personalized policies to simple shapes, e.g., imposing the signals from the regulator to be linear combinations to some objective quantity, such as energy consumption (as in [Ar18]).
We will then study if the regulator can limit overfitting deceptive behavior by boosting data borrowing (Task2) via increasing, when the regulator estimates the agent utility, the weight given to population-related parameters and decreasing the agent-specific parameters. A drawback of this approach would be a loss in personalization. In the extreme, a non-personalized policy, such as a "random coefficient" [Da19], is extremely robust to deceptive agents but is not able to adapt to the peculiarity of the agents. We will thus study in this task what seems to be a fundamental trade-off between personalization and robustness to deception.

Task6: Result analysis

We will validate our findings in simple synthetic scenarios via the simulator of Task4. In addition, we will adjust the model utility functions in order to match the observations from the stated preference survey conducted in Task3 and reflect the sensitivity to various incentives and elasticity of demand [Be19].

REFERENCES

[Ar18] Araldo, Ben-Akiva et al. "System-level optimization of multi-modal transportation networks" TRR (2019)

[Ar23] Araldo et al., "Personalized Incentives with Constrained Regulator's Budget" TransportmetricaA (2023)

[As21] Ascarza et al. "Eliminating unintended bias in personalized policies using bias-eliminating adapted trees" Proceedings National Academy of Sciences 2022

[Ba00] Orme, "Comparing hierarchical Bayes draws and randomized first choice for conjoint simulations" SawtoothSoftware (2000)

[Be91] Ben-Akiva et al. "Analysis of the reliability of preference ranking data" Journal of Business Research (1991)

[Be19] Ben-Akiva, McFadden, Train. "Foundations of stated preference elicitation: Consumer behavior and choice-based conjoint analysis" Foundations and Trends in Econometrics (2019)

[Da19] Ben-Akiva, et al. "Online discrete choice models: Applications in personalized recommendations" Sec. 4.2, Decision Support Systems (2019)

[Dr19] Drutsa et al., “Optimal Pricing in Repeated Posted-Price Auctions”, NeurIPS (2019)

[Gan20] Birmpas et al. “Optimally deceiving a learning leader in Stackelberg games” NeurIPS (2020)

[Gn00] Gneezy et al. “Pay enough or don't pay at all.” The quarterly journal of economics (2000)

[IE21] “Do we need to change our behaviour to reach net zero by 2050?,” International Energy Agency (2021)

[IM23] “Public Perceptions of Climate Mitigation Policies,” International Monetary Fund (2023)

[IP22] “Demand, services and social aspects of mitigation,” Supplement, IPCC, Sec.5.SM.2 (2022)

[IP23] “Climate Change 2023 Synthesis,” IPCC, Fig.4.4 (2023)

[Mo15] Mohri et al., “Revenue optimization against strategic buyers” NeurIPS (2015)

[Ro13] Rostamizadeh et al., “Learning prices for repeated auctions with strategic buyers” NeurIPS (2013)

[Ro19] Rotemberg. “Equilibrium effects of firm subsidies” American Economic Review (2019)

[Ta22] Wang et al. “Coordinating followers to reach better equilibria” AAAI Conference (2022)

[Xu21] Dawkins et al. “The Limits of Optimal Pricing in the Dark” NeurIPS (2021)