Research Statement

Yingkai Li

Modern online platforms like Google and Meta have access to vast amounts of consumer data, which they use to optimize interactions with strategic agents through various algorithms. A key distinction between designing algorithms for online platforms and traditional algorithms is that in the former, the inputs are provided by strategic agents, making them sensitive to the chosen algorithm. Therefore, economic concepts and methodologies play a vital role in addressing strategic effects in computational problems in this context. Additionally, the information contained in consumer data is critical to both algorithm design for online platforms and the decisions made by platform users. In these applications, the platform has control over data collection and sharing, as well as the employed algorithms. It's important to note that these design aspects must be considered together to achieve optimal social objectives, such as maximizing the platform's revenue or consumer surplus. My research focuses on this interdisciplinary area of computer science and economics, analyzing the joint design of optimal data policies and algorithms for online platforms.

First, we will highlight several applications of the joint design problem and review my past work in those applications. Then, we will delve into a detailed discussion of my future research agendas for each of these directions.

- *Optimal Principal Learning.* In online marketplaces, platforms can selectively choose which type of data to collect from their users. The traditional wisdom suggests that, in the absence of regulation, platforms always benefit from collecting more data, as it enables them to make better-informed business decisions. However, a crucial factor in this scenario is that these platforms typically lack the capacity to commit to how they will employ the collected data in future interactions. The additional data collection can have external effects on users, potentially distorting their behavior and, consequently, negatively impacting the platform's payoffs. For example, Amazon users may be aware of the platform's adoption of big-data discriminatory-pricing algorithms based on past user data like purchase or browsing histories. This added concern may deter users from making purchases on Amazon, which is an undesirable outcome for the platform. In Clark and Li (2023), we address this issue by outlining conditions under which the platform consistently favors gathering more user information and conditions under which the platform strictly benefits from abstaining from collecting certain user data. One primary rationale for the platform's advantage in abstaining from free information collection is to credibly constrain themselves from exploiting the users.

- *Optimal Information Disclosure.* Online platforms with access to large datasets on product qualities and consumer preferences can strategically leverage this data by revealing information through advertisements. These advertisements can influence how agents perceive the values of products, subsequently affecting their purchasing decisions. In the context of single-item auction settings, although optimizing either the information policy or the auction format, when the other is held constant, can be solved in polynomial time, our work in Cai et al. (2023b) reveals that the joint design problem, where both are considered simultaneously, is NP-hard. We then provide a PTAS algorithm for this problem. In multi-item auction settings, the optimal mechanisms, when agents have exogenous information structures, are known to be complex. These mechanisms often involve intricate menus of lotteries, and the menu complexity for achieving approximate optimality can be unbounded, even with only two items for sale (Hart and Nisan, 2019). As part of the ongoing work Cai et al. (2023a), we show that under optimal information disclosure, item pricing is the exact optimal mechanism when there are two items and is approximately optimal when there are at least three items. This finding provides strong support for the practical use of item pricing mechanisms.

- *Costly Information Acquisition.* Previous discussions have centered around applications where platforms have access to consumer data at no cost or negligible costs compared to its value. However, there are also scenarios in which data is not free, and platforms may rely on workers to acquire valuable data at a cost. For example, platforms using machine learning algorithms may depend on workers to collect labeled training data for parameter tuning. The quality of the data significantly impacts algorithm performance, so designers must carefully structure worker compensation to incentivize them to exert effort to obtain high-quality data. We show that the optimal mechanisms in cases where the agent has binary effort choices, deciding to either work or shirk, can be computed efficiently via a polynomial size linear program (Li et al., 2022). However, when the agent's effort choice becomes multi-dimensional, we show that computing the optimal mechanism is NP-hard (Hartline et al., 2023). In both settings, we introduce a simple mechanism that can be computed efficiently and achieves constant approximations to the optimal solution.

Research Agenda

Computation and Complexity. The computational problems in these joint design problems remain largely unresolved. As pointed out by Clark and Li (2023), these problems may not even fall within the class of NP since the space of information structures and mechanisms that the principal needs to jointly design can be unbounded and may not admit simple representations. Moreover, when the principal can jointly design the information structures and the mechanisms, the optimization problem for the principal is inherently non-convex. In fact, as indicated by Cai et al. (2023b), this problem corresponds to an infinite-dimensional quadratic program. Consequently, standard techniques for computing optimal solutions from both information design and mechanism design do not apply in these models. Therefore, it is intriguing to investigate whether there exist PTAS algorithms or polynomial-time constant approximation algorithms for all the general models mentioned above. Below, we list several important computational problems that are worth investigating for the joint design problems.

1. Identify the complexity class for the joint design problems. Do they belong to NP, EXPTIME or anything else?

2. Design FPTAS, PTAS, or constant approximation algorithms for these problems, or prove that it is computationally intractable.

3. Bound the worst-case cardinalities of the signal space. Note that the classic approach of using Carathéodory Theorem does not apply in the joint design problem for bounding the cardinalities (Clark and Li, 2023).

Robustness. The classic work in this literature has primarily focused on ideal Bayesian environments where both the principal and the agents have common knowledge about the uncertainties in the environments. While this assumption allows us to derive novel insights from characterizations of optimal mechanisms, it is unlikely to hold in practical, real-world scenarios. In modern data markets, it is, therefore, of great importance to understand how to design mechanisms when either the principal, the agents, or both of them lack full knowledge of the environments and rely on statistical models or learning algorithms to generate informative predictions.

In cases where the principal has ambiguity over the environments, we aim to extend the framework of benchmark design and prior-independent approximation introduced in Hartline et al. (2020) to our model, where the principal can jointly design information structures and mechanisms.

1. Design polynomial time algorithms for computing the (approximately) optimal prior-independent information structure and mechanisms when the principal does not know the underlying distribution.

2. Identify natural prior-free benchmarks for the joint design problems, design prior-free optimal information structure and mechanisms for the chosen benchmarks, and illustrate their robustness properties.

When the agents have ambiguity over the environments, it is crucial to first formalize the behavior of the agent before analyzing the optimal design for the principal. One plausible approach is to assume that the agent behaves in a manner similar to a statistician. This agent, akin to a statistician, has access to a dataset and employs statistical models to estimate the environments. Given any information structure and mechanism designed by the principal, such an agent would make choices to maximize their worst-case performance based on their statistical estimations.

1. Design polynomial time algorithms for computing the (approximately) optimal information structure and mechanisms when the agents adopt certain family of statistical models.

Simple Mechanisms. The design and analysis of simple mechanisms in complex environments are fundamental challenges within the Econ-CS community. Historically, the focus has been primarily on auction environments (e.g., Feng et al., 2019, 2023), but more recently, attention has shifted toward contract design settings (e.g., Alon et al., 2023). An intriguing question is whether simple information structures or simple mechanisms can offer approximately optimal solutions to the joint design problems. It's important to note that the format of simple mechanisms can vary depending on the context. The primary objective here is to identify natural simple mechanisms for each application and analyze their worst-case performance.

1. In the context of optimal principal learning and optimal advertisement, identify the worst case approximation ratios for fully revealing information structures and fully uninformative information structures.

2. In the context of optimal principal learning, identify the worst-case approximation ratios when focusing on mechanisms that are dominant strategy incentive compatible.

3. In the context of optimal advertisement, identify the worst case approximation for selling items separately or selling via VCG mechanisms when there are multiple agents and multiple buyers.

4. In the context of costly information acquisition, identify the worst case approximation for elicitation without screening when the agent has exogenous private information.

Fairness and Privacy. In current data markets, two significant concerns are fairness across different identity groups when using the data and privacy issues, which safeguard consumers against welfare loss due to information leakage. These concerns introduce additional complexities into the design of optimal mechanisms and information policies for online platforms. For instance, fairness on information sharing differ from traditional work on fair division of physical goods, where information is freely replicable and can be easily provided to all parties without additional costs. In contrast, when sharing information, there are externalities among agents' utilities (see Gan et al., 2023, for an additional discussion on allocation with externalities). Thus, in these applications, it is necessary to propose appropriate notions of fairness for information sharing and then consider the design of optimal information-sharing policies while accommodating fairness constraints. Similarly, for privacy concerns, it is essential to propose appropriate privacy notions and redesign the mechanisms for online platforms while adhering to privacy constraints.

1. Propose proper notion of fairness (e.g., envy-free under swap by Velez, 2016) and design polynomial time algorithms for computing the (approximate) optimal information structure and mechanisms under fairness constraints.

2. Propose proper notion of privacy (e.g., differential privacy by Dwork, 2006) and design polynomial time algorithms for computing the (approximate) optimal information structure and mechanisms under privacy constraints.

References

Alon, T., Duetting, P., Li, Y., and Talgam-Cohen, I. (2023). Bayesian analysis of linear contracts. In *Proceedings of the 24th ACM Conference on Economics and Computation (EC)*, page 66.

Cai, Y., Li, Y., and Wu, J. (2023a). (Approximate) optimality of item pricing under optimal disclosure. *working paper*.

Cai, Y., Li, Y., and Wu, J. (2023b). Computing optimal information and optimal auction. *working paper*.

Clark, D. and Li, Y. (2023). Optimal mechanism design with endogenous principal learning. *working paper*.

Dwork, C. (2006). Differential privacy. In *International colloquium on automata, languages, and programming (ICALP)*, pages 1–12.

Feng, Y., Hartline, J. D., and Li, Y. (2019). Optimal auctions vs. anonymous pricing: Beyond linear utility. In *Proceedings of the 2019 ACM Conference on Economics and Computation (EC)*, pages 885–886. ACM.

Feng, Y., Hartline, J. D., and Li, Y. (2023). Simple mechanisms for non-linear agents. In *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 3802–3816.

Gan, J., Li, B., and Li, Y. (2023). Your college dorm and dormmates: Fair resource sharing with externalities. *Journal of Artificial Intelligence Research*, 77:793–820.

Hart, S. and Nisan, N. (2019). Selling multiple correlated goods: Revenue maximization and menu-size complexity. *Journal of Economic Theory*, 183:991–1029.

Hartline, J., Johnsen, A., and Li, Y. (2020). Benchmark design and prior-independent optimization. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 294–305. IEEE.

Hartline, J. D., Shan, L., Li, Y., and Wu, Y. (2023). Optimal scoring rules for multi-dimensional effort. In *The Thirty Sixth Annual Conference on Learning Theory (COLT)*, pages 2624–2650.

Li, Y., Hartline, J. D., Shan, L., and Wu, Y. (2022). Optimization of scoring rules. In *Proceedings of the 23rd ACM Conference on Economics and Computation (EC)*, pages 988–989.

Velez, R. A. (2016). Fairness and externalities. *Theoretical Economics*, 11(1):381–410.