Mechanism Learning with Mechanism Induced Data

Tie-Yan Liu
Microsoft Research
tyliu@microsoft.com

Wei Chen
Microsoft Research
wche@microsoft.com

Tao Qin
Microsoft Research
taoqin@microsoft.com

Abstract

Machine learning and game theory are two important directions of AI. The former usually assumes data is independent of the models to be learned; the latter usually assumes agents are fully rational. In many modern Internet applications, like sponsored search and crowdsourcing, the two basic assumptions are violated and new challenges are posed to both machine learning and game theory. To better model and study such applications, we need to go beyond conventional machine learning and game theory (mechanism design), and adopt a new approach called *mechanism learning with mechanism induced data*. Specifically, we propose to learn a behavior model from data to describe how "real" agents play the complicated game, instead of making the full-rationality assumption. Then we propose to optimize the mechanism by using the learned behavior models to predict the future behaviors of agents in response to the new mechanism. Because the above process couples mechanism learning and behavior learning in a loop, new algorithms and theories are needed to perform the task and guarantee the asymptotical performance. As shown in this paper, there are many interesting research topics along this direction, many of which are still open problems, waiting for researchers in our community to deeply investigate.

1 Introduction

Machine learning and game theory are two important branches of AI. Both directions have been well developed in the past decades. Machine learning mainly focuses on learning models from data, based on the assumption that data are generated independent of the model, and game theory mainly focuses on equilibrium analysis and mechanism design, usually based on the assumption that strategic agents have fully rational behaviors.

In this Internet era, new applications like sponsored search, crowdsourcing, and app store create brand new monetization channels for companies, and also change the daily life of every individual. In these applications, a huge number of human agents (e.g., advertisers in sponsored search, workers in crowdsourcing, and app developers in app store) interact with each other governed by a complicated (and possible evolving) mechanism. On one hand, because of human nature, these agents behave strategically to be better off, and thus generate huge data that are dependent of the mechanism. On the other hand, their behaviors are far from fully rational due to their constrained capability and limited information access. As a result, neither conventional machine learning nor game theory can model these applications very well and provide us the right tool to effectively optimize their mechanisms. We simply need to develop new frameworks and theory for this purpose. This is exactly the motivation of our paper.

1.1 Motivating Applications

As mentioned above, many Internet applications including sponsored search, crowdsourcing, and app store can be regarded as dynamic systems that involve multi-party interactions. *Users* arrive at the system with their particular needs; *agents* (who compete or collaborate with each other) provide products or services that could potentially satisfy users' needs; and the *platform* employs a complicated mechanism (e.g., an auction) to match agents with users and extract revenue from this procedure. Afterwards, the platform may provide agents with certain signals as their performance indicator (which we usually call Key Performance Indicators, or KPIs for short). In order to be better off during this economic process, self-interested agents may strategically adjust their behaviors (e.g., strategically reveal the information about their services or products) in response to the mechanism (or more accurately to the signals they receive from the platform since the mechanism is usually invisible to them). Specifically, in a sponsored search, platform, users, and agents correspond to the search engine, search users, and advertisers respectively; in crowdsourcing, platform, users, and agents correspond to crowdsourcing platform (e.g., Amazon Mechanical Turk), employers, and workers respectively; and in app stores, platform, users and agents correspond to app store, app users, and app developers respectively.

Driven by the crowd of strategic agents and the power of the Internet, the abovementioned applications exhibit their large scale, fast pace, and high dynamics. In such situations, while agents are actively adjusting their behaviors to be better off, they could never perform in a perfectly rational manner due to limited information access and constrained capability. For example, in sponsored search (Qin, Chen, and Liu 2014), the advertisers receive KPI for their ads, and adjust their bids accordingly. The KPIs, and therefore the response to the KPI, depend on both the bids of other advertisers and the auction mechanism (Athey and Nekipelov 2010; Pin and Key. 2011). However, since the KPIs only reveal partial information of the system (i.e., the advertisers cannot know others' exact bids and the parameter in the mechanism, and can hardly learn them in an effective way give the large number of competitors in the auction, the complex broad match between queries and keywords, and the fast-paced auction process driven by billions of queries per day), it is almost impossible for advertisers' to make their behavior changes a best response (Duong and Lahaie 2011; Rong, Qin, and An 2014). As a result, the assumptions used in both conventional machine learning and game theory become rather fragile and cannot be used to explain real behaviors of the agents in these applications of our interest.

1.2 Our Proposal

To tackle aforementioned challenge, we propose a new research direction called *mechanism learning with mechanism induced data (MLMID)*. The key of this proposal is as follows:

(1) Instead of assuming the agent behaviors to be fully rational, we propose to learn a more real model from data to describe their behaviors. We call such models *data-driven behavior models*. For instance, we can employ a Markov model, which depends on much weaker assumptions, i.e., agents behaviors only depend on their previous actions and accessible information in a finite period of history.

(2) Instead of assuming the behavior data to be independent of the mechanism, we regard the data as being induced by the mechanism, and propose to use a bilevel framework to learn the optimal mechanism. That is, we first use the behavior model learned from historical data to predict agents' future behaviors in response to a given mechanism, and then learn the optimal mechanism on its basis.

(3) We analyze the theoretical properties of the above learning process by considering the loop formed by behavior learning and mechanism learning. In particular, we may need to defined a nested capacity measure that consider both the behavior model space and the mechanism space. We also need to characterize the generalization error with respect to both the number of training instances for behavior learning, and the number of loops for optimizing the mechanism based on the predicted behaviors.

Regarding the above proposal, there are a number of interesting open questions to explore. For example, how to collect appropriate training data and test data? How to define meaningful empirical loss and expected loss? How to effectively learn a data-driven behavior model? How to learn the mechanism given that it is coupled with a data-driven behavior model? How to define appropriate surrogate loss function for efficient optimization? How to ensure the generalization ability of the learned mechanism learned? Which solution (e.g., equilibrium) concept shall we use while conducting game theoretical analysis for the learned mechanism?

We truly believe that by answering the above questions, we can gain deep understanding of the modern Internet applications, and obtain many new and interesting research topics. We show the big picture of MLMID in Figure 1, and will make detailed discussions in the remaining parts of this paper.

2 Related Work

As can be seen from the above descriptions, our proposal leverages both machine learning and game theory, and let them interact with each other. We have noticed that in many previous works, machine learning also meets game theory. However, some of these works still assume the i.i.d. data generation and/or full-rationality; some do not leverage data at all, and some others are not concerned with mechanism design. Therefore, they cannot model the MLMID problem of our interest as well as our proposal. In this regard, we consider our proposal as a good complement to the existing literature of machine learning plus game theory, which establishes a stronger connection to real applications.

Here we briefly discuss the relationship between our proposal and some related works.

Learning in games (Fudenberg 1998; Mossel and Roch 2007; Mannor, Shamma, and Arslan 2007; Yao, Chen, and Liu 2012) studies how an agent optimizes his/her behaviors through learning under the assumption that the mechanism of a game is fixed and does not depend on agents' behaviors. This is clearly different from our problem, since we want to learn the mechanism based on agents' behavior data.

Mechanism design via machine learning (Dütting et al. 2012; Balcan et al. 2005) reduces the optimization problem for mechanism design to existing machine learning techniques, e.g., support vector machines. It differs from our problem in two aspects: (1) It assumes the full rationality of agents' behaviors and targets at designing incentive compatible mechanisms. (2) It does not leverage data for learning.

Machine learning for mechanism design also targets at mechanism design by learning with data. However, it is different from our problem in the following aspect: (1) Some works (Babaioff, Sharma, and Slivkins 2009; Babaioff, Kleinberg, and Slivkins 2010; Cole and Roughgarden 2014)assume the full rationality of agents' behaviors and targets at designing incentive compatible mechanisms. (2) Some works (Xu, Qin, and Liu 2013; He et al. 2013b; Hummel and McAfee 2014; Cole and Roughgarden 2014; Yang et al. 2014) assume the data to be i.i.d., while in our problem data is induced by the mechanism.

Game theory for machine learning, including strategyproof classification (Meir, Procaccia, and Rosenschein 2012) and incentive compatible regression (Dekel, Fischer, and Procaccia 2008), focuses on how to make machine learning algorithms incentive compatible. In our problem, we do not consider incentive compatibility since agents are not fully rational.

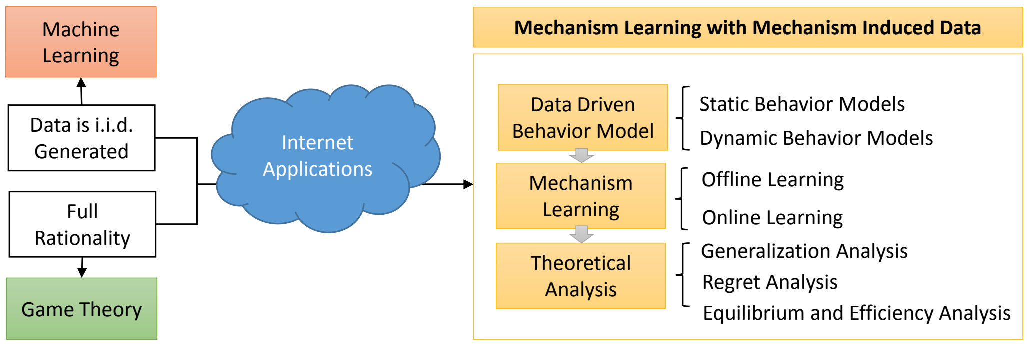

Figure 1: Mechanism Learning with Mechanism Induced Data

3 Mechanism Learning

The focus of mechanism learning is to adopt new machine learning techniques to optimize the mechanism of the platform, based on the learned data-driven behavior models of the agents. In this section, we make detailed elaborations on major research topics in this new direction.

3.1 Data Driven Behavior Models

In this subsection, we discuss the data-driven behavior models and how to effectively learn their parameters.

Static behavior models. In this kind of models, the behaviors of an agent can be described by a static parametric function whose mathematical form is fixed and does not change over time. The mathematical form might be just a regression function, or a parameterized version of some existing behavior models studied in the literature (Duong and Lahaie 2011). The input to the behavior function include all kinds of information that the agent can access. The output of the behavior function is the behavior change. Please note that adopting static behavior models does not mean that the behaviors are static, because the input to this "static" function may be dynamic. Given the data generated by the agents in a certain application, we can learn the parameters of the behavior function and further use them to predict the future behaviors of the agents. Actually, there have been some meaningful attempts on learning static behavior models in the recent literature. For example, in (Tian et al. 2014) and (He et al. 2013a), it is assumed that the model is Markovian, and its transition probability is a truncated Gaussian function, then the parameters of the model are learned by means of maximum likelihood estimation. In (Tian et al. 2014), generalization analysis is conducted for the model and conditions are given that can guarantee the generalization ability of the model. In (Xu et al. 2013), the best response model is parameterized by introducing the willingness, capability, and constraints of the agents, and then the parameters of the model is learned also by means of maximum likelihood estimation.

Dynamic behavior models. In this kind of models, there might not exist a static function that describe how agents change their behaviors. Instead, the assumption is that agents will perform online learning when interacting with the platform, users, and other agents. Therefore, the behavior models of the agents will be dynamic and evolve over time. For this line of research, the related attempts are very limited. Because of this, there is actually quite a large space to explore. For example, since there are both stochastic factors (e.g., users and other uncertain factors in the ecosystem) and adversarial factors (e.g., both the behaviors of other agents and the mechanism of the platform may change strategically), conventional online learning algorithms (e.g., multi-armed bandits) might not suffice. New methodologies and theoretical frameworks might be needed.

3.2 Mechanism Learning based on Learned Behavior Model

To perform the task of mechanism learning, one could collect historical data including the agent behavior data and the user data in advance, and then optimize the mechanism on these data. In (Zhu et al. 2009; Cui et al. 2011; Mohri and Medina 2013), the authors assume agents' behaviors or the behavior distribution will not change when the mechanism changes, and apply machine learning techniques to learn the optimal algorithm. However, as mentioned before, the i.i.d. assumption does not hold in practice, and when we update the mechanism during the learning process, the agent behavior data we collected in advance will become obsolete.

To tackle the problem, we propose to use the collected agent behavior data just for the training of the data-driven behavior model, and then use this model to generate new behavior data for the agents during the process of mechanism learning. One existing attempt along this direction is (He et al. 2013a). They propose a new framework that involves a bi-level empirical risk minimization (ERM): It first learns a Markov behavior model to characterize how agents change their behaviors, and then optimizes the mechanism by simulating agents' behavior changes in response to the mechanism based on the learned Markov model. Although this framework has demonstrated promising empirical results, its optimization is somehow inefficient due to the complex process of simulating new behavior data and the non-smooth objective function. It remains an open question how to find appropriate surrogate objective function and design efficient optimization algorithms for this bilevel ERM framework.

Furthermore, there is one assumption behind the aforementioned approach, i.e., the collected behavior data is comprehensive enough and the behavior model learned from it can be valid for any new mechanism. However, sometimes this assumption might not hold (especially considering that agents may perform online learning and their behaviors can be very dynamic and complicated). In this situation, it would be better to adopt the online version of mechanism learning. Specially, we collect our desired behavior data of the agents in a progressive and adaptive manner within a real application, and probe the mechanisms by minimizing the regret on the collected data. With this online setting, we have the power to change the mechanism of the system and impact the data generalization process. To our knowledge, there is no formal attempt along this direction yet, and there are many related questions to be answered, e.g., how to collect behavior data and probe different mechanisms? And how to combine static/dynamic behavior models with online mechanism learning?

3.3 Theoretic Analysis for Mechanism Learning

The theoretical study of mechanism learning involves two perspectives and three subtopics. From the perspective of machine learning, we can conduct (1) generalization analysis for offline mechanism learning and (2) regret analysis for online mechanism learning; from the perspective of game theory, we can conduct (3) equilibrium and efficiency analysis for such a dynamic system in which the mechanism is optimized based on the agents' behavior models and the agents' behavior models are learned from non-i.i.d. data.

Generalization analysis for offline mechanism learning is highly non-trivial because of the loop between behavior data and mechanism. Very recently, the first generalization analysis regarding MLMID is conducted in (Li et al. 2015), which decomposes the overall generalization error into the behavior learning error and the mechanism learning error. The former relates to the process of learning a Markov behavior model from data, and the latter relates to the process of learning the optimal mechanism based on the learned behavior model. For the behavior learning error, a non-asymptotic error bound is obtained by considering the gap between transition frequency and transition probability of a Markov chain. For the mechanism learning error, a new concept called nested covering number of the mechanism space is used to obtain an uniform convergence bound. Specifically, the mechanism space is first partitioned into subspaces (i.e., a cover) according to the similarity between the stationary distributions of the data induced by mechanisms. In each subspace, the data distribution is similar and therefore one can substitute the data sample associated with each mechanism by a common sample without affecting the expected risk by much. Second, for each mechanism subspace, a uniform convergence bound is derived based on its covering number (Anthony and Bartlett 2009) by using the generalization analysis techniques developed for mixing sequences. While the above analysis takes a meaningful step forward, it is still very preliminary and a lot of works need to be further done. For example, according to the analysis, the generalization guarantee heavily relies on an algorithmic trick introduced by (He et al. 2013a), called δ-sample sharing. This makes the analysis rather narrow, and may not be able to explain other types of algorithms for MLMID.

Regret analysis for online mechanism learning is very different from the analysis for most existing online learning tasks. Comparing with classical online learning task with stochastic environment (Shalev-Shwartz 2011), online mechanism learning involves much more complex agents behaviors; comparing with adversarial online learning which does not utilize any information about agents behaviors (Cesa-Bianchi and Lugosi 2006), the data-driven behavior model in our proposal should help us obtain results with more insights. Thus, regret analysis for online mechanism learning is interesting and promising, and may be very critical for the algorithm design for online mechanism learning. To our knowledge, this is still unexplored space.

Equilibrium and efficiency analysis is a difficult task for a dynamic system in which the mechanism is learned from mechanism induced data, the data is not i.i.d., and the agents' behaviors are not fully rational. It is known that when the learned behavior model takes certain forms such as the quantal response model, although without the full rationality assumption, we can still get some meaningful results regarding the equilibrium of some specific game with fixed mechanisms such as sponsored search auctions (Rong, Qin, and An 2014). However, for other kinds of behavior models learned from data, e.g., the one based on the truncated Gaussian function (He et al. 2013a), and the mechanisms learned based on the behavior models, we do not have any clear understanding of their corresponding equilibria. Many questions remain open and need to answer. For example, which solution concept shall we use to analyze the equilibrium of a dynamic system in which the mechanism and agents' behavior models are dependent and both evolving over time? Shall we define some new equilibrium concepts without the full rationality assumption on agents' behaviors? If there are multiple equilibria, how does the mechanism perform in terms of social welfare in the best-case and the worst-case equilibria? And how about the average case?

4 Conclusion

Motivated by the interaction between agents and the mechanism in many Internet applications, we describe a new research direction, called mechanism learning with mechanism induced data, in which we need to learn agents' behaviors based on historical data induced from the mechanism and then optimize/learn the mechanism based on the learned behavior model. Machine learning and game theory methodologies meet each other in this new direction, and both need to be refined.

In our opinion, the new direction has both big research potential to enlarge the scope of AI and practical impact to help us better understand modern Internet applications. We hope that this paper will draw more attention from the community and attract more researchers to get into this direction.

References

Athey, S., and Nekipelov, D. 2010. A structural model of sponsored search advertising auctions. In *Sixth ad auctions workshop*.

Babaioff, M.; Kleinberg, R. D.; and Slivkins, A. 2010. Truthful mechanisms with implicit payment computation. In *Proceedings of the 11th ACM conference on Electronic commerce*, 43–52. ACM.

Babaioff, M.; Sharma, Y.; and Slivkins, A. 2009. Characterizing truthful multi-armed bandit mechanisms. In *Proceedings of the 10th ACM conference on Electronic commerce*, 79–88. ACM.

Balcan, M.-F.; Blum, A.; Hartline, J. D.; and Mansour, Y. 2005. Mechanism design via machine learning. In *Foundations of Computer Science, 2005. FOCS 2005. 46th Annual IEEE Symposium on*, 605–614. IEEE.

Cesa-Bianchi, N., and Lugosi, G. 2006. *Prediction, learning, and games*, volume 1. Cambridge University Press Cambridge.

Cole, R., and Roughgarden, T. 2014. The sample complexity of revenue maximization. In *ACM Symp. on Theory of Computing. To appear*.

Cui, Y.; Zhang, R.; Li, W.; and Mao., J. 2011. Bid landscape forecasting in online ad exchange marketplace. In *KDD '11 Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining.* ACM Press.

Dekel, O.; Fischer, F.; and Procaccia, A. D. 2008. Incentive compatible regression learning. In *Proceedings of the nineteenth annual ACM-SIAM symposium on Discrete algorithms*, 884–893. Society for Industrial and Applied Mathematics.

Duong, Q., and Lahaie, S. 2011. Discrete choice models of bidder behavior in sponsored search. In *Internet and Network Economics*. Springer. 134–145.

Dütting, P.; Fischer, F.; Jirapinyo, P.; Lai, J. K.; Lubin, B.; and Parkes, D. C. 2012. Payment rules through discriminant-based classifiers. In *Proceedings of the 13th ACM Conference on Electronic Commerce*, 477–494. ACM.

Fudenberg, D. 1998. *The theory of learning in games*, volume 2. MIT press.

He, D.; Chen, W.; Wang, L.; and Liu, T.-Y. 2013a. A game-theoretic machine learning approach for revenue maximization in sponsored search. In *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence*.

He, D.; Chen, W.; Wang, L.; and Liu, T.-Y. 2013b. Online learning for auction mechanism in bandit setting. *Decision Support Systems* 56:379–386.

Hummel, P., and McAfee, P. 2014. Machine learning in an auction environment. In *Proceedings of the 23rd international conference on World wide web*, 7–18. International World Wide Web Conferences Steering Committee.

Li, H.; Tian, F.; Chen, W.; Qin, T.; and Liu, T.-Y. 2015. Generalization analysis for game-theoretic machine learning. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*.

Mannor, S.; Shamma, J. S.; and Arslan, G. 2007. Online calibrated forecasts: Memory efficiency versus universality for learning in games. *Machine Learning* 67(1-2):77–115.

Meir, R.; Procaccia, A. D.; and Rosenschein, J. S. 2012. Algorithms for strategyproof classification. *Artificial Intelligence* 186:123–156.

Mohri, M., and Medina, A. M. 2013. Learning theory and algorithms for revenue optimization in second-price auctions with reserve. *arXiv preprint arXiv:1310.5665*.

Mossel, E., and Roch, S. 2007. Slow emergence of cooperation for win-stay lose-shift on trees. *Machine Learning* 67(1-2):7–22.

Pin, F., and Key., P. 2011. Stochasitic variability in sponsored search auctions: Observations and models. In *EC '11 Proceedings of the 12th ACM conference on Electronic commerce.* ACM Press.

Qin, T.; Chen, W.; and Liu, T.-Y. 2014. Sponsored search auctions: Recent advances and future directions. *ACM Transactions on Intelligent Systems and Technology (TIST), to appear*.

Rong, J.; Qin, T.; and An, B. 2014. Quantal response equilibrium for sponsored search auctions: Computation and inference. In *Tenth ad auction workshop*.

Shalev-Shwartz, S. 2011. Online learning and online convex optimization. *Foundations and Trends in Machine Learning* 4(2):107–194.

Tian, F.; Li, H.; Chen, W.; Qin, T.; Chen, E.; and Liu, T.-Y. 2014. Agent behavior prediction and its generalization analysis. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*.

Xu, H.; Gao, B.; Yang, D.; and Liu, T.-Y. 2013. Predicting advertiser bidding behaviors in sponsored search by rationality modeling. In *Proceedings of the 22nd international conference on World Wide Web*, 1433–1444. International World Wide Web Conferences Steering Committee.

Xu, M.; Qin, T.; and Liu, T.-Y. 2013. Estimation bias in multi-armed bandit algorithms for search advertising. In *Advances in Neural Information Processing Systems*, 2400–2408.

Yang, R.; Ford, B.; Tambe, M.; and Lemieux, A. 2014. Adaptive resource allocation for wildlife protection against illegal poachers. In *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*, 453–460. International Foundation for Autonomous Agents and Multiagent Systems.

Yao, L.; Chen, W.; and Liu, T.-Y. 2012. Convergence analysis for weighted joint strategy fictitious play in generalized second price auction. In *Internet and Network Economics*. Springer. 489–495.

Zhu, Y.; Wang, G.; Yang, J.; Wang, D.; Yan, J.; Hu, J.; and Chen, Z. 2009. Optimizing search engine revenue in sponsored search. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, 588–595. ACM.