Generalization Analysis for Game-Theoretic Machine Learning

Anonymous Author(s)
Affiliation
Address
email

Abstract

For Internet applications like sponsored search, cautions need to be taken when using machine learning to optimize their mechanisms (e.g., auction) since self-interested agents in these applications may change their behaviors (and thus the data distribution) in response to the mechanisms. To tackle this problem, a framework called game-theoretic machine learning (GTML) was recently proposed, which first learns a Markov behavior model to characterize agents behaviors, and then learns the optimal mechanism by simulating agents' behavior changes in response to the mechanism. While GTML has demonstrated practical success, its generalization analysis is challenging because the behavior data are non-i.i.d. (dependent on the mechanism). To address this challenge, we decompose the generalization error for GTML into the behavior learning error and the mechanism learning error. To characterize the first error, we derive a new upper bound for the gap between transition frequency and transition probability of a Markov chain, and then apply the Hoeffding inequality for Markov chains. To characterize the second error, we derive a uniform convergence bound based on the so-called *nested covering number* of the mechanism space and the generalization analysis techniques developed for mixing sequences. To the best of our knowledge, this is the first work on the generalization analysis of GTML, and we believe it has general implications to the theoretical analysis of other complicated machine learning problems.

1 Introduction

Many Internet applications, such as sponsored search and crowdsourcing, can be regarded as dynamic systems that involve multi-party interactions. Specifically, *users* arrive at the system at random with their particular needs; *agents* report the information about their products or services that could potentially satisfy users' needs; and the *platform* employs a mechanism to match agents with users and extract revenue from this process. Afterwards, users may give feedback to the platform about their satisfactions; and the platform may provide agents with some signals as their performance indicator. Since both the information reported by the agents and the mechanism will affect the payoff of the agents, self-interested agents may strategically adjust their reporting behaviors in response to the mechanism (or more accurately the signals they receive since the mechanism is invisible to them). Take sponsored search as an example. When a user submits a query to the search engine (the platform), the search engine runs an auction to determine a ranked list of ads based on the bid prices reported by the advertisers (the agents). If the user clicks on (gives feedback to) an ad, the corresponding advertiser will be charged by a certain amount of money. After a few rounds of auctions, the search engine will provide the advertisers with some signals on the auction outcome, e.g., the average rank positions of their ads, the numbers of clicks, and the total payments. Based on such signals, the advertisers may adjust their bidding behaviors to be better off in the future.

It is clear that the mechanism plays a central role in these Internet applications. It determines the satisfaction of the users, the payoffs of the agents, and the revenue of the platform. Therefore, how to optimize the mechanism becomes an important research topic. A number of research works have used machine learning to optimize the mechanism, including [1] [2] [3] [4] [5] [6] [7]. These works could be categorized into three types.

First, some researchers assume that the agents are fully rational and investigate the Nash (or dominant-strategy) equilibrium of the mechanism. For example, [6] proposes a machine learning framework to optimize the second-price auction in sponsored search. In this case, the dominant strategy for fully rational advertisers is to truthfully reveal their valuations through the bid prices and therefore their bidding behaviors have no dynamics. Second, some researchers assume that the behaviors of the agents are i.i.d. and independent of the mechanism, and then make use of the historical behavior data to optimize future mechanisms. For example, [3] and [5] apply machine learning algorithms to optimize the auction mechanism based on the historical bidding data of the advertisers. Third, some other researchers believe that the behaviors of the agents are neither fully rational nor i.i.d., instead, they are dependent on the mechanism through a data-driven Markov model. In particular, [8] and [9] assume that the behavior change of the agents depends on their historical behaviors and received signals in a finite number of previous time periods, and they propose learning a Markov model from data to characterize this dependency. Please note that the assumption in the third type of works is more general, and can cover the other two types as its special cases. According to [10], Nash (and also dominant-strategy) equilibrium in many games can be achieved by best-response behaviors, with which an agent determines the next action by maximizing his/her payoff based on the current action profile and mechanism. It is clear that the best-response behaviors are Markovian. Furthermore, it is also clear that the i.i.d. behaviors are special cases of Markov behaviors.

Based on the Markov assumption on agent behaviors, [7][8] propose a new framework for mechanism optimization, called game-theoretic machine learning (GTML). The GTML framework involves a bi-level empirical risk minimization (ERM): it first learns a Markov 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. The GTML framework has demonstrated promising empirical results, however, its generalization analysis is very challenging as compared to conventional machine learning approaches: while conventional machine learning assumes an unknown but fixed data distribution, in GTML, the data distribution may dynamically change in response to the mechanism. As a result, conventional generalization analysis techniques could not be directly applied.

In this paper, we present our formal analysis on the generalization ability of GTML. Specifically, we propose decomposing the generalization error for GTML into the behavior learning error and the mechanism learning error. The former relates to the process of learning a Markov behavior model from data; the latter relates to the process of learning the optimal mechanism based on the learned behavior model. For the former error, we derive a new upper bound for the gap between transition frequency and transition probability of a Markov chain, and then apply the Hoeffding inequality for Markov chains. For the latter error, we make use of the so-called *nested covering number* of the mechanism space. Specifically, we first partition the mechanism space into subspaces (i.e., a cover) according to the similarity between the stationary data distributions 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 revenue by much. Second, for each mechanism subspace, we derive a uniform convergence bound using its covering number and the generalization analysis techniques developed for mixing sequences. Finally we take generalized second price (GSP) auctions with reserve prices [6][11][12] as an example to show how the above bound can be concretely computed.

To the best of our knowledge, this is the first work that performs formal generalization analysis on GTML, and we believe the methodologies we use have their general implications to the theoretical analysis of other complicated machine learning problems as well.

2 Game-Theoretic Machine Learning Framework

In this section, we give a description of the game-theoretic machine learning (GTML) framework.

2.1 Mechanisms in Internet Applications

Internet applications such as sponsored search and crowdsourcing can be regarded as dynamic systems involving interactions between multiple parties, e.g., users, agents, and platform. For example, in sponsored search, the search engine (platform) ranks and shows ads to users and charges the advertisers (agents) if their ads are clicked by users, based on the relevance degrees of the ads and the bid prices reported by the advertisers. In crowdsourcing, companies like Mechanical Turk (platform) match workers (agents) to employers (users) and pay the workers for their efforts and extract revenue from this process, based on the quality of their job completion (reviewed by the employers) and the demanded salaries reported by the workers. In these examples, users' needs/feedback (e.g., queries, clicks, job, review) can be regarded as i.i.d. samples, however, agents' behaviors (e.g., reported bid prices and demanded salaries) may be strategic. This is because agents usually have clear utilities in their minds, and they may change behaviors in order to maximize their utilities given the understandings on the mechanism used by the platform. As a result, the distribution of the agents' behaviors might be dependent on the mechanism, instead of being i.i.d. sampled.

Mathematically, we denote the space of mechanisms as $\mathcal{A}$, and assume it to be a compact metric space with distance $d_{\mathcal{A}}$. We denote the space of user need/feedback as $\mathcal{U}$ and the space of agent behaviors as $\mathcal{B}$. For a mechanism $a \in \mathcal{A}$, at the t-th time period, agents' behavior profile is $b_t^a \in \mathcal{B}$, and a user $u_t \in \mathcal{U}$ arrives at the system. The platform matches the agents to the user and charges them according to mechanism a. After that, the platform will provide some signals $h_t \in \mathcal{H}$ (e.g., the number of clicks on the ads and the review on the quality of job completion) to the agents as an indication of their performances. Since h_t may be affected by agents' behavior profile b_t^a, mechanism a, and user data u_t, we denote $h_t = sig(a, b_t^a, u_t)$, where $sig : \mathcal{A} \times \mathcal{B} \times \mathcal{U} \to \mathcal{H}$ is a function generating the signals for agents. After observing h_t, agents will change their behavior to b_{t+1}^a to be better off in the future.

2.2 Markov Agent Behavior Model

In order to describe how agents change their behaviors, the authors of [8] and [9] proposed a Markov behavior model. The key assumption made by the Markov model is that any agent only has a limited memory, and his/her behavior change only depends on his/her previous behaviors and signals in a finite number of time periods. To ease the discussion and without loss of too much generality, we assume the behavior model to be first-order Markovian. Formally, given the signal h_t, the distribution of agents' next behavior profile can be written as follows,

$$P(b_{t+1}^a = \cdot | b_t^a, ..., b_1^a; u_t, ..., u_1) = P(b_{t+1}^a = \cdot | b_t^a, ..., b_1^a; h_t, ..., h_1) = P(b_{t+1}^a | b_t^a, h_t) := M_{h_t}(b_t^a, b_{t+1}^a),$$

where M_h is the transition probability matrix of the behavior profile, given the signals $h \in \mathcal{H}$.

As mentioned in the introduction, the Markov behavior model is very general and can cover other types of behavior models studied in the literature. The Markov behavior model can cover the i.i.d. behaviors by simply setting the transition probability to be a fixed distribution independent of the signals and the previous behaviors. It can also cover the best-response behaviors by setting the transition probability to be a delta distribution whose support is the behavior with the largest payoff computed from the previous behavior profile and mechanism.

2.3 Bi-Level Empirical Risk Minimization

In this subsection, we introduce the bi-level empirical risk minimization (ERM) algorithm proposed in [7]. The first-level ERM corresponds to behavior learning, i.e., learning the Markov behavior model (the transition probability matrixes $M_h(\cdot, \cdot), h \in \mathcal{H}$) from training data containing signals and corresponding behavior changes. The second-level ERM corresponds to mechanism learning, i.e., learning the mechanism with the minimum empirical risk defined with both the behavior model learned at the first level and the training data containing users' needs/feedback.

For behavior learning, suppose we have T_1 samples of historical behaviors and signals $\{b_t, h_t\}_{t=1}^{T_1}$. The goal is to learn the transition matrix $M_h(\cdot, \cdot)$ from these data. In [7] and [9], both parametric and non-parametric approaches were adopted for behavior learning. With the parametric approach, one assumes the transition probability to take a certain mathematical form, e.g., $M_h(b, b') \propto \exp(-(b' - \langle w, (b, h)\rangle))$, where $\langle \cdot, \cdot \rangle$ denotes the inner product of two vectors and parameter w is learned by minimizing the empirical likelihood loss. With the non-parametric approach, one directly estimates

each entry $M_h(b, b')$ by counting the frequencies of the event $(b_t = b, b_{t+1} = b')$ out of the event $(b_t = b)$ given signal h. No matter which approach is used, we denote the learned behavior model as $\hat{M}_{T_1}$ for ease of reference.

For mechanism learning, suppose we have T_2 samples of user data $\{u_t\}_{t=1}^{T_2}$ and a Markov behavior model $\hat{M}_{T_1}$, learned as above. The goal is to learn an optimal mechanism to minimize the empirical risk (or equivalently, maximize the empirical revenue) on the user data, denoted as $L(a, b, u)$. For this purpose, for each mechanism in the mechanism space, one generates T_2 samples of behavior data $\{b_t^a\}_{t=1}^{T_2}$ in a sequential manner using the Markov model $\hat{M}_{T_1}$. Together with the T_2 samples of user data, the empirical risk can be computed. To improve the computational efficiency of mechanism learning, in [7], the authors introduce a technique called δ-sample sharing. Specifically, given $\delta > 0$, in the optimization process, if the distance between a new mechanism a and another mechanism a' whose behavior data is already generated is smaller than δ (i.e., $d_{\mathcal{A}}(a, a') \leq \delta$), then one will not generate behavior data for a any more, but instead reuse the behavior data previously generated for mechanism a'. Therefore, we denote the sample for mechanism a as $\{b_t^{s(a,\delta)}\}_{t=1}^{T_2}$, where $s(a, \delta)$ is equal to a itself or another mechanism satisfying $d_{\mathcal{A}}(a, s(a, \delta)) \leq \delta$. Consequently, the empirical risk can be represented as $\frac{1}{T_2}\sum_{t=1}^{T_2} L(a, b_t^{s(a,\delta)}, u_t)$. For ease of reference, we denote this empirical risk as $\mathcal{R}_{T_2}(a, \hat{M}_{T_1}, \delta)$. By minimizing $\mathcal{R}_{T_2}(a, \hat{M}_{T_1}, \delta)$, one can obtain an empirically optimal mechanism:

$$\hat{a}_{T_2} = \arg\min_{a \in \mathcal{A}} \mathcal{R}_{T_2}(a, \hat{M}_{T_1}, \delta).$$

While GTML and the bi-level ERM algorithm have demonstrated their practical success [8], their theoretical properties are not yet well understood. In particular, given that GTML is more complicated than conventional machine learning (in GTML the behavior data are dependent on the mechanism), conventional generalization analysis techniques cannot be directly applied and new methodologies need to be proposed.

3 Generalization Analysis for Game-Theoretic Machine Learning

In this section, we first give a formal definition to the generalization error of the bi-level ERM algorithm for GTML, and discuss how to derive a meaningful upper bound for this generation error.

According to [9], for a behavior model M (including the true Markov behavior model M^* and the model $\hat{M}_{T_1}$ obtained by the behavior learning algorithm), under some mild conditions (e.g., M is irreducible and aperiodic), the process (b_t^a, u_t) is a uniformly ergodic Markov chain for arbitrary mechanism a. Then given mechanism a and behavior model M, there exists a stationary distribution for (b_t^a, u_t), which we denote as $\pi(a, M)$. We define the risk for each mechanism $a \in \mathcal{A}$ as the expected loss with respect to the stationary distribution of this mechanism under the true behavior model M^*, i.e., $\mathcal{R}(a, M^*) = E_{\pi(a,M^*)}L(a, b^a, u)$. The optimal mechanism minimizing this risk is denote as a^*, i.e., $a^* = \arg\min_{a \in \mathcal{A}} \mathcal{R}(a, M^*)$. We consider the gap between the risk of the mechanism $\hat{a}_{T_2}$ learned by the bi-level ERM algorithm and the risk of the optimal mechanism a^*, i.e., $\mathcal{R}(\hat{a}_{T_2}, M^*) - \mathcal{R}(a^*, M^*)$. We call this gap the (generalization) error for the bi-level ERM algorithm, or simply the (generalization) error for GTML.

To ease the analysis, we decompose the generalization error for GTML into two parts, as shown in the following Theorem (due to space restrictions, we leave the proof to supplementary materials).

Theorem 3.1. *The generalization error of the bi-level ERM algorithm for GTML can be bounded as:*

$$\mathcal{R}(\hat{a}_{T_2}, M^*) - \mathcal{R}(a^*, M^*) \leq 2KC(M^*)||M^* - \hat{M}_{T_1}||_\infty + 2\sup_{a \in \mathcal{A}} |\mathcal{R}(a, \hat{M}_{T_1}) - \mathcal{R}_{T_2}(a, \hat{M}_{T_1}, \delta)|, \quad (1)$$

where K is an upper bound for loss L, and $C(M^*)$ is a non-negative constant depending on M^*.

For ease of reference, we define the first term $||M^* - \hat{M}_{T_1}||_\infty$ in the right-hand side of inequality (1) as the error of behavior learning (or *behavior error* for short) and the second term $\sup_{a \in \mathcal{A}} |\mathcal{R}(a, \hat{M}_{T_1}) - \mathcal{R}_{T_2}(a, \hat{M}_{T_1}, \delta)|$ as the error of mechanism learning (or *mechanism error* for short). We will derive upper bounds for both errors in the following subsections.

3.1 Bound for Behavior Error

In this section we derive an upper bound for behavior error. Suppose the behavior space and signal space are both finite, i.e., $\mathcal{B} = \{B_1, \cdots, B_{|\mathcal{B}|}\}$ and $\mathcal{H} = \{H_1, \cdots, H_{|\mathcal{H}|}\}$. It is shown in [9] that $\{b_{t+1}, b_t, h_t\}$ forms a time-homogeneous Markov chain, and under regular conditions for its transition matrix, the Markov chain is uniformly ergodic and there exists N_0, such that the elements in the N_0-step transition probability matrix of $\{b_{t+1}, b_t, h_t\}$ are all positive (for ease of reference, we denote the minimum element in this matrix as δ_0). Since the mechanism is fixed, we omit all the super scripts a_0 in $b_t^{a_0}$ if without confusion. As mentioned in the introduction, both non-parametric and parametric approaches have been used to estimate the Markov transition probability of the behavior model. In this section, we focus on the non-parametric approach.[1]

In non-parametric behavior learning, the transition probability $M_{H_j}(B_i, B_k)$ is estimated by the conditional frequency of the event $b_{t+1} = B_k$ given that $b_t = B_i$ and $h_t = H_j$, i.e., $\hat{M}_{H_j}(B_i, B_k) = \frac{T_1(ijk)}{T_1(ij)}$ where $T_1(ijk) = \sum_{t=1}^{T_1} \mathbb{1}_{\{b_{t+1}=B_k, b_t=B_i, h_t=H_j\}}$ and $T_1(ij) = \sum_{t=1}^{T_1} \mathbb{1}_{\{b_t=B_i, h_t=H_j\}}$. The difficulty in analyzing the error of above estimate comes from sum of random variables in the denominator of conditional frequency. To tackle the challenge, we first derive an upper bound for the gap between conditional transition frequency and conditional transition probability, which does not involve such a sum of random variables, then apply the Hoeffding inequality for uniformly ergodic Markov chain [13] to this upper bound. In this way, we manage to obtain a behavior error bound, as shown in the following theorem.

Theorem 3.2. *For $\forall \epsilon > 0$, we have for $T_1 > \frac{2N_0(|\mathcal{B}|+1)}{|\mathcal{B}||\mathcal{H}|\delta_0\tilde{C}\epsilon}$,*

$$P(||\hat{M}_{T_1} - M^*||_\infty \geq \epsilon) \leq 2|\mathcal{H}||\mathcal{B}|^2(|\mathcal{B}|+1)\exp(-\frac{[\tilde{C}T_1\delta_0|\mathcal{B}||\mathcal{H}|\epsilon - 2N_0(|\mathcal{B}|+1)]^2}{2T_1N_0^2(|\mathcal{B}|+1)^2}),$$

where $\delta_0, N_0, \tilde{C}$ are positive constants.

Proof. According to the definition ∞-norm of matrix, we have

$$||\hat{M}_{T_1} - M^*||_\infty = \max_{H_j} ||\hat{M}_{H_j} - M_{H_j}||_\infty = \max_{H_j} \max_{1\leq i\leq|\mathcal{B}|} \sum_{k=1}^{|\mathcal{B}|} |\frac{T_1(ijk)}{T_1(ij)} - \frac{P_{ijk}}{P_{ij}}|. \tag{2}$$

where $P_{ijk} = P(b_{t+1} = B_k, b_t = B_i, h_t = H_j)$, $P_{ij} = P(b_t = B_i, h_t = H_j)$ and P is the stationary distribution of (b_{t+1}, b_t, h_t). We assume if $T_1(ij) = 0$, then $T_1(ijk)/T_1(ij) = 0$ and the same assumption holds for P_{ijk}/P_{ij}. Then we have

$$\begin{aligned}
&\left|\hat{M}_{H_j}(B_i, B_k) - M_{H_j}(B_i, B_k)\right| \leq \left|\frac{\frac{1}{T_1}T_1(ijk)}{\frac{1}{T_1}T_1(ij)}(1 - \frac{\frac{1}{T_1}T_1(ij)}{P_{ij}})\right| + \left|\frac{\frac{1}{T_1}T_1(ijk) - P_{ijk}}{P_{ij}}\right| \\
\leq &\left|\frac{P_{ij} - \frac{1}{T_1}T_1(ij)}{P_{ij}}\right| + \left|\frac{\frac{1}{T_1}T_1(ijk) - P_{ijk}}{P_{ij}}\right| \leq \sum_{1\leq l\leq|\mathcal{B}|} \left|\frac{\frac{1}{T_1}T_1(ijl) - P_{ijl}}{P_{ij}}\right| + \left|\frac{\frac{1}{T_1}T_1(ijk) - P_{ijk}}{P_{ij}}\right|.
\end{aligned} \tag{3}$$

Combining inequalities (2) and (3), we obtain

$$\begin{aligned}
&P(||\hat{M}_{T_1} - M^*||_\infty \geq \epsilon) \leq |\mathcal{H}||\mathcal{B}|^2 \max_{1\leq i,k\leq|\mathcal{B}|} \max_{1\leq j\leq|\mathcal{H}|} P(|\frac{T_1(ijk)}{T_1(ij)} - M_{H_j}(B_i, B_k)| \geq \frac{\epsilon}{|\mathcal{B}|}) \\
\leq &|\mathcal{H}||\mathcal{B}|^2(|\mathcal{B}|+1) \max_{1\leq i,l\leq|\mathcal{B}|} \max_{1\leq j\leq|\mathcal{H}|} P(|\frac{1}{T_1}T_1(ijl) - P_{ijl}| \geq \frac{P_{ij}\epsilon}{|\mathcal{B}|(|\mathcal{B}|+1)}) \\
\leq &2|\mathcal{H}||\mathcal{B}|^2(|\mathcal{B}|+1)\exp(-\frac{(T_1\delta_0|\mathcal{B}||\mathcal{H}|\tilde{C}\epsilon - 2N_0(|\mathcal{B}|+1))^2}{2T_1N_0^2(|\mathcal{B}|+1)^2}), \text{for } T_1 > \frac{2N_0(|\mathcal{B}|+1)}{|\mathcal{B}||\mathcal{H}|\delta_0\tilde{C}\epsilon}.
\end{aligned} \tag{4}$$

where $\tilde{C} = \min_{ij} P_{ij}$, and the last $\leq$ holds according to Hoeffding inequality for uniformly ergodic Markov chains [13]. □

[1] Note that in [9], the authors gave a uniform convergence bound for generalization error of the parametric behavior learning approach. However, their setting is different from that of GTML: while [9] only considers the behavior learning given a fixed mechanism a_0, in GTML the behavior model will be applied to generate behavior data for new mechanisms. Given that the behavior data induced by different mechanisms might have different distributions, the optimal model learned by the parametric approach over the data distribution induced by $a_0$may not be appropriate for the data distribution induced by other mechanisms, yielding learning bias for the parametric approach. To handle this problem, one may need to collect training data from multiple mechanisms. We will leave the corresponding discussions to the future work.

3.2 Bound for Mechanism Error

In this section, we bound the mechanism error by using a new concept called *nested cover number* for the mechanism space. We first give the definition of this concept, then prove a uniform convergence bound for mechanism learning on its basis. Finally, we take GSP auctions with reserve price as an example to show how the bound can be concretely computed.

3.2.1 Nested Cover Number of Mechanism Space

In this subsection, we define the nested cover for mechanism space. The nested cover contains two layers of covers: the first-layer cover is defined for the entire mechanism space based on the distance between stationary data distributions induced by the mechanisms. The second-layer cover is defined for each partition (subspace) obtained in the first layer based on the distance between the losses of the mechanisms projected onto finite common data samples.

First, we construct the first-layer cover for the mechanism space $\mathcal{A}$. In mechanism learning, the learned Markov behavior model $\hat{M}_{T_1}$ is used to generate the behavior data for different mechanisms. For simplicity, we denote the stationary distribution of the generated data as $\pi(a, \hat{M}_{T_1})$ (or π_a for simplification) and the set of stationary distributions for $\mathcal{A}$ as $\pi(\mathcal{A})$. We define the *(induced) total variance distance* on $\mathcal{A}$ as the total variance distance on $\pi(\mathcal{A})$, i.e., for $\forall a, a' \in \mathcal{A}$, $d_{TV}(a, a') = d_{TV}(\pi_a, \pi_{a'})$. For $\forall \epsilon > 0$, the smallest ϵ-cover of $\mathcal{A}$ w.r.t. the total variance distance is $cover^{\epsilon}_{TV} = \{g^{\epsilon}_1, \cdots, g^{\epsilon}_i, \cdots\}$, where $g^{\epsilon}_i \in \mathcal{A}$. That is, $\mathcal{A} \subseteq \cup_i B(g^{\epsilon}_i, \epsilon)$, where $B(g^{\epsilon}_i, \epsilon)$ is the ϵ-balls of g^{ϵ}_i with respect to the (induced) total variance distance. We define the first-layer covering number as the cardinality of $cover^{\epsilon}_{TV}$, denoted as $\mathcal{N}_{TV}(\epsilon, \mathcal{A})$. Based on $cover^{\epsilon}_{TV}$, we can obtain a partition for $\mathcal{A}$, denoted as $\{\mathcal{A}^{\epsilon}_i\}$, where $\mathcal{A}^{\epsilon}_i$ is an ϵ-partition of $\mathcal{A}$. When $\mathcal{B}$ and $\mathcal{U}$ are both finite, it is easy to see $\mathcal{N}_{TV}(\epsilon, \pi_{\mathcal{A}}) < \infty$. When the mapping from mechanism to its stationary distribution is L-Lipschitz continuous, if $\mathcal{A}$ is compact, then $\mathcal{N}_{TV}(\epsilon, \mathcal{A}) < \infty$; otherwise for $\forall \delta > 0$, a and $s(a, \delta)$ belong to the same δL-partition of $\mathcal{A}$.

Second, we consider the loss for each mechanism subspace $L \circ \mathcal{A}^{\epsilon}_i := \{L \circ a : \mathcal{B} \times \mathcal{U} \to [-K, 0] | L \circ a(b^a_t, u_t) = L(a, b^a_t, u_t), a \in \mathcal{A}^{\epsilon}_i\}$, and define its covering number with respect to the T_2 *common samples* $\{X^{\epsilon}_{i,t}\}^{T_2}_{t=1}$, where $X^{\epsilon}_{i,t} = \{b^{g^{\epsilon}_i}_t, u_t\}$ and $\{b^{g^{\epsilon}_i}_t\}^{T_2}_{t=1}$ are generated by mechanism g^{ϵ}_i. Again, we define the second-layer cover as the smallest ϵ'-cover of $L \circ \mathcal{A}^{\epsilon}_i|_{\{X^{\epsilon}_{i,t}\}^{T_2}_{t=1}}$ under the l_1 distance, i.e., $cover^{\epsilon'}_1(L \circ \mathcal{A}^{\epsilon}_i|_{\{X^{\epsilon}_{i,t}\}^{T_2}_{t=1}})$, and define the second-layer covering number $\mathcal{N}_1(\epsilon', L \circ \mathcal{A}^{\epsilon}_i, T_2)$ as its maximum cardinality with respect to the sample $\{X^{\epsilon}_{i,t}\}^{T_2}_{t=1}$.

In summary, the nested cover and covering number for a mechanism space are defined as follows:

Definition 3.3. *Suppose $\mathcal{A}$ is a mechanism space, we define its nested cover as $\left\{cover^{\epsilon}_{TV}(\mathcal{A}), \{cover^{\epsilon'}_1(L \circ \mathcal{A}^{\epsilon}_i|_{\{X^{\epsilon}_{i,t}\}^{T_2}_{t=1}})\}\right\}$, and its nested covering numbers as $\{\mathcal{N}_{TV}(\epsilon, \mathcal{A}), \{\mathcal{N}_1(\epsilon', L \circ \mathcal{A}^{\epsilon}_i, T_2)\}\}$.*

3.2.2 Uniform Convergence Bound for Mechanism Learning

In this subsection, we derive a uniform convergence bound for the ERM algorithm for mechanism learning. We first relate the uniform convergence bound for the entire mechanism space to that for the subspaces constructed according to the first-layer cover. In each subspace, we assume Markov chain (b^a_t, u_t) is stationary. Considering that uniform ergodic Markov chains are β-mixing [14], we make use of the *independent block technique* for mixing sequences [15] to transform the original problem based on dependent samples to that based on independent blocks. Finally, we apply the symmetrization technique and Hoeffding inequality to obtain the desired bound.

Theorem 3.4. *Suppose that the mapping from $\mathcal{A}$ to $\pi_{\mathcal{A}}$ is L-Lipschitz, and $\mathcal{A}$ is compact. For $\epsilon > 0$, and $0 < \delta < \epsilon/(KL)$, the following inequality holds:*

$$
\begin{aligned}
&P(\sup_{a \in \mathcal{A}} |\mathcal{R}_{T_2}(a, \hat{M}_{T_1}, \delta) - \mathcal{R}(a, \hat{M}_{T_1})| \geq \epsilon) \\
\leq &\mathcal{N}_{TV}(\delta L, \mathcal{A}) \max_{1 \leq i \leq \mathcal{N}_{TV}(\delta L, \mathcal{A})} \left[16\mathcal{N}_1((\epsilon - KL\delta)/16, L \circ \mathcal{A}^{\delta L}_i, T_2) \exp(-\frac{(\epsilon - \delta LK)^2 T_2}{256K^2 m}) + \frac{T}{m}\beta(g^{\delta L}_i, m)\right],
\end{aligned} \tag{5}
$$

where $m \in \mathbb{N}$ such that $T/2m \in \mathbb{N}$, K is the upper bound for the loss function L, and $\beta(g_i^{\delta L}, m)$ is the mixing rate of the process $\{X_{i,t}^{\delta L}\}$.

Proof. Denote $\eta = \delta L$, and $\tau = T/2m$. Since mapping from $\mathcal{A}$ to $\pi_{\mathcal{A}}$ is Lipschitz continuous and $\mathcal{A}$ is compact, thus $N = \mathcal{N}_{TV}(\eta, \mathcal{A}) < \infty$. $\forall a \in \mathcal{A}_i^\eta$, we have $d_{\mathcal{A}}(a, g_i^\eta) < \delta$. For sake of simplicity and w.l.o.g., we assume $s(a, \delta) = g_i^\eta$. $\forall a \in \mathcal{A}_i^\eta$, let $\mathcal{R}(a, \hat{M}_{T_1}, \delta) := E_{\pi(g_i^\eta, \hat{M}_{T_1})} L(a, X_{i,t}^\epsilon)$. We have $|\mathcal{R}(a, \hat{M}_{T_1}) - \mathcal{R}(a, \hat{M}_{T_1}, \delta)| \leq K||\pi(a, \hat{M}_{T_1}) - \pi(g_i^\eta, \hat{M}_{T_1})|| \leq KLd_{\mathcal{A}}(a, g_i^\eta) \leq K\eta$. Then,

$$P(\sup_{a\in\mathcal{A}} |\mathcal{R}_{T_2}(a, \hat{M}_{T_1}, \delta) - \mathcal{R}(a, \hat{M}_{T_1})| \geq \epsilon) \leq N \max_{1\leq i\leq N} P(\sup_{a\in\mathcal{A}_i^\eta} |\mathcal{R}_{T_2}(a, \hat{M}_{T_1}, \delta) - \mathcal{R}(a, \hat{M}_{T_1})| \geq \epsilon)$$

$$\leq \quad N \max_{1\leq i\leq N} P(\sup_{a\in\mathcal{A}_i^\eta} |\mathcal{R}_{T_2}(a, \hat{M}_{T_1}, \delta) - \mathcal{R}(a, \hat{M}_{T_1}, \delta)| \geq \epsilon - K\eta). \tag{6}$$

Divide $\{X_{i,t}^\eta\}$ into 2τ blocks, each of which consists of m consecutive samples. For $1 \leq j \leq \tau$, we define $H_j = \{t : 2(j-1)m + 1 \leq t \leq (2j-1)m\}$, and $T_j = \{t : (2j-1)m + 1 \leq t \leq (2j)m\}$. We introduce i.i.d. ghost blocks $\{\tilde{X}_{i,t}^\eta : t \in H_j\}$ for blocks $\{X_{i,t}^\eta : t \in H_j\}_{j=1}^\tau$. Therefore,

$$P(\sup_{a\in\mathcal{A}_i^\eta} |\mathcal{R}_T(a, \hat{M}_{T_1}, \delta) - \mathcal{R}(a, \hat{M}_{T_1}, \delta)| \geq \epsilon - K\eta)$$

$$\leq \quad 2P(\sup_{a\in\mathcal{A}_i^\eta} |\frac{1}{\tau}\sum_{j=1}^{\tau}\sum_{t\in H_j} L(a, \tilde{X}_{i,t}^\eta) - E\sum_{t\in H_j} L(a, \tilde{X}_{i,t}^\eta)| \geq m(\epsilon - K\eta)) + 2\tau\beta(g_i^\eta, m)$$

$$\leq \quad 16\mathcal{N}_1((\epsilon - K\eta)/16, L \circ \mathcal{A}_i^\eta, T_2) \exp(-\frac{(\epsilon - K\eta)^2\tau}{128K^2}) + 2\tau\beta(g_i^\eta, m) \tag{7}$$

The last $\leq$ holds since $\{\sum_{t\in H_j} L(a, \tilde{X}_{i,t}^\eta)\}_{j=1}^\tau$ are i.i.d.. Combining inequalities (6) and (7), we complete the proof of the theorem. □

One may note that the selection of m affects the bound given in Theorem 5. The following corollary shows how to set m when the common sample forms an algebraically mixing process.

Corollary 3.5. *If common sample $\{X_{i,t}^{\delta L}\}$ forms an algebraically mixing sequence, i.e., $\beta(g_i^{\delta L}, m) \leq \beta_0 m^{-\gamma_i}$, where $\beta_0, \gamma_i \geq 0$, then the optimal $m = C(T_2^{\frac{1}{1+s}})$, where $0 < s < \min_{1\leq i\leq \mathcal{N}_{TV}(\eta,\mathcal{A})} \gamma_i$ and C is positive constant. The error bound for mechanism learning is as below:*

$$P(\sup_{a\in\mathcal{A}} |\mathcal{R}_{T_2}(a, \hat{M}_{T_1}, \delta) - \mathcal{R}(a, \hat{M}_{T_1})| \geq \epsilon)$$

$$\leq \mathcal{N}_{TV}(\delta L, \mathcal{A}) \max_{1\leq i\leq \mathcal{N}_{TV}(\delta L,\mathcal{A})} \left[16\mathcal{N}_1((\epsilon - KL\delta)/16, L \circ \mathcal{A}_i^{\delta L}, T_2) \exp(-\frac{(\epsilon - \delta LK)^2}{256K^2C} T_2^{\frac{s}{1+s}}) + \frac{1}{C} T_2^{\frac{s-\gamma_i}{1+s}} \beta_0\right].$$

Remark 1: While the δ-sample sharing technique was originally proposed to improve efficiency, it seems that it plays an important role in generalization ability according to our proof. Then a question is whether this technique is necessary for generalization ability. Our answer is yes if $\mathcal{A}$ is infinite. Let us consider a special case in which $\mathcal{U} = \{u\}$ and $\pi_a \equiv \pi, \forall a \in \mathcal{A}$, i.e., the behavior model does not rely on the signals. If δ-sample sharing is not used, for finite T,

$$P(\sup_{a\in\mathcal{A}} |\frac{1}{T}\sum_{t=1}^{T} L(a, b_t^a, u_t) - EL(a, b_t^a, u_t)| \geq \epsilon) = 1 - \prod_{a\in\mathcal{A}} (1 - P(|\frac{1}{T}\sum_{t=1}^{T} L(a, b_t^a, u_t) - EL(a, b_t^a, u_t)| = 1.$$

This implies that mechanism learning without δ-sample sharing does not have generalization ability.

Remark 2: An assumption made in our proof is that the map from $\mathcal{A}$ to $\pi_{\mathcal{A}}$ is Lipschitz continuous. However, sometimes this assumption might not hold. In this case, we propose a modification to the original δ-sample sharing technique. The modification comes from the observation that the first-layer cover is constructed based on the total variance distance between stationary distributions of mechanisms. Therefore, in order to ensure a meaningful cover, we could let two mechanisms share the same data sample if the estimates of their induced stationary distributions (instead of their parameters) are similar. Please refer to the supplementary materials for details of this modification and a proof why it can bypass the discontinuity challenge. Note that the modified δ-sample sharing technique no longer has efficiency advantage since it involves the generation of behavior data for every mechanism examined during the training process, however, it ensures the generalization ability of the mechanism learning algorithm, which is desirable from the theoretic perspective.

3.2.3 Example: GSP Auctions with Reserve Price

In this section, we take GSP auction with a query-dependent reserve price [6][11][12] as an example to give a bound for its second layer covering number.

When a reserve price $r \in \mathcal{R}^+$ is used, the GSP auction runs in the following manners. First, the search engine ranks the ads according to their bid prices (here we follow the common practice to absorb the click-through rate of an ad into its bid price to ease the notations), and will show to the users those ads whose bid prices are higher than the reserve price. If the ad on the i-th position (denoted as ad_i) is clicked by a user, the search engine will charge the corresponding advertiser by the maximum of the bid price of ad_{i+1} and the reserve price r. For sake of simplicity and without loss of generality, we will only consider two ad slots. Let the binary vector $c = \{c_1, c_2\}$ indicate whether ad_1 and ad_2 are clicked by users. Then the user data include two components, i.e., $u = (q, c)$, where $q \in Q$ is the query issued by the user and c records user's click feedback. Denote the bid profile of the shown ads as $(b^{(1),q}, b^{(2),q})$ (for simplicity we sometimes omit q in the notation). We consider a query-dependent reserve price, i.e., the auction family is $\mathcal{A} = \{a : Q \to \mathbb{R}^+\}$. For a mechanism a, the revenue of the search engine can be represented as:

$$Rev(a, b, u) = a(q)c_1 \mathbb{1}_{\{b^{(2)} \le a(q) \le b^{(1)}\}} + (b^{(2)}c_1 + a(q)c_2)\mathbb{1}_{\{b^{(3)} \le a(q) \le b^{(2)}\}} + (b^{(2)}c_1 + b^{(3)}c_2)\mathbb{1}_{\{a(q) \le b^{(3)}\}}. \tag{8}$$

and the loss is $L(a, b, u) = -Rev(a, b, u)$. We give an upper bound of the second-layer covering number in the following theorem. (Please refer to the supplementary materials for its proof.)

Theorem 3.6. *For GSP auctions with reserve price, the covering number for $L \circ \mathcal{A}_i^\epsilon$ can be bounded by the pseudo-dimension (P-dim) of the reserve price function class. To be specific, we have:*

$$\mathcal{N}_1(\epsilon', L \circ \mathcal{A}_i^\epsilon, T_2) \le \left(\frac{eT_2K}{\epsilon'}\right)^{16|\mathcal{B}|P\text{-}dim(\mathcal{A}_i^\epsilon)} \le \left(\frac{eT_2K}{\epsilon'}\right)^{16|\mathcal{B}|P\text{-}dim(\mathcal{A})}, \forall T_2 > 4|\mathcal{B}|P\text{-}dim(\mathcal{A}) \tag{9}$$

3.3 The Total Error Bound

By combining Theorem 3.1, Theorem 3.2 and Theorem 3.4, we obtain the total error bound for GTML as shown in the following theorem.

Theorem 3.7. *For bi-level ERM algorithm in GTML, for $\epsilon > 0$, and $0 < \delta < \epsilon/(KL)$, we have the following generalization error bound:*

$$\begin{aligned}
&P(\mathcal{R}(\hat{a}_{T_2}, M^*) - \mathcal{R}(a^*, M^*) \ge \epsilon) \\
\le &\mathcal{N}_{TV}(\delta L, \mathcal{A}) \max_{1 \le i \le \mathcal{N}_{TV}(\delta L, \mathcal{A})} \left[16\mathcal{N}_1((\epsilon - 4KL\delta)/64, L \circ \mathcal{A}_i^{\delta L}, T_2) \exp\left(-\frac{(\epsilon/4 - KL\delta)^2 T_2}{256K^2 m}\right) + \frac{T_2}{m}\beta(g_i^{\delta L}, m)\right] \\
&+2|\mathcal{H}||\mathcal{B}|^2(|\mathcal{B}|+1)\exp\left(-\frac{[\tilde{C}T_1\delta_0|\mathcal{B}||\mathcal{H}|\epsilon - 8KC(M^*)N_0(|\mathcal{B}|+1)]^2}{32T_1N_0^2K^2(C(M^*))^2(|\mathcal{B}|+1)^2}\right),
\end{aligned}$$

where $m \in \mathbb{N}$ such that $T_2/2m \in \mathbb{N}$, K is the upper bound for the loss function L, $\beta(g_i^{\delta L}, m)$ is the mixing rate of the process $\{X_{i,t}^{\delta L}\}$, and $\tilde{C}, \delta_0, N_0, C(M^*)$ are positive constants.

4 Conclusion and Future Work

In this paper, we have given a formal generalization analysis to the game-theoretic machine learning (GTML) framework, which involves a bi-level learning process (i.e., mechanism learning and behavior learning). The challenges of generalization analysis for GTML lies in the dependency between the behavior data and the mechanism. To tackle the challenge, we first bound the error of behavior learning by leveraging the Hoeffding inequality for Markov Chains, and then introduce a new notion called *nested covering number* and bound the errors of mechanism learning on its basis. Our theoretical analysis not only enriches the understanding on machine learning algorithms in complicated dynamic systems with multi-party interactions, but also provides some practical algorithmic guidance to mechanism design for these systems.

As for future work, we plan to investigate whether and how parametric behavior learning algorithms can also lead to reasonable generalization ability, since parametric algorithms usually have their computational advantages over non-parametric approaches. We would also like to extend the idea of δ-sample sharing and apply it to improve the mechanisms in other real applications, such as mobile apps and social networks.

References

[1] S. Lahaie and D. M. Pennock, "Revenue analysis of a family of ranking rules for keyword auctions," in *Proceedings of the 8th ACM Conference on Electronic Commerce*, EC '07, (New York, NY, USA), pp. 50–56, ACM, 2007.

[2] D. Garg and Y. Narahari, "An optimal mechanism for sponsored search auctions on the web and comparison with other mechanisms," *IEEE Transactions on Automation Science and Engineering*, vol. 6, no. 4, pp. 641–657, 2009.

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

[4] F. Radlinski, A. Broder, P. Ciccolo, E. Gabrilovich, V. Josifovski, and L. Riedel, "Optimizing relevance and revenue in ad search: A query substitution approach," in *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '08, (New York, NY, USA), pp. 403–410, ACM, 2008.

[5] Y. Zhu, G. Wang, J. Yang, D. Wang, J. Yan, and Z. Chen, "Revenue optimization with relevance constraint in sponsored search," in *Proceedings of the Third International Workshop on Data Mining and Audience Intelligence for Advertising*, pp. 55–60, ACM, 2009.

[6] A. M. Medina and M. Mohri, "Learning theory and algorithms for revenue optimization in second price auctions with reserve," in *Proceedings of the Thirty-First International Conference on Machine Learning*, pp. 262–270, 2014.

[7] D. He, W. Chen, L. Wang, and T.-Y. Liu, "A game-theoretic machine learning approach for revenue maximization in sponsored search," *CoRR*, vol. abs/1406.0728, 2014.

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

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

[10] D. Fudenberg, *The theory of learning in games*, vol. 2. MIT press, 1998.

[11] B. Edelman, M. Ostrovsky, and M. Schwarz, "Internet advertising and the generalized second price auction: Selling billions of dollars worth of keywords," tech. rep., National Bureau of Economic Research, 2005.

[12] D. Easley and J. Kleinberg, *Networks, crowds, and markets*. Cambridge Univ Press, 2010.

[13] P. W. Glynn and D. Ormoneit, "Hoeffding's inequality for uniformly ergodic markov chains," *Statistics & probability letters*, vol. 56, no. 2, pp. 143–146, 2002.

[14] J. Doob, *Stochastic processes*. Wiley publications in statistics, Wiley, 1990.

[15] B. Yu, "Rates of convergence for empirical processes of stationary mixing sequences," *The Annals of Probability*, pp. 94–116, 1994.