Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence

Agent Behavior Prediction and Its Generalization Analysis

Fei Tian*
University of Science
and Technology of China
tianfei@mail.ustc.edu.cn

Haifang Li*
Chinese Academy of Sciences
lihaifang@amss.ac.cn

Wei Chen
Microsoft Research
wche@microsoft.com

Tao Qin
Microsoft Research
taoqin@microsoft.com

Enhong Chen
University of Science and Technology of China
cheneh@ustc.edu.cn

Tie-Yan Liu
Microsoft Research
tyliu@microsoft.com

Abstract

Machine learning algorithms have been applied to predict agent behaviors in real-world dynamic systems, such as advertiser behaviors in sponsored search and worker behaviors in crowdsourcing. Behavior data in these systems are generated by live agents: once systems change due to the adoption of prediction models learnt from behavior data, agents will observe and respond to these changes by changing their own behaviors accordingly. Therefore, the evolving behavior data will not be identically and independently distributed, posing great challenges to theoretical analysis. To tackle this challenge, in this paper, we propose to use *Markov Chain in Random Environments* (MCRE) to describe the behavior data, and perform generalization analysis of machine learning algorithms on its basis. We propose a novel technique that transforms the original time-variant MCRE into a higher-dimensional time-homogeneous Markov chain, which is easier to deal with. We prove the convergence of the new Markov chain when time approaches infinity. Then we obtain a generalization bound for the machine learning algorithms on the behavior data generated by the new Markov chain. To the best of our knowledge, this is the first work that performs the generalization analysis on data generated by complex processes in real-world dynamic systems.

1 Introduction

In this Internet era, more and more data are generated by self-interested agents in interactive systems. For example, in sponsored search, advertisers generate a large volume of bidding log data in their daily competitions with each other in attracting search users to click their ads; in crowdsourcing, workers generate a lot of behavior data when competing with other workers in getting tasks from the employers, and when completing the tasks assigned to them.

In many real cases including the aforementioned ones, there are three kinds of players in the systems: platform, users, and self-interested agents. Platform is the owner of the system, who designs the mechanism of the system and takes care of its execution. Users arrive at the platform in random, with their particular needs to be fulfilled. Agents behave strategically in order to attract the attention of the users so as to realize their own utilities. Taking both user needs and agent behaviors into consideration, the platform matches users with agents, extracts revenue from this procedure, and gives agents feedback on their performances (which depend on both the behaviors of agents and randomness in users). Upon the feedback, agents will adjust their behaviors in order to be better off in the future. To design a good mechanism, it is very important for the platform to understand and predict agent behaviors. With accurate prediction of agent behaviors, the platform can also provide tools to help agents to optimize their performances and therefore attract more agents to the system. Thus, predicting agent behaviors is an important task for the platform. For ease of reference, we call the problem of predicting agent behaviors in an interactive system *"agent behavior prediction (ABP)"*.

1.1 Examples of ABP

Here we take sponsored search systems as an example for illustration. More examples can be found in the full version of the paper(Tian et al. 2014).

In a sponsored search system, platform, users, and agents correspond to the search engine, search users, and advertisers respectively. Advertiser behaviors are the bid prices on their ads. When a search user issues a query, the search engine will run a GSP auction (Edelman, Ostrovsky, and Schwarz 2005) among all advertisers who bid on the query, rank their ads according to the product of the bid price and predicted click-through rate, and then charge the winning advertisers if the user clicks on their ads. After a period of time, the search engine will provide feedback to the advertisers about their performances (which we usually call Key Performance Indicators, or KPIs for short). The KPIs usually contain numbers of impressions and clicks, average rank positions, and costs per click of their ads. Many advertisers will adjust their bid prices based on the feedback they receive, either by themselves or with the help of third-party search engine marketing companies. By logging the bid prices in a

*This work was done when the first two authors were visiting Microsoft Research Asia.

sufficiently long period of time, the search engine can predict how advertisers behave, and consequently enhance its click prediction algorithm and auction mechanism.

1.2 Generalization Analysis for ABP

Because of its importance, ABP has been studied in many works, including (Cary et al. 2007; Pin and Key. 2011; Zhou and Lukose. 2007; Xu et al. 2013; He et al. 2013). Some of them (Xu et al. 2013; He et al. 2013) have adopted machine learning techniques and attempted to learn an agent behavior model by means of empirical risk minimization (ERM) on the behavior logs. Empirical results have shown that these machine learning techniques can significantly outperform previous non-learning approaches. However, despite the experimental success, it still remains an open question whether the use of ERM algorithms in behavior prediction is theoretically sound, and whether certain generalization ability of such algorithms can be guaranteed.

As far as we know, the answers to the above questions are unclear yet. This is mainly because of the complication of the corresponding theoretical analysis. As aforementioned, the behavior data are generated by self-interested agents, and dependent on both their previous behaviors and the random user factors in the system. As a result, the behavior data have quite complex statistical properties, and the generalization analysis in such a setting goes beyond the state of the art of statistical learning theory (Vapnik 1998; Devroye 1996; Yu 1994).

More detailed introduction of related work towards generalization analysis for ABP can be found in the full version of the paper(Tian et al. 2014).

1.3 Our Results

In order to analyze the ERM algorithms on agent behavior prediction, we propose a set of new techniques in this paper.

First, we model the generation process of the behavior data using a so-called *Markov Chain in Random Environments* (MCRE), whose transition matrix is time-variant (depending on the random environments). Take sponsored search as an example. After the current-round auction, the advertiser observes his/her KPIs, which depend on both the bids of all the advertisers and the random clicks of the users. Based on the KPIs, the advertiser will determine how to set his/her own bid for the next round of auction and for different KPI values, the conditional distribution of his/her bid at the next round will be different. In this sense, the sequence of advertiser bids can be regarded as an MCRE.

Second, considering that it is difficult to perform generalization analysis on MCRE, we propose a transformation that maps the original MCRE to a higher dimensional time-homogenous Markov chain. Although the new Markov chain involves more variables, it is more *regular* and thus easier to deal with from the perspective of generalization analysis. We prove that the new Markov chain will converge when time approaches infinity and a Hoeffding-style inequality holds for the empirical process associated with it.

Third, by exploiting the covering number technique, we derive a uniform convergence bound for the ERM algorithms on the data generated by the new Markov chain (and thus by the original MCRE due to the equivalence transformation). As a consequence, we prove that the ERM algorithms on the data generated by MCRE have their theoretical guarantees, which explains their good empirical performances reported in previous works. To the best of our knowledge, this is the first work in the literature that performs formal generalization analysis on the agent behavior prediction problem.

2 Agent Behavior Prediction

In this section we give a formal description of the Agent Behavior Prediction (ABP) problem. We first show that the generation process for the behaviors of self-interested agents can be described by a Markov Chain in Random Environments (MCRE), and then formulate ABP as an optimization problem.

2.1 Agent Behaviors: Markov Chain in Random Environments

The dynamic interactive systems mentioned in the introduction share some common properties. (1) The behaviors of an agent only depend on a finite number of his/her historical actions due to the limited memory of human being. In other words, the behaviors have Markovian properties. (2)Random users' behaviors are independent and identically distributed (i.i.d), for example, in sponsored search, there are two aspects associated with users: queries issued by users, and users' click patterns on ad ranking lists. It is clear that queries can be regarded as i.i.d. random variables. Click patterns are defined with respect to all possible ad ranking lists and they are also independent of agent behaviors(which only determine the selected ranking list), thus can be regarded as i.i.d. (3) The behavior change of an agent is mainly affected by the feedback given by the platform. Since the feedback depends on the users randomly arriving at the system, the behavior change is not governed by a constant rule, but instead by some random factors.

Taking all the three properties into consideration, we can regard agent behaviors as generated by a Markov Chain in Random Environments (MCRE)(Cogburn 1980). Note that this generation process is much more complicated than a simple i.i.d. sampling process or a time-homogeneous Markov process.

Before formally describing the MCRE process for agent behaviors, we make some assumptions. First, we assume that both the behavior space and feedback space are finite. This assumption is reasonable, since in many applications the behaviors of the agents are either characterized by categorical profiles (e.g., expertise and functionalities) or bounded and associated with a minimum unit (e.g., bidding price, payment requirements, and available time slots). The same reason holds for the finiteness of the feedback space, since feedback usually takes discrete values (e.g., number of clicks, ratings, and number of reviews) as well. Second, we assume there is a deterministic function that generates the feedback for a given agent i based on the behaviors of all the agents and the random arrival of the users. This assumption is also reasonable because the feedback is usually provided by the platform using a deterministic algorithm.

With the above assumptions, now let us describe the generation process of the agent behavior according to (He et al. 2013). Suppose there are N agents. Let $\mathbb{B}$ and $\mathbb{H}$ be the behavior space and feedback space of a single agent respectively, and $\mathbb{U}$ be the random factor space induced by users. We use the mapping $\eta_i : \mathbb{U} \times \mathbb{B}^N \rightarrow \mathbb{H}$ to denote the deterministic function that generates the feedback for agent i, and $\eta : \mathbb{U} \times \mathbb{B}^N \rightarrow \mathbb{H}^N$ to denote the joint feedback for all the agents. At the beginning of the $(t+1)$-th time period, agent i receives feedback $h_t^i = \eta_i(u_t, b_t)$ about his/her behaviors in the t-th time period. Based on the feedback, agent i may change his/her behavior to b_{t+1}^i in order to be better off. That is, given the Markov property,

$$P(b_{t+1}^i | b_1, ..., b_t; u_1, ..., u_t) = P(b_{t+1}^i | b_t^i; h_t^i) \quad (1)$$

Note that the above equation implies the one-step Markov property, which is motivated here mainly due to ease of statement. Our analysis in this paper can be extended to higher order Markov chain as well, without too many modifications.

Given the feedback h_t, all the agents change their behaviors independently, so we have

$$\begin{aligned} P(b_{t+1} | b_1, ..., b_t; u_1, ..., u_t) &= \prod_{i=1}^{N} P(b_{t+1}^i | b_t^i; h_t^i) \\ &= P(b_{t+1} | b_t, h_t) \end{aligned} \quad (2)$$

This indicates that $\{b_t\}$ is an MCRE (Cogburn 1980), where the environmental process is $\{h_t\}$. According to (2), the one-step transition matrix of an MCRE is time-variant and depends on the environmental variable.

2.2 Learning Agent Behavior Model

There exist some related works that leverage the empirical risk minimization (ERM) framework to learn the agent behavior model. Mathematically speaking, given a training set containing the behaviors of agents and the feedback they received in T rounds, $\{(h_1, b_1); (h_2, b_2); ...; (h_T, b_T)\}$, one aims to learn a function f: $\mathbb{H}^N \times \mathbb{B}^N \rightarrow \mathbb{B}^N$, which takes the behaviors and feedback at the current round as inputs and predicts the behavior at the next round. To this end, one minimizes the empirical risk on the training set: $\min_{f \in \mathcal{F}} \sum_{t=1}^{T} l\left(f\left(h_t, b_t\right); b_{t+1}\right)$, where l measures the loss between the predicted behavior and the real behavior in the training data. For example, l can be the $0-1$ classification loss: $l\left(f\left(b_t, h_t\right); b_{t+1}\right) = 1_{f(h_t, b_t) \neq b_{t+1}}$, and can also be some surrogate loss functions.

This ERM framework covers the algorithms for predicting advertiser behaviors in many previous works including (He et al. 2013) and (Xu et al. 2013). For example, in (He et al. 2013), a truncated Gaussian function is used to model the Markov transition probabilities and a negative likelihood function is used as the loss function l. For another example, in (Xu et al. 2013), a compound function that considers the willingness, capability, and constraint of an advertiser is used as the model f, and again the negative likelihood of the historical behaviors is used as the loss function f.

As mentioned in the introduction, the ERM algorithms led to experimental success in the ABP tasks. However, it is unclear whether these algorithms have desired theoretical guarantee. In particular, it is unknown (1) qualitatively whether the ABP problem is learnable through an ERM process, and (2) quantitatively what is the relationship between generalization error and the size of the training data. The reason why the answers to these important questions are missing lies in that statistical learning theory mainly addresses learning problems with data generated by an i.i.d. sampling or a β-mixing Markov chain. This motivates us to formally study the learning theory with respect to the data generated by more complicated stochastic processes like MCRE.

3 Generalization Bounds for ABP

In this section, we perform generalization analysis on the ERM algorithms for agent behavior prediction. Our main result is stated in Theorem (6), which is proved in three steps: (1) constructing a new Markov chain of higher dimensionality but with more regular properties than the original MCRE; (2) proving the convergence of the empirical loss to the expected loss when the data are generated by this new Markov chain; (3) proving a uniform convergence bound by further leveraging the technique of covering number. For ease of reference, we use a Notation Table 1 in the end of the paper to summarize all notations used in this section.

3.1 Constructing a Higher-Dimensional Markov Chain

The difficulty of analyzing the ERM algorithms when the data are generated by an MCRE lies in its time-variant transition probabilities. To tackle the challenge, we construct a higher-dimensional chain $\mathbb{M} = \{(h_t, b_t, b_{t+1}); t \geq 0\}$, by grouping *correlated* variables together. Let $M_k(m, n)$ be one-step transition probability of $\{b_t\}$ from state m to n under the environment $k \in \mathbb{H}^N$. For convenience, we set $z_t = (h_t, b_t, b_{t+1})$ to be the t-th state of $\mathbb{M}$. Since the state space $\mathbb{H}^N \times \mathbb{B}^{2N}$ of the chain is finite, we label all the state values as $1, 2...Z$, i.e., $\mathbb{M}$ takes value in state space $[Z] := \{1, 2, ..., Z\}$.

The following Lemma (1) and Theorem (2) show that the new Markov chain is time-homogeneous and has a stationary distribution under some mild assumptions.

Lemma 1 *We assume the random factors caused by users u_t are i.i.d., then the constructed Markov chain $\mathbb{M} = \{(h_t, b_t, b_{t+1}); t \geq 0\}$ is time-homogeneous.*

Proof *Since both feedback space and behavior space are finite, the set of their three dimensional Cartesian product values is also finite.*

To show that the process is a time-homogeneous Markov chain, we consider any two states $(j, p, q), (k, m, n) \in \mathbb{H}^N \times \mathbb{B}^{2N}$ and write their one-step transition probability as:

$$
\begin{aligned}
&P(h_t = k, b_t = m, b_{t+1} = n| \\
&(h_s, b_s, b_{s+1})_{s=1}^{t-2}; \quad h_{t-1} = j, b_{t-1} = p, b_t = q) \\
=&P(h_t = k, b_t = m, b_{t+1} = n|h_{t-1} = j, b_{t-1} = p, b_t = q) \\
=&\begin{cases} 0 & for \quad m \neq q \\ P(u \in \mathbb{U} : \eta(u, m) = k)M_k(m, n) & for \quad m = q \end{cases}
\end{aligned} \tag{3}
$$

From (3), we know that the value of (h_t, b_t, b_{t+1}) only depends on (h_{t-1}, b_{t-1}, b_t). Therefore the process is a Markov chain. Furthermore, the one-step transition probability only depends on the states $(j, p, q), (k, m, n)$ and is independent of time index t, and therefore the transition probability is time-invariant. ∎

For ease of statement, we will use M to denote the probability transition matrix in (3). As we can see from (3), lots of elements in M are zero. Therefore it is not straightforward to judge whether Markov chain $\mathbb{M}$ will converge or not. Theorem (2) shows that under some mild assumptions the convergence can be achieved.

Theorem 2 *The Markov chain $\mathbb{M}$ composed by $\{z_t\}_{t=1}^T := \{h_t, b_t, b_{t+1}\}_{t=1}^T$ has a stationary distribution under the following assumptions:* ***(A.1)*** *for every fixed value of the feedback $h_t = k$, the Markov process with transition matrix M_k is irreducible and aperiodic;* ***(A.2)*** *the feedback function η and the random user distribution satisfy $\forall m \in \mathbb{B}, k \in \mathbb{H}, P(u \in \mathbb{U} : \eta(u, m) = k) > 0$.* [1]

To prove this theorem, since the state of (h_t, b_t, b_{t+1}) is finite and $\mathbb{M}$ is time-homogeneous, we only need to prove that Markov chain $\mathbb{M}$ is irreducible and aperiodic, which is shown in the following two lemmas respectively.

Lemma 3 *Under the assumption $(A.1)$ and $(A.2)$, the Markov chain $\mathbb{M}$ is irreducible.*

Proof *According to the definition of irreducibility, we need to show that any two states $(k, m, n) \in \mathbb{H}^N \times \mathbb{B}^{2N}$ and $(j, p, q) \in \mathbb{H}^N \times \mathbb{B}^{2N}$ are accessible to each other. To prove that, we show three simple facts in the following:*

- *According to assumption $(A.2)$, it is possible to produce feedback signal k by a one step transition from state (j, p, q). i.e. $\exists x \in \mathbb{B}, s.t.P(z_{t+1} = (k, q, x)|z_t = (j, p, q)) > 0$.*
- *In Markov chain M_k where the feedback signal is fixed to be k, there exists $d \in \{1, 2, \cdots\}$, such that we can build a d-step transition path from behavior profile x to behavior profile m, followed by a one step transition to behavior profile n.*
  The existence of the d-step path from q to m is due to the irreducibility of M_k in assumption $(A.1)$. To see that it is possible to transit from state m to n by one step in Markov chain M_k, note that if $M_k(m, n) = 0$, we can simply erase the state (k, m, n) from the state space of Markov chain $\mathbb{M}$, which does not affect our results, therefore we only need to consider the case in which $M_k(m, n) > 0$.
- *According to assumtion $(A.2)$, in Marokov chain $\mathbb{M}$, it is possible to observe $d + 1$ consecutive states in which the $d + 1$ feedback signals are all k.*
  Combining the three points, we can obtain:

$$
\begin{aligned}
&P(z_{t+d+2} = (k, m, n)|z_t = (j, p, q)) \\
\geq&P(z_{t+d+2} = (k, m, n)|z_{t+1} = (k, q, x)) \\
&\times P(z_{t+1} = (k, q, x)|z_t = (j, p, q)) \\
=&M_k(m, n) \cdot P(u \in \mathbb{U} : \eta(u, m) = k) \\
&\times \cdots \times M_k(q, x) \cdot P(u \in \mathbb{U} : \eta(u, q) = k) \\
>&0.
\end{aligned} \tag{4}
$$

Therefore state (k, m, n) is reachable from state (j, p, q). Similarly, we can also prove that state (j, p, q) is reachable from state (k, m, n). Since (k, m, n) and (j, p, q) are arbitrarily chosen, we actually prove the irreducibility of the Markov chain $\mathbb{M}$. ∎

Lemma 4 *Under the assumption $(A.1)$ and $(A.2)$, the Markov chain $\mathbb{M}$ is aperiodic.*

Proof *Since the Markov chain is irreducible, all states in the chain have the same period. Therefore, in order to prove the aperiodicity, we just need to show that for a given state $(k, m, m) \in \mathbb{H}^N \times \mathbb{B}^{2N}$, its period is one. According to the first assumption in the lemma, $\forall d \geq 1$ satisfying $M_k^{(d)}(m, m) > 0$. Therefore we can build a d-step path in Markov chain M_k: $m \to m_1 \to m_2 \cdots \to m_{d-1} \to m$, such that the transition probability in every step is positive, i.e., $M_k(m, m_1) > 0$, $M_k(m_1, m_2) > 0$, ..., $M_k(m_{d-1}, m) > 0$. As a result,*

$$
\begin{aligned}
&P(z_{t+d} = (k, m, m)|z_t = (k, m, m)) \\
\geq&P(z_{t+d} = (k, m, m)|z_{t+d-1} = (k, m_{d-1}, m))\times \\
&P(z_{t+d-1} = (k, m_{d-1}, m)|z_{t+d-2} = (k, m_{d-2}, m_{d-1})) \\
&\times \cdots \times P(z_{t+1} = (k, m, m_1)|z_t = (k, m, m)) \\
=&M_k(m_{d-1}, m) \cdot P(u \in \mathbb{U} : \eta(u, m_{d-1}) = k) \\
&\times M_k(m_{d-2}, m_{d-1}) \cdot P(u \in \mathbb{U} : \eta(u, m_{d-2}) = k) \\
&\times \cdots \times M_k(m, m_1) \cdot P(u \in \mathbb{U} : \eta(u, m) = k) \\
>&0.
\end{aligned} \tag{5}
$$

Hence, a d-step transition path with positive probability in Markov chain $\mathbb{M}$ can be built as $(k, m, m) \to (k, m, m_1) \to (k, m_1, m_2) \cdots \to (k, m_{d-1}, m) \to (k, m, m)$. Then by the definition of period [2]*, Markov chain $\mathbb{M}$ has the same period as Markov chain M_k, which is one.* ∎

3.2 Convergence Bound

In the previous subsection, we have constructed a new Markov chain and proved its convergence. In this subsection, we show that these results can be used to analyze the

[1] Condition (A.2), which seems not very intuitive, basically says that every possible value of the feedback is *reachable* from every value of the behavior, if there are a very large number of random users arriving at the platform and they have very high dynamics and variety.

[2] In Markov chain P, state i's period is defined as the g.c.d. of all $d \in 1, 2, \cdots$ satisfying $P^{(d)}(i, i) > 0$.

convergence rate of the empirical risk for a specified behavior prediction model, namely f.

Let us start from a formal definition of the problem. Given the T-round training data $S = \{(h_1, b_1, b_2); (h_2, b_2, b_3); \cdots; (h_T, b_T, b_{T+1})\} = (z_1, z_2, \cdots, z_T)$, we define the T-round empirical risk of f with respect to S as follows:

$$err_S^T(f) = \frac{1}{T}\sum_{t=1}^{T} l(f(b_t, h_t), b_{t+1}) = \frac{1}{T}\sum_{t=1}^{T} l(f, z_t), \quad (6)$$

where we assume l to be upper bounded by a constant $B > 0$.

We then define the T-round expected risk of f as $err_M^T(f) = \frac{1}{T}\sum_{t=1}^{T} E_{z_t\sim\pi_0 M^t} l(f, z_t)$, where π_0 is the initial distribution of Markov chain $\mathbb{M}$. According to Theorem(2), the limit of $err_M^T(f)$ exists:

$$err_\pi(f) = \lim_{T\to\infty} err_M^T(f) = \sum_{z\in[Z]} l(f, z)\cdot \pi_z, \quad (7)$$

where π is the stationary distribution of $\mathbb{M}$, and $err_\pi(f)$ is the *real* expected risk of f.

Next we investigate how well the T-round empirical risk $err_S^T(f)$ approximates $err_\pi(f)$. For this purpose, we leverage the Hoeffding Inequality for uniformly ergodic Markov Chains (Glynn and Ormoneit 2002), which is rephrased as below for completeness.

Proposition 1 ***(Hoeffding Inequality for uniformly ergodic Markov Chains)*** *Let $X = (X_n : n \geq 0)$ be a Markov Chain taking values in a state space S, if the following assumption holds:* ***(A.3)*** *there exists a probability measure ϕ on S, $\lambda > 0$, and an integer $m \geq 1$ s.t. $\forall x \in S$, $P(X_m \in \cdot|X_0 = x) \geq \lambda\phi(\cdot)$, then for a function $g : S \to \mathcal{R}$ with its norm defined as $||g|| = sup\{|g(x)| : x \in S\} < \infty$, define $S_T = \frac{1}{T}\sum_{t=1}^{T} g(X_t)$, for $T > 2||g||m/(\lambda\epsilon)$, we have*

$$P(|S_T - E(S_T)| \geq \epsilon) \leq 2exp(-\frac{\lambda^2(T\epsilon - 2||g||m/\lambda)^2}{2T||g||^2m^2}) \quad (8)$$

where the expectation $E(S_T)$ is taken on the stationary distribution of X.

In order to leverage Proposition (1), we need to check whether its assumption $(A.3)$ holds in our problem. For this purpose, we note the fact that for an irreducible, aperiodic, and finite-state Markov chain with time-invariant transition probability matrix P, there must exist N such that $\forall n \geq N$, all elements of n-step transition matrix $P^{(n)}$ are non-zero(Lemma 6.6.3 in (Durrett 2010)). Accordingly in our setting, for Markov chain $\mathbb{M}$, there exists N_0 such that $\forall 1 \leq i, j \leq Z, M_{i,j}^{(N_0)} > 0$. Denote δ as the minimum element in $M^{(N_0)}$, i.e., $\delta = \min_{1\leq i,j\leq Z} M_{i,j}^{(N_0)} > 0$. Then if we set $m = N_0$, $\lambda = Z\delta$ and set ϕ to be the uniform distribution on $[Z]$, it is easy to see that $(A.3)$ holds. Further noticing that $||g||$ in (8) is B in our setting, where $g = l(f; z_t)$ and B is the upperbound of function l, we can leverage Proposition (1) to obtain desired convergence bound as Theorem (5) shows.

Theorem 5 ***Convergence of Empirical Loss to Expected Loss*** *Let $f : \mathbb{H}^N \times \mathbb{B}^N \to \mathbb{B}^N$ be the behavior prediction function, $err_S^T(f)$ and $err_\pi(f)$ be the empirical loss and expected loss respectively, as defined in equation (6) and (7). For any $\epsilon > 0$ and $T > 2BN_0/(Z\delta\epsilon)$, we have*

$$\begin{aligned} &P(\left|err_S^T(f) - err_\pi(f)\right| \geq \epsilon) \\ \leq &2exp(-\frac{Z^2\delta^2\,(T\epsilon - 2BN_0/(Z\delta))^2}{2TB^2N_0^2}) \end{aligned} \quad (9)$$

Theorem (5) basically states that when the sample size T is large enough, the empirical risk $err_S^T(f)$ will converge to the long-term expected risk $err_\pi(f)$, and the convergence rate is exponential in the sample size.

3.3 Uniform Convergence Bound

In this subsection, we prove a uniform convergence bound for the ABP problem based on covering number,[3] as shown in the following theorem.

Theorem 6 ***Uniform Convergence Theorem.*** *Let $\mathcal{F} = \{f : \mathbb{H}^N \times \mathbb{B}^N \to \mathbb{B}^N\}$ be the behavior function space and $l \circ \mathcal{F}$ be the composite function class of the loss function l acting on $\mathcal{F}$. For a behavior function $f \in \mathcal{F}$, denote $err_S^T(f)$ and $err_\pi(f)$ as its empirical loss and expected loss respectively, as defined in equation (6) and (7). For any $\epsilon > 0$, we have*

$$\begin{aligned} &P(\sup_{f\in\mathcal{F}} |err_S^T(f) - err_\pi(f)| \geq \epsilon) \\ \leq &8\mathcal{N}(\frac{\epsilon}{8}, l\circ\mathcal{F}, 2T)exp(-\frac{Z^2\delta^2\,(T\epsilon - 16BN_0/(Z\delta))^2}{128TB^2N_0^2}) \end{aligned} \quad (10)$$

for $T \geq max\,(T_0,\ 16BN_0/(Z\delta\epsilon))$, where T_0 satisfies $\frac{Z^2\delta^2(T_0\epsilon - 4BN_0/(Z\delta))^2}{8T_0B^2N_0^2} \geq \ln 4$, and $\mathcal{N}(\frac{\epsilon}{8}, l\circ\mathcal{F}, 2T)$ is the $\epsilon/8$-covering number of $l \circ \mathcal{F}$. N_0 and δ are the parameters of Markov chain $\mathbb{M}$ (see Notation Table 1), B is the upper bound of loss function $l(f, z)$, Z is the state number of Markov chain $\mathbb{M}$.

The proof of the theorem contains two steps. For the first step, we employ the symmetrization technique to reduce the probability of the uniform convergence bound to a probability involving only two sample sets. For the second step, we further reduce the case to a finite function class by using the covering number theory.

Lemma 7 ***Symmetrization Lemma.*** *Denote $\tilde{S} = (\tilde{z}_1, \tilde{z}_2, ..., \tilde{z}_T)$ as a set of ghost samples sampled from Markov chain $\mathbb{M}$, then we have,*

$$\begin{aligned} &P(\sup_{f\in\mathcal{F}} \left|err_S^T(f) - err_\pi(f)\right| \geq \epsilon) \\ \leq &2P(\sup_{f\in\mathcal{F}} \left|err_S^T(f) - err_{\tilde{S}}^T(f)\right| \geq \frac{\epsilon}{2}) \end{aligned} \quad (11)$$

[3]Covering number is one of the common ways to measure the complexity of a function class. Specifically, covering number $\mathcal{N}(\epsilon, l\circ\mathcal{F}, T)$ is defined as $\max_{\{z_t\}_{t=1}^T\in[Z]^T} \mathcal{N}(\epsilon, l\circ\mathcal{F}|_{\{z_t\}_{t=1}^T}, d_T)$ in which $\mathcal{N}(\epsilon, l\circ\mathcal{F}|_{\{z_t\}_{t=1}^T}, d_T)$ is the minimum capacity of ϵ-cover of $(l\circ\mathcal{F})$'s projection on data $\{z_t\}_{t=1}^T$, w.r.t. the distance metric between $x \in R^T, y \in R^T$ defined as $d_T(x, y) := \max_{1\leq t\leq T} |x_t - y_t|$.

for T large enough to satisfy $\frac{Z^2\delta^2(T\epsilon-4BN_0/(Z\delta))^2}{8TB^2N_0^2} \geq \ln 4$.

The proof of the lemma can be found in the full version of the paper(Tian et al. 2014).

Proof of Theorem (6)

Fixing any sample data $(S, \tilde{S})$, we set $z_{T+t} = \tilde{z}_t, \quad t = 1, 2, ..., T$ for simplicity. Pick a subset of $\mathcal{F}$ $\mathcal{G} \subseteq \mathcal{F}$ such that $l \circ \mathcal{G}$ is an $\epsilon/8$-cover of $l \circ \mathcal{F}$ with respect to the metric $d_{2T}(l \circ f_1, l \circ f_2) := \max_{1 \leq t \leq 2T} |l(f_1, z_t) - l(f_2, z_t)|$. Pick $f \in \mathcal{F}$ such that $|err_S^T(f) - err_{\tilde{S}}^T(f)| \geq \epsilon/2$. According to the definition of $\mathcal{G}$, there exists a function $g \in \mathcal{G}$ such that $\max_{1 \leq t \leq 2T} |l(f, z_t) - l(g, z_t)| < \frac{\epsilon}{8}$ holds. In fact, as shown below, such a g satisfies $|err_S^T(g) - err_{\tilde{S}}^T(g)| \geq \epsilon/4$:

$$\begin{aligned}
&\frac{\epsilon}{2} \leq \left|err_S^T(f) - err_{\tilde{S}}^T(f)\right| \\
\leq& |err_S^T(f) - err_S^T(g)| + |err_S^T(g) - err_{\tilde{S}}^T(g)| \\
&+ |err_{\tilde{S}}^T(g) - err_{\tilde{S}}^T(f)| \\
\leq& |err_S^T(g) - err_{\tilde{S}}^T(g)| + \frac{1}{T}\sum_{t=1}^{2T} |l(f, z_t) - l(g, z_t)| \\
\leq& |err_S^T(g) - err_{\tilde{S}}^T(g)| + 2\max_{1 \leq t \leq 2T} |l(f, z_t) - l(g, z_t)| \\
<& |err_S^T(g) - err_{\tilde{S}}^T(g)| + \frac{\epsilon}{4}
\end{aligned} \tag{12}$$

The above result enables us to reduce the problem to the case of finite function class. Combining it with Inequality (11), we obtain

$$\begin{aligned}
&P(\sup_{f \in \mathcal{F}} \left|err_S^T(f) - err_{\tilde{S}}^T(f)\right| \geq \epsilon/2) \\
\leq& P(\max_{g \in \mathcal{G}} \left|err_S^T(g) - err_{\tilde{S}}^T(g)\right| \geq \epsilon/4) \\
\leq& \mathcal{N}(\epsilon/8, l \circ \mathcal{F}, 2T) \max_{g \in \mathcal{G}} P(\left|err_S^T(g) - err_{\tilde{S}}^T(g)\right| \geq \epsilon/4)
\end{aligned} \tag{13}$$

Further considering the result stated in Theorem (5), we obtain

$$\begin{aligned}
&P(\left|err_S^T(g) - err_{\tilde{S}}^T(g)\right| \geq \epsilon/4) \\
\leq& P(\left|err_S^T(g) - err_{\pi}(g)\right| \geq \epsilon/8) \\
&+ P(\left|err_{\tilde{S}}^T(g) - err_{\pi}(g)\right| \geq \epsilon/8) \\
\leq& 4exp(-\frac{Z^2\delta^2\left(T\epsilon - 16BN_0/(Z\delta)\right)^2}{128TB^2N_0^2})
\end{aligned} \tag{14}$$

for $T > 16BN_0/(Z\delta\epsilon)$.

By combining inequalities (11), (13), and (14), we finally prove the theorem. ∎

Remark: Please note that for most regular function class, the covering number $\mathcal{N}(\epsilon, l \circ \mathcal{F}, T)$ defined in Theorem 6 can be polynomially bounded. For example,

- If the loss function l is Lipschitz-continuous in its first argument with Lipschitz constant $L > 0$, then for any T, we have $\mathcal{N}(\epsilon, l \circ \mathcal{F}, T) \leq \mathcal{N}(\epsilon/L, \mathcal{F}, T)$.
- Since we assume the behavior set to be finite, the ABP problem can actually be regarded as a multi-class classification problem, where the class number is $|\mathbb{B}|$. In this case, the covering number $\mathcal{N}(\frac{\epsilon}{8}/L, \mathcal{F}, 2T)$ can be bounded by the growth function of $\mathcal{F}$, defined as $\max_{\{z_t\}_{t=1}^T \in [Z]^{2T}} \left|\mathcal{F}|_{\{z_t\}_{t=1}^{2T}}\right|$. Moreover, the growth function is bounded by $(\frac{2Te(|\mathbb{B}|+1)^2}{2d})^d$ (Bendavid et al. 1995), where d is the Natarajan dimension of $\mathcal{F}$ (Natarajan 1989). Thus $\mathcal{N}(\frac{\epsilon}{8}/L, \mathcal{F}, 2T)$ is at most in T's polynomial order.

4 Conclusion and Future Work

In this paper, we have studied the generalization ability of ERM algorithms for agent behavior prediction. In particular, we first develop a new technique that transforms MCRE to a higher-dimensional but more regular Markov chain and then give a uniform generalization bound based on the new Markov chain. As for the future work, we plan to investigate the joint learning problem of the optimal mechanism of the platform and the optimal prediction model of agent behaviors. Generalization analyses for these two cases will be even more challenging, and the corresponding results will also have more profound impact on adopting machine learning techniques in real-world interactive systems.

Notation	Meaning
$\mathbb{U}, \mathbb{B}, \mathbb{H}$	Respectively the random user factors space, agent behavior space and feedback space
u_t, b_t, h_t	Respectively the random users, agents joint behavior and feedback at round t
N	Number of agents in the system
η	Feedback function outputting feedbacks to all agents. η maps from $\mathbb{U} \times \mathbb{B}^N$ to $\mathbb{H}^N$
M_k	Agents joint behavior transition probability matrix under joint feedback k
$\mathbb{M}$	The new construucted higher dimensional chain, $\mathbb{M} = \{(h_t, b_t, b_{t+1}); t \geq 0\}$
Z	$\mathbb{M}$'s state number. We assume that M takes states from $[Z] = \{1, 2, ..., Z\}$
z_t	$\mathbb{M}'$'s state value at round t, $z_t = \{h_t, b_t, b_{t+1}\}$
M	Transition probability matrix of $\mathbb{M}$
π	$\mathbb{M}$'s stationary distribution
N_0	The elements of N_0-step transition probability matrix $M^{(N_0)}$ of $\mathbb{M}$ are all positive
δ	Minimum element of $M^{(N_0)}$
$\mathcal{F}$	Function class of behavior prediction functions, $\mathcal{F} \subset \{f : \mathbb{H}^N \times \mathbb{B}^N \to \mathbb{B}^N\}$
l	Loss function w.r.t. behavior prediction function f and data, e.g, $l(f, z_t) = 1_{f(h_t, b_t) \neq b_{t+1}}$
B	Upper bound for loss function l
$S, \tilde{S}$	T-round training data and ghost data sampled from $\mathbb{M}$ respectively, $S = (z_t)_{t=1}^T$, $\tilde{S} = (\tilde{z}_t)_{t=1}^T$
$err_S^T(f), err_\pi(f)$	Respectively denotes the T-round empirical loss on S and expected loss of an agent behavior prediction function f

Table 1: Notations

5 Acknowledgement

This work is partially supported by grant from the National Science Foundation for Distinguished Young Scholars of China (Grant No. 61325010). We thank Di He for his valuable suggestions on the detailed proof technique. Thanks to the AAAI anonymous reviewers for their comments to make the paper more exact and clear.

References

Bendavid, S.; Cesabianchi, N.; Haussler, D.; and Long, P. M. 1995. Characterizations of learnability for classes of (0,..., n)-valued functions. *Journal of Computer and System Sciences* 50(1):74–86.

Cary, M.; Das, A.; Edelman, B.; Giotis, I.; Heimerl, K.; Karlin, A. R.; Mathieu, C.; and Schwarz., M. 2007. Greedy bidding strategies for keyword auctions. In *EC '07 Proceedings of the 8th ACM conference on Electronic commerce.* ACM Press.

Cogburn, R. 1980. Markov chains in random environments: the case of markovian environments. *The Annals of Probability* 908–916.

Devroye, L. 1996. *A probabilistic theory of pattern recognition*, volume 31. springer.

Durrett, R. 2010. *Probability: theory and examples*, volume 3. Cambridge university press.

Edelman, B.; Ostrovsky, M.; and Schwarz, M. 2005. Internet advertising and the generalized second price auction: Selling billions of dollars worth of keywords. Technical report, National Bureau of Economic Research.

Glynn, P. W., and Ormoneit, D. 2002. Hoeffding's inequality for uniformly ergodic markov chains. *Statistics & probability letters* 56(2):143–146.

He, D.; Chen, W.; Wang, L.; and Liu, T.-Y. 2013. A game-theoretic machine learning approach for revenue maximization in sponsored search.

Natarajan, B. K. 1989. On learning sets and functions. *Machine Learning* 4(1):67–97.

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.

Tian, F.; Li, H.; Chen, W.; Qin, T.; Chen, E.; and Liu, T.-Y. 2014. Agent behavior prediction and its generalization analysis. Preprint, Arxiv preprint arXiv:1404.4960.

Vapnik, V. N. 1998. Statistical learning theory.

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.

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

Zhou, Y., and Lukose., R. 2007. Vindictive bidding in keyword auctions. In *ICEC '07 Proceedings of the ninth international conference on Electronic commerce.* ACM Press.