Finding One's Best Crowd: Online Learning By Exploiting Source Similarity

Yang Liu and Mingyan Liu
Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor
1301 Beal Avenue, Ann Arbor, Michigan 48109
{youngliu,mingyan}@umich.edu

Abstract

We consider an online learning problem (classification or prediction) involving disparate sources of sequentially arriving data, whereby a user over time learns the best set of data sources to use in constructing the classifier by exploiting their similarity. We first show that, when (1) the similarity information among data sources is known, and (2) data from different sources can be acquired without cost, then a judicious selection of data from different sources can effectively enlarge the training sample size compared to using a single data source, thereby improving the rate and performance of learning; this is achieved by bounding the classification error of the resulting classifier. We then relax assumption (1) and characterize the loss in learning performance when the similarity information must also be acquired through repeated sampling. We further relax both (1) and (2) and present a cost-efficient algorithm that identifies a *best crowd* from a potentially large set of data sources in terms of both classifier performance and data acquisition cost. This problem has various applications, including online prediction systems with time series data of various forms, such as financial markets, advertisement and network measurement.

Introduction

The ability to learn (classify or predict) accurately with sequentially arriving data has many applications. Examples include predicting future values on a prediction market, weather forecasting, TV ratings, and ad placement by observing user behavior. The subject of learning in such contexts has been extensively studied. Past literature is heavily focused on learning by treating each source or object's historical data separately, see e.g., [10, 13, 11] for single source multi-armed bandit problems for learning the best options of returned rewards, [9] for a support vector machine based forecasting for financial time series data, [6] for a model predicting spammers using a network's past statistics, and [8] for forecasting stock price index, among other.

More recent development has increasingly been focusing on improving learning through integrating data from multiple sources with similar statistics, see e.g., [7] for wind power prediction using both temporal and spatial information. The idea of increasing sample spaces by exploiting similarity proves to be helpful especially when the data arrives slowly, e.g., weather reports generated a few times per day. This idea naturally arises when different data sources are physically correlated, e.g., wind turbines on the same farm, or environmental monitoring sensors located within close proximity. However, it also fits well in the emerging context of crowdsourcing, where different sources (e.g., Amazon Mechanical Turks) contribute to a common data collection objective (e.g., labeling a set of images), and exploiting multiple data sources can improve the quality of crowdsourced data. For instance the idea of aggregating selectively data from a crowd to make prediction more accurate is empirically demonstrated and referred to as finding a "smaller but smarter crowd" in [4, 5].

In this paper we seek to make the notion of a "smarter" crowd quantitatively precise and develop methods to systematically identify and utilize this crowd. Specifically, we consider a problem involving K (potentially-)disparate data sources, each of which may be associated with a user. A given user can use its own data to achieve a certain learning (prediction, classification) objective but is interested in improving its performance by tapping into other data sources, and can request data from other sources at a cost. Accordingly, decisions need to be made judiciously on which sources of data should be used so as to optimize its learning accuracy. This implies two challenges: (1) we need to be able to measure the similarity/disparity between two sources in order to differentiate which sources are more useful toward the learning objective, and (2) we need to be able to determine the best set of sources given the measured similarity. Prior work most relevant to the present study is [3], where the problem of combining static IID data sources is analyzed. There are however a number of key differences: 1) in [3] the similarity information is assumed known a priori and the cost of obtaining data is not considered. 2) The results in [3] are established pre-collected IID data, while we focus on an online learning setting with Markovian data sources. In addition, the methodology we employ in this paper is quite different from [3] which draws mainly from VC theory [15], while our study is based on both VC theory and the multi-armed bandit (MAB) literature [2].

We will start by establishing bounds on the expected learning error under ideal conditions, including that (1) the similarity information between data sources is known a pri-

ori, and (2) data from all sources are available for free. We then relax assumption (1) and similarly establish the bounds on the error when such similarity information needs to be learned over time. We then relax both (1) and (2) and design an efficient online learning algorithm that simultaneously makes decisions on requesting and combining data for the purpose of training the predictor, and learning the similarity among data sources. We again show that this algorithm achieves a guaranteed performance uniform in time, and the additional cost with respect to the minimum cost required to achieve optimal learning rate diminishes in time. Moreover, the obtained bounds show clearly the trade-off between learning accuracy and the cost to obtain additional data. This provides useful information for system designers with different objectives. To our best knowledge this is the first study on online learning by exploiting source similarity with provable performance guarantees. Unless otherwise specified, all proofs can be found in the Appendices contained in supplementary materials.

Problem Formulation

Learning with multiple data sources

Consider K sources of data each associated with a unique user, indexed by $\mathcal{D} = \{1,2,\cdots,K\}$, which we also refer to as the whole crowd of sources. The sources need not be governed by identical probability distributions. Data samples arrive in discrete time to each user; the sample arriving at time t for user i is denoted by $z_i(t) = (x_i(t), y_i(t))$, $t = 1,2,\cdots$, with $x_i(t)$ denoting the features and $y_i(t)$ denoting the labels. At each time t, $x_i(t)$ is revealed first followed by a prediction on $y_i(t)$ made by the user, after which $y_i(t)$ is revealed and $z_i(t)$ is added to the training set. For simplicity of exposition, we will assume $x_i(t)$ to be a scalar; however our analysis easily extends to more complex forms of data, including batch arrivals. The objective of each user is to train a classifier to predict $y_i(t)$ using collected past data, and after prediction at time t, $y_i(t)$ will be revealed and can be used for training in the future steps. As a special case, when the target is to predict for future, $y_i(t)$ can be taken as $x_i(t+1)$. For analytical tractability we will further assume that the data arrival processes $\{x_i(t)\}_t$, $\forall i$, are mutually independently (but not necessarily identical), and each is given by a first order[1] finite-state positive recurrent Markov chain, with the corresponding transition probability matrix denoted by P^i on the state space $\mathcal{X}^i$ ($|\mathcal{X}^i| < \infty$). Denote by $P^i_{x,y}$ the transition probability from state x to y under P^i, and by π^i its stationary distribution on $\mathcal{X}^i$. For simplicity we will assume that $\mathcal{X}^1 = \mathcal{X}^2 = ... = \mathcal{X}^K = \mathcal{X}$, though this assumption can be easily relaxed, albeit with more cumbersome notation. The motivation for such modeling choice is by observing that for many applications the sequentially arriving data does not follow IID distribution as has been studied in the literature; consider e.g., weather conditions. Suppose labels $y_i(t) \in \mathcal{Y}^i$ and again for simplicity let us assume $\mathcal{Y}^1 = \mathcal{Y}^2 = ... = \mathcal{Y}^K = \mathcal{Y}$, and $|\mathcal{Y}| < \infty$. Denote $y^* := \max_{y\in\mathcal{Y}} |y|$.

[1] A high order extension is also straightforward.

For the classification job, a straightforward approach would be for each user i to build a classifier/predictor by using past observations of its own data up to time t: $\{z_i(1),...,z_i(t)\}$. Denote the classifier by f_i for user i, and a loss function $\mathcal{L}$ to measure the classification error. For instance $\mathcal{L}$ can be taken as the squared loss function $\mathcal{L}(f_i, z_i(t)) = [y_i(t) - f_i(x_i(t))]^2$. With the definition of loss function, the classification task for a user is to find the classifier that best fits its past observations:

$$f_i(t) = \text{argmin}_{f\in\mathcal{F}} \sum_{n=1}^{t} \mathcal{L}(f, z_i(n)) \,, \tag{1}$$

where we have used $\mathcal{F}$ to denote the set of all models of classifier (hypothesis space). For example, $\mathcal{F}$ could contain the classical linear regression models.

The idea we seek to explore in this paper is to construct the classifier f_i by utilizing similarity embedded among data sources, i.e., we ask whether f_i should be a function of all sources' past data and not just i's own, and if so how should such a classifier be constructed. Specifically, if we collect data from a set Ω_k of sources and use them as if they were from a single source, then the best classifier is given by

$$f_{\Omega_k}(t) = \text{argmin}_{f\in\mathcal{F}} \sum_{j\in\Omega_k} \sum_{n=1}^{t} \mathcal{L}(f, z_j(n)) \,. \tag{2}$$

It was shown in [3] that the expected error of the above classifier is bounded by a function of certain source similarity measures; the higher the similarity the lower the error bound.

Our interest is in constructing the best classifier for any given user i by utilizing other data sources. To do so we will need to measure the similarity or discrepancy between sources and to judiciously use data from the right set of sources. We will accomplish this by decomposing the problem into two sub-problems, the first is to use a similarity measure to determine a preferred set Ω_k^* to use, and the second is to construct the classifier using data from this set.

Pair-wise similarity between data sources

We first introduce the notion of cross-classification error, which is the expected loss when using classifier f_j (trained using source j's data) on user i's data and can be formally defined as $r_i(f_j) = E_i[\mathcal{L}(f_j, z_i)]$ where the expectation is with respect to user i's source data distribution. In principle, this could be used to measure the degree of similarity between two data sources i and j. However, this definition is not easy to work with as it involves a classifier that is only implicitly given in (1). Instead, we introduce a notion of similarity between two data sources i and j, that satisfies the following two conditions: (1) it can be obtained from the statistics of two respective data sources, and (2) it satisfies the following bound:

$$r_i(f_j) \le \beta_1(1 - S_{i,j}) + \beta_2 \,, \tag{3}$$

where $\beta_1, \beta_2 \ge 0$ are normalization constants and $0 \le S_{i,j} \le 1$ denotes the similarity measure; the higher this value the more similar two sources. The relationship captured in Eqn (3) between the error function and similarity can also take

on alternate forms; we adopt this simple linear relationship for simplicity of exposition. The following example shows the existence of such a measure.

Suppose for each user i, corresponding to each state/feature $x \in \mathcal{X}$, labels $y \in \mathcal{Y}$ is generated according a probability measure Q_x^i and denote each probability as $Q_{x,y}^i$ and $\sum_{y\in\mathcal{Y}} Q_{x,y}^i = 1$. Consider the following example. Take $\mathcal{L}$ as the squared loss and $S_{i,j}$ as:

$$S_{i,j} = 1 - \max_{x\in\mathcal{X},y\in\mathcal{Y}} |Q_{x,y}^i - Q_{x,y}^j|^2 \ . \tag{4}$$

Then we can show[2] that, by setting $\beta_1 := 2\sum_{y\in\mathcal{Y}} y^2$ and $\beta_2 := 2\sum_{x\in\mathcal{X}} \pi_x^i \cdot \sum_{y\in\mathcal{Y}} Q_{x,y}^i \cdot (\sum_{\hat{y}\in\mathcal{Y}} Q_{x,\hat{y}}^i \hat{y} - y)^2$, i.e. two times the intrinsic classification error with user i's own (perfect) data, which is independent with other sources j, the choice of $S_{i,j}$ satisfies both conditions. We note that the choice of such an S is not unique. For example, we could also take $S_{i,j}$ to be

$$S_{i,j} = 1 - \sum_{x\in\mathcal{X}} \pi_x^i \cdot \sum_{y\in\mathcal{Y}} |Q_{x,y}^i - Q_{x,y}^j|^2 \ ,$$

while setting $\beta_1 := 2(y^*)^2$. Later we will argue that an S that leads to a tighter bound can help achieve a better performance in classification. As it shall become clearer later when such similarity information needs to be estimated, the trade-off between selecting a tighter and looser similarity measure comes from the fact that tighter similarity may incur more learning error as it requires the evaluation of more terms.

Without loss of generality, for the remainder of our discussion we will focus on user 1. We will also denote $s_i := \min\{S_{1,i}, S_{i,1}\}, \forall i$. While the definition given in (4) is symmetric in i and j such that $S_{1,i} = S_{i,1}$, this needs not be true in general under alternate definitions of similarity. Note that $s_1 = 1$. We will then relabel the users in decreasing order of their similarity to user 1: $1 = s_1 \geq s_2 .... \geq s_K \geq 0$.

Solution with Complete Information

As mentioned earlier, the problem of finding the best set of data sources to use and that of finding the best classifier given this set are inherently coupled and strictly speaking need to be jointly optimized, resulting in significant challenges. The approach we take in this paper is as follows. We will first derive an upper bound on the error of the classifier given in (2) when applied to user 1, for a set of k independent Markov sources; this bound is shown to be a function of k and their similarity with user 1. This bound is then optimized to obtain the best set. Below we derive this upper bound assuming (1) the similarity information is known and (2) data is free, i.e., at time t all past and present samples from all sources are available to user 1.

Upper bounding the learning error

First notice we have the following convergence results for positive recurrent Markov Chain we consider in the current paper [14],

$$||\tilde{\pi}^i(t) - \pi^i||_{\text{TV}} \leq C_{\text{MC}} \cdot (\lambda_2^i)^t \ ,$$

where C_{MC} is some positive constant, $\tilde{\pi}_x^i(t)$ is the expected empirical distribution of state x for data source i's Markov chain upto time t for user i and π_x^i denotes its stationary distribution, and $0 < \lambda_2^i < 1$ is the second largest eigenvalue which specifies the mixing speed of the process. The total variation distance $||p-q||_{\text{TV}}$ between two probability measures p and q that are defined on $\mathcal{X}$ is defined as follows

$$||p-q||_{\text{TV}} := \max_{S\in 2^{\mathcal{X}}} |\sum_{x\in S} (p(x) - q(x))| \ . \tag{5}$$

Denote $\rho_{k(t)}(t) := \max\mathcal{L} \cdot C_{\text{MC}} \frac{\sum_{i\in k(t)} (\lambda_2^i)^t}{|k(t)|}$, where $\max\mathcal{L}$ is the maximum value attained by the loss function. Throughout the paper we denote $[k] := \{1, 2, ..., k\}$ as the ordered and continuous set up to k, and $k(t)$ for any other un-ordered set invoked at time t and use $|k(t)|$ to denote its size. For squared loss function we have the following results:

Theorem 1. *At time t, with probability at least $1 - O(\frac{1}{t^2})$ the error of a classifier $f_{[k]}(t)$ constructed using data from k sources of similarity $s_i, i \in k(t)$ can be bounded as*

$$r_1(f_{k(t)}(t)) \leq \underbrace{4\min_{f\in\mathcal{F}} r_1^{\text{IID}}(f) + 6\beta_2}_{\text{Term 1}} + \underbrace{6\beta_1 \frac{\sum_{i\in k(t)}(1-s_i)}{|k(t)|}}_{\text{Term 2}}$$
$$+ \underbrace{\rho_{k(t)}(t)}_{\text{Term 3}} + \underbrace{8y^*(2\sqrt{2d} + y^*)\sqrt{\frac{\log|k(t)|t}{|k(t)|t}}}_{\text{Term 4}} \ , \tag{6}$$

where d is the VC dimension for $\mathcal{F}$, and $r_1^{\text{IID}}(f)$ is the expected prediction error when the data are generated according to an IID process.

Denote the upper bound for $r_1(\cdot)$ in Eqn. (6) with set $k(t)$ of data sources (after ordering based on their similarity with user 1) at time t by $\mathcal{U}_{k(t)}(t)$. The results may be viewed as an extension to the previous one from [3] where static and IID data sources were considered. This upper bound can serve as a good guide for the selection of such a set and in particular the best choice of $|k(t)|$ given estimated values of s_i's. Note that Terms 1 is independent of this selection and it is a function of the baseline error of the classification problem, Term 2 is due to the integration of disparate data sources, Term 3 comes from the mixing time of a Markov source, and Term 4 arises from imperfect estimation and decision using a finite number of samples ($|k(t)|t$ samples up to time t).

Below we first point out the key steps in the proof that differ from that in [3] (full proof is in the supplementary materials), and then highlight the properties of this bound.

Main steps in the proof Our analysis starts with connecting Markovian data sources to IID sources so that the classical VC theory [15] and corresponding results can apply. The idea is rather simple: by the ergodicity assumption on the arrival process, the estimation error converges to that of IID data sources as shown in [1]. In particular, we can bound the difference in error when applying a predictor $f \in \mathcal{F}$ to a Markovian vs. an IID source (with distribution being the

[2] Please refer to supplementary materials.

same as the steady state distribution of the Markov chain) at time t, constructed with available data as follows:

$$
\begin{aligned}
&|r_i(f(t)) - r_i^{\text{IID}}(f(t))| \\
&= |\sum_{x\in\mathcal{X}} \tilde{\pi}_x^i E_{y\sim\mathcal{Y}}[\mathcal{L}(f(t),(x,y))] - \sum_{x\in\mathcal{X}} \pi_x^i E_{y\sim\mathcal{Y}}[\mathcal{L}(f(t),(x,y))]| \\
&\leq \max \mathcal{L} \cdot C_{\text{MC}}(\lambda_2^i)^t \ .
\end{aligned}
$$

We impose α-triangle inequality on the error function $\forall i,j,k$, of the corresponding data sources $r_i(f_j) \leq \alpha \cdot [r_i(f_k) + r_k(f_j)]$, where $\alpha \geq 1$ is a constant. When $\mathcal{L}$ is the squared loss function, we have $\alpha = 2$, following Jensen's inequality. Then $\forall f$

$$\frac{r_1(f)}{k} \leq \frac{\alpha \cdot [r_1(f_i) + r_i(f)]}{k} \ .$$

Sum over all $i \in k(t)$ we have

$$r_1(f) \leq \frac{\alpha\beta_1}{|k(t)|} \cdot \sum_{i\in k(t)} (1-s_i) + \alpha\beta_2 + \alpha \cdot \bar{r}_{k(t)}(f) \ ,$$

where $\bar{r}_{k(t)}(f) = \frac{\sum_{i\in k(t)} r_i(f)}{|k(t)|}$ is the average regret by applying f onto the $|k(t)|$ data sources. Due to the bias of mixing time for Markovian sources we have the following fact :

$$\bar{r}_{k(t)}(f) \leq \bar{r}_{k(t)}^{\text{IID}}(f) + \rho_{k(t)}(t).$$

The rest of the proof focuses on bounding $\bar{r}_{k(t)}^{\text{IID}}(f)$, i.e., the expected prediction error on IID data sources, which is similar in spirit to that presented in [3].

Properties of the error bound The upper bound $\mathcal{U}_{k(t)}(t)$ has the following useful properties.

Proposition 2. *For sources ordered in decreasing similarity $s_1 \geq s_2 \cdots$, $\frac{\sum_{i=1}^k s_i}{k}$ is non-increasing in k.*

This is straightforward to see by noting that

$$\frac{\sum_{i=1}^{k+1} s_i}{k+1} - \frac{\sum_{i=1}^k s_i}{k} = -\sum_{i=1}^k \frac{s_i}{k(k+1)} + \frac{s_{k+1}}{k+1} = \sum_{i=1}^k \frac{s_{k+1} - s_i}{k(k+1)} \leq 0 \ .$$

Terms 3 and 4 both decrease in time. While Term 4 converges at the order of $O(1/\sqrt{t})$, Term 3 converges with geometric rate, which is much faster than Term 4 and can be ignored for now. We then know because of the use of multiple sources, Term 4 decrease $|k(t)|$ times faster, leading to a better bound. This shows how the use of multiple sources fundamentally changes the behavior of the error bound.

The upper bound also suggests that the optimal selection is always to choose those with the highest similarity, which leads to a linear search for the optimal number k. Based on above discussions, the trade-off comes from the fact a larger k returns a smaller average similarity term $\sum_{i=1}^k s_i/k$ (and thus a larger $\sum_{i=1}^k (1-s_i)/k$), while with more data we have a faster convergence of Term 4. Define the optimal set of sources at time t as the one minimizing the bound $\mathcal{U}_{k(t)}(t)$, and denote it by $k^*(t)$. We then have the following fact,

Proposition 3. *When $\{s_i\}_{i\in\mathcal{D}}$ is known, $\exists\, t_o$, such that $\forall t \geq t_o$, if $i \in k^*(t)$ then $i \in k^*(n), \forall t_o \leq n \leq t$.*

This implies that if a data source is similar enough to be included at t, then it would have been included in previous time steps as well except for a constant number of times. This also motivates us to observe a threshold or phase transitioning phenomenon in selecting each user's best crowd. This result is also crucial in proving Theorem 6 where it helps establish bounded number of missed sampling for an optimal data source in an adaptive algorithm.

Proposition 4. *A set of tighter similarity measures S returns better worst case performance.*

Consider two such similarity measures s' and s with $s'_i \geq s_i$ (with at least one strict inequality). Suppose at any time t and optimal set of crowd for s is $k(t)$, then simply by selecting $k(t)$ for s' we achieve a better worst case performance (a smaller $\sum_{i=1}^k (1-s_i)/k$ in upper bound).

Overhead of Learning Similarity

As we show in the previous section, once the optimal set of data sources is determined, the classification/prediction performance is bounded. However in a real crowdsourcing system, neither of the two assumptions may be valid. In this section we relax the first assumption and consider a more realistic setting where the similarity information remains unknown a-priori and can only be learned through shared data. In this regards we need to estimate the similarity information $\{s_i\}_{i\neq 1}$ while making decision of which set of data sources to use.

The learning process works in the following way. At step t, we first estimate similarity $\tilde{s}_i$ according to the following:

$$\tilde{s}_i = 1 - \max_{x\in\mathcal{X}, y\in\mathcal{Y}} |\tilde{Q}_{x,y}^i(t) - \tilde{Q}_{x,y}^1(t)|^2 \ ,$$

where $\tilde{Q}_{x,y}^i(t) := \frac{n_{i,x\to y}(t)}{n_{i,x}(t)}$ are the estimated transition probability matrices with $n_{i,x}(t)$ denoting the number of times user i is sampled to be in state $x \in \mathcal{X}$ up to time t and $n_{i,x\to y}(t)$ denoting the number of observed samples from data source i being in (x,y). Different from the previous Section, now since $\{s_i\}_{i\neq 1}$ is unknown, in order to select data sources, the estimate of the upper bound $\mathcal{U}_{k(t)}(t)$ becomes a function of $\{\tilde{s}_i\}$: $\mathcal{U}_{k(t)}(t;\{\tilde{s}_i\}_{i\in k(t)})$, which is obtained by simply substituting all s terms in $\mathcal{U}_{k(t)}(t)$ with $\tilde{s}$. Denote the terms that are being affected by choosing set $k(t)$ in $\mathcal{U}_{k(t)}(t;\{\tilde{s}_i\}_{i\in k(t)})$ as follows:

$$\tilde{\mathcal{U}}_{k(t)}^{tr}(t) = 6\beta_1 \frac{\sum_{i\in k(t)}(1-\tilde{s}_i)}{k} + 8y^*(2\sqrt{2d} + y^*)\sqrt{\frac{\log|k(t)|t}{|k(t)|t}} \ .$$

Note we are omitting $\rho_{k(t)}(t)$ as it is on a much smaller order and will not affect our results order-wise.

Then the learning algorithm first orders all data sources according to $\{\tilde{s}_i\}$. And then choses $\tilde{k}^*(t)$ by a linear search such that

$$\tilde{k}^*(t) = \arg \max_{[k], 1\leq k\leq K} \tilde{\mathcal{U}}_{[k]}^{tr}(t) \ .$$

We have the following results.

Theorem 5. *At time t, with probability at least $1-O(\frac{1}{t^2})$ the error of trained classifier $f_{\tilde{k}^*(t)}(t)$ using $\tilde{k}^*(t)$ data sources can be bounded as follows*

$$r_1(f_{\tilde{k}^*(t)}(t)) \leq \mathcal{U}_{k^*(t)}(t) + O(\sqrt{\frac{\log t}{t}}) \,. \tag{7}$$

Clearly from above results we see there is an extra $O(\sqrt{\log t/t})$ term capturing the loss of learning the similarity information.

A Cost-efficient Algorithm

Now we relax the second restriction on data acquisition. In reality data acquisition from other sources are costly. In our study, we explicitly model this aspect whereby at each time step a user may request data from another user at a unit cost of c. This modeling choice not only reflects reality, but also allows us to examine the tradeoff between a user's desire to to keep its overall cost low while keeping its prediction performance high. We present a cost-efficient algorithm with performance guarantee. As one may expect, with less data the prediction accuracy will degrade. But the number of unnecessary data will also be bounded from above.

A cost-efficient online algorithm

Denote by $n_i(t)$ the number of collected samples from source i up to time t and $N_{k(t)}(t) = \sum_{i\in k(t)} n_i(t)$. Notice in this section $n_i(t) \neq t$ in general. Denote $D(t) := O(t^z)$; z will be referred to as the exploration constant satisfying $0 < z < 1$. Later we will show how z controls the trade-off between data acquisition and classification accuracy. Again denote by $n_{i,x}(t)$ the number of times user i is sampled to be in state $x \in \mathcal{X}$ up to time t and construct the following set at each time t :

$$O(t) = \{i : i \in \mathcal{D}, \exists\, x \in \mathcal{X}, n_{i,x}(t) < D(t)\} \,.$$

We name the algorithm as `K-Learning`, which consists mainly of the following two steps (run by user 1):

Exploration: At time t, if any data source has a state x that has been observed (from requested data) for less than $D(t)$ times, i.e., if $O(t)$ is non-empty, then the algorithm enters an exploration phase and collects data from *all* sources $k_2(t) = \mathcal{D}$ and predicts via its own data $k_1(t) = \{1\}$. The prediction at exploration phase is *conservative* since without enough sampling user 1 cannot be confident in calculating its optimal set of similar sources, in which case the user would rather limit itself to its own data.

Exploitation: If $O(t)$ is empty at time t then the algorithm enters an exploitation phase, whereby it first estimates similarity measures of all sources. For our analysis we will use the same definition given earlier: $\tilde{s}_i(t) = 1 - \max_{x\in\mathcal{X},y\in\mathcal{Y}} |\tilde{Q}^i_{x,y}(t) - \tilde{Q}^1_{x,y}(t)|^2$. The algorithm then calculates $k_1(t)$ using the estimated bound $\tilde{\mathcal{U}}^{tr}_{k_1(t)}(t)$, and uses data from this set $k_1(t)$ of sources for training the classifier, while requesting data from set $k_2(t)$, where $k_2(t)$ is set to be: $k_2(t) := \text{argmax}_{k'(t)\subseteq D}\{|k'(t)| :$

$$\tilde{\mathcal{U}}^{tr}_{k'(t)}(t) \in [\tilde{\mathcal{U}}^{tr}_{k_1(t)}(t) - \sqrt{\frac{\log t}{t^z}}, \tilde{\mathcal{U}}^{tr}_{k_1(t)}(t) + \sqrt{\frac{\log t}{t^z}}]\} \,.$$

Algorithm 1 `K-Learning`

Initialization:
Set $t = 1$ and similarity $\{\tilde{s}_i(1)\}_{i\in\mathcal{D}}$ to some value in $[0,1]$; $n_{i,x}(t) = 1$ for all i and x.
loop:
Calculate $O(t)$.
if $O(t) \neq \emptyset$ **then**
    Explores, sets $k_1(t) = \{1\}, k_2(t) = \mathcal{D}$.
else
    Exploit, orders data sources according to $\{\tilde{s}_i(t)\}_{i\in\mathcal{D}}$ and computes $k_1(t)$ that minimizes $\tilde{\mathcal{U}}^{tr}_{k_1(t)}(t)$, which is solved using the linear search property, and the current estimates $\{\tilde{s}_i(t)\}_{i\in\mathcal{D}}$. Set $k_2(t)$ as $k_2(t) := \text{argmax}_{k'(t)\subseteq D}\{|k'(t)| : \tilde{\mathcal{U}}^{tr}_{k'(t)}(t) \in [\tilde{\mathcal{U}}^{tr}_{k_1(t)}(t) - \sqrt{\frac{\log t}{t^z}}, \tilde{\mathcal{U}}^{tr}_{k_1(t)}(t) + \sqrt{\frac{\log t}{t^z}}]\}$.
end if
Construct classifier $f_{k_1(t)}$ using data collected from sources in $k_1(t)$. Request data from $k_2(t)$.
$t := t+1$ and update $\{n_{i,x}(t)\}_{i,x}$, $\{\tilde{s}_i(t)\}_{i\in\mathcal{D}}$ using collected samples.
goto *loop*.

Notice when calculating $k_2(t)$ we set a tolerance region (due to imperfect estimation of $\tilde{\mathcal{U}}^{tr}_{k_1(t)}(t)$) so that a sample data from an optimal data source will not be missed with high probability.

Performance of `K-Learning`

There are three types of error in the learning performance: (1) Error due to exploration, in which case the error comes from conservative training due to no enough sampling. Due to technical difficulties, we approximate the error (compared to the performance with optimal classifier) by the worst case performance loss, that is the performance difference in upper bounds. (2) Prediction error associated with incorrect computation of $k_1(t)$ (i.e., $k_1(t) \neq k^*(t)$) in exploitation due to imperfect estimates on $\{s_i\}_{i\neq 1}$. (3) Prediction error from sub-sampling effects. This is because even though under the case that $k_1(t) = k^*(t)$, i.e., $k^*(t)$ is correctly identified, due to incomplete sampling, $\exists\, i > 1, n_i(t) < t$, $\hat{\mathcal{U}}_{k_1(t)} \neq \mathcal{U}_{k_1(t)}$, where $\hat{\mathcal{U}}_{k_1(t)}$ is the upper bound for the classification error with collected data: this can be similarly derived following the proof of Theorem 1 and results in [3]:

$$\hat{\mathcal{U}}_{k(t)}(t) = 4\min_{f\in\mathcal{F}} r_1^{\widetilde{\text{IID}}}(f) + 6\beta_2 + 6\beta_1 \frac{\sum_{i\in k(t)} n_i(t)(1-s_i)}{N_{k(t)}(t)}$$
$$+ \tilde{\rho}_{k(t)}(t) + 8y^*(2\sqrt{2d} + y^*) \cdot \sqrt{\frac{\log N_{k(t)}(t)}{N_{k(t)}(t)}} \,,$$

where

$$\tilde{\rho}_{k(t)} := \max \mathcal{L} \cdot C_{\text{MC}} \frac{\sum_{i\in k(t)} (\lambda_2^i)^{n_i(t)}}{|k(t)|} \,,$$

and $\min_{f\in\mathcal{F}} r_1^{\widetilde{\text{IID}}}(f)$ is error rate over a biased data distribution due to incomplete sampling, compared to the tar-

get IID distribution. We emphasize that the difference between $\mathcal{U}_{k(t)}(t;\{\tilde{s}_i\}_{i\in k(t)})$ and $\hat{\mathcal{U}}_{k(t)}(t)$: $\mathcal{U}_{k(t)}(t;\{\tilde{s}_i\}_{i\in k(t)})$ is the estimation of upper bound $\mathcal{U}_{k(t)}(t)$ with estimated similarity information $\tilde{s}$, while $\hat{\mathcal{U}}_{k(t)}(t)$ bounds actual error of the learning task at each step. In $\mathcal{U}_{k(t)}(t)$ and $\mathcal{U}_{k(t)}(t;\{\tilde{s}_i\}_{i\in k(t)})$, full samples are assumed to have been collected for each data source in $k(t)$, i.e., $n_i(t)=t$. However this is not true for $\hat{\mathcal{U}}_{k(t)}(t)$, except for $n_1(t)$ the data source for user 1 itself. Also due to dis-continuous sampling for Markovian data, the sampled data distribution is biased which results in $\min_{f\in\mathcal{F}} \widetilde{r}_1^{\text{IID}}(f)$. The main gist of bounding this discrepancy is that due to Proposition 3 we are able to bound the missed samples for a data source appearing in the optimal set.

A subtle difference between the results in this section and the previous one is the performance of the classifier trained during an *exploration* phase is simply the one using user 1's own data, which is bounded away from the optimal performance bound (via data sources $k^*(t)$). Denote the worse case performance loss (difference in performance upper bound) in exploration phases upto time t as $R_e(t)$, that is

$$R_e(t)=\sum_{n=1}^{t}\mathbf{1}_{O(n)\neq\emptyset}\cdot|\,\mathcal{U}_{[1]}(t)-\mathcal{U}_{k^*(t)}(t)|\,. \tag{8}$$

This is a quantity we are interested in determining for exploration phases. For exploitation phases, we evaluate the prediction/classification performance as the ones with classifier $f_{k_1(t)}(t)$.

Theorem 6. *At time t,*

- *The number of exploration phases is bounded as follows,*

$$E[\sum_{n=1}^{t}\mathbf{1}_{O(n)\neq\emptyset}]\leq O(t^z)\,.$$

Further the per round performance loss due to exploration phases $\frac{E[R_e(t)]}{t}$ is bounded as follows: with probability being at least $1-O(e^{-Ct^z})$ where $C>0$ is a constant,

$$\frac{E[R_e(t)]}{t}\leq O(\sqrt{z\cdot\log t}\cdot t^{z/2-1})\,.$$

- *If t is an exploitation phase, with probability being at least $1-O(\frac{1}{t^2})$ we bound the average prediction error for classifier $f_{k_1(t)}(t)$ with data sources $k_1(t)$ as follows,*

$$r_1(f_{k_1(t)}(t))\leq\mathcal{U}_{k^*(t)}(t)+O(\sqrt{\frac{\log t}{t^z}})+O(\log t\cdot t^{-2/3})\,.$$

Note on the bound:

- $O(\sqrt{z\cdot\log t}\cdot t^{z/2-1})$ is the average error invoked by exploration. This term is diminishing with t, that is the average amount of exploration error is converging to 0. $O(\sqrt{\log t/t^z})$ is the learning error incurred in exploitation phases, which is in analogy to the $O(\sqrt{\log t/t})$ term as shown in the bound proved in Theorem 5. $O(\log t\cdot t^{-2/3})$ is also incurred in exploitation phases. This is a unique error term associated with sub-sampling of Markovian data: due to (1) missed sampling and (2) discontinuous sampling.
- It should be noted that the prediction error term $O(\sqrt{\log t/t^z})$ decrease with z for $0<z<1$. That is with a higher z, a tighter bound can be achieved. With $z\to 1$ (number of samples cannot go beyond t at time t), we can show the prediction error term converges to $O(\sqrt{\log t/t})$, which is consistent with the results we reported in last section. Also it worths pointing out $O(\log t\cdot t^{-2/3})$ is generally on a smaller order compared to $O(\sqrt{z\cdot\log t}\cdot t^{z/2-1})$ and $O(\sqrt{\log t/t^z})$: simply set z to be $z>2/3$.
- This observation also sheds lights on establishing the tightness of this bound for z close to 1, as $O(\sqrt{\log t/t})$ is the uniform convergence bound as proved in statistical learning theory [15].

Cost analysis

To capture the effectiveness of cost saving, we define the following difference in cost:

$$\text{Cost measure}\ :\ R_c(t)=c\sum_{n=1}^{t}\sum_{i=1}^{K}\mathbf{1}_{i\notin k^*(n),i\in k_2(n)}\,.$$

$R_c(t)$ will be referred to as the cost measure, which quantifies the amount of data requests from non-optimal data sources. We have the following main results.

Theorem 7. *At time t, we have*

$$E[R_c(t)]\leq O(ct^z)\,.$$

Notes on the bound:

- First of all note that $E[R_c(t)]=o(t)$ when $t<1$ and thus $E[R_c(t)]/t\to 0$ as $t\to\infty$. This demonstrates the cost saving property of our algorithm as the average number of redundant data request is converging to 0.
- Clearly z controls the trade-offs between prediction accuracy $r_1(f_{k_1(t)}(t))$ and data acquisition cost regret $E[R_c(t)]$. A higher z leads to a more frequent sampling scheme and thus higher cost regret, while with a small z the sampling is conservative which leads to higher prediction error.

Acknowledgement

This material is based on research sponsored by the NSF under grant CNS-1422211 and the Department of Homeland Security (DHS) Science and Technology Directorate, Homeland Security Advanced Research Projects Agency (HSARPA), Cyber Security Division (DHS S&T/HSARPA/CSD), BAA 11-02 via contract number HSHQDC-13-C-B0015.

Conclusion

In this paper we consider a problem of finding best set of data for each user to enhance its online learning (be it a classification or prediction problem) performance when facing

disparate sources of sequentially arriving samples. We first establish learning error when similarity information among users are known and data can be collected without cost. We then extend the results to the case when such information is unknown a priori. Lastly we propose and analyze a cost-efficient algorithm to help users adaptively distinguish between similar and dis-similar data sources. and aggregate and request data appropriately for the purpose of training predictor and saving budget. We establish its performance guarantee and show the algorithm helps avoid requesting redundant data from sources that are helpless (or even harmful) and thus saves cost.

References

[1] Adams, T. M., and Nobel, A. B. 2010. Uniform convergence of vapnikchervonenkis classes under ergodic sampling. *The Annals of Probability* 38(4):1345–1367.

[2] Auer, P.; Cesa-Bianchi, N.; and Fischer, P. 2002. Finite-time analysis of the multiarmed bandit problem. In *Machine learning*, volume 47, 235–256. Springer.

[3] Crammer, K.; Kearns, M.; and Wortman, J. 2008. Learning from multiple sources. *The Journal of Machine Learning Research* 9:1757–1774.

[4] Galesic, M., and Barkoczi, D. 2014. Wisdom of small crowds for diverse real-world tasks. In *Available at SSRN 2484234*.

[5] Goldstein, D. G.; McAfee, R. P.; and Suri, S. 2014. The wisdom of smaller, smarter crowds. In *Proceedings of the fifteenth ACM conference on Economics and computation*, 471–488. ACM.

[6] Hao, S.; Syed, N. A.; Feamster, N.; Gray, A. G.; and Krasser, S. 2009. Detecting spammers with snare: Spatio-temporal network-level automatic reputation engine. In *Presented as part of the 18th USENIX Security Symposium (USENIX Security 09)*. Montreal, Canada: USENIX.

[7] He, M.; Yang, L.; Zhang, J.; and Vittal, V. A spatio-temporal analysis approach for short-term forecast of wind farm generation. *IEEE Trans. Power Syst*.

[8] Hyup Roh, T. 2007. Forecasting the volatility of stock price index. In *Expert Systems with Applications*, volume 33, 916–922. Elsevier.

[9] Kim, K.-j. 2003. Financial time series forecasting using support vector machines. In *Neurocomputing*, volume 55, 307–319. Elsevier.

[10] Lai, T. L., and Robbins, H. 1985. Asymptotically Efficient Adaptive Allocation Rules. In *Advances in Applied Mathematics*, volume 6, 4–22.

[11] Langford, J., and Zhang, T. 2007. The Epoch-Greedy Algorithm for Multi-armed Bandits with Side Information. In *NIPS*.

[12] Lezaud, P. 1998. Chernoff-type bound for finite markov chains. *The Annals of Applied Probability* 8(3):849–867.

[13] Lu, T.; Pl, D.; and Pal, M. 2010. Contextual multi-armed bandits. In *Journal of Machine Learning Research*, volume 9, 485–492.

[14] Rosenthal, J. S. 1995. Convergence rates for markov chains. *SIAM Review* 37(3):pp. 387–405.

[15] Vapnik, V. N. 1995. *The Nature of Statistical Learning Theory*. New York, NY, USA: Springer-Verlag New York, Inc.

Appendices

Example of S

We show $S_{i,j} = 1 - \max_{x\in\mathcal{X},y\in\mathcal{Y}} |Q^i_{x,y} - Q^j_{x,y}|^2$ while setting $\beta_1 := 2\sum_{y\in\mathcal{Y}} y^2$ and $\beta_2 := 2\sum_{x\in\mathcal{X}} \pi^i_x \cdot \sum_{y\in\mathcal{Y}} Q^i_{x,y} \cdot (\sum_{\hat{y}\in\mathcal{Y}} Q^i_{x,\hat{y}}\hat{y} - y)^2$ is a feasible similarity measure according to our definition. For squared loss the optimal predictor is given by the conditional expectation; we thus have the following:

$$
\begin{aligned}
r_i(f_j) &= \sum_{x\in\mathcal{X}} \pi^i_x \cdot \sum_{y\in\mathcal{Y}} Q^i_{x,y} \cdot (\sum_{\hat{y}\in\mathcal{Y}} Q^j_{x,\hat{y}}\hat{y} - y)^2 \\
&= \sum_{x\in\mathcal{X}} \pi^i_x \cdot \sum_{y\in\mathcal{Y}} Q^i_{x,y} \cdot (\sum_{\hat{y}\in\mathcal{Y}} Q^j_{x,\hat{y}}\hat{y} - \sum_{\hat{y}\in\mathcal{Y}} Q^i_{x,\hat{y}}\hat{y} + \sum_{\hat{y}\in\mathcal{Y}} Q^i_{x,\hat{y}}\hat{y} - y)^2 \\
&\le 2\sum_{x\in\mathcal{X}} \pi^i_x \cdot \sum_{y\in\mathcal{Y}} Q^i_{x,y} \cdot (\sum_{\hat{y}\in\mathcal{Y}} Q^j_{x,\hat{y}}\hat{y} - \sum_{\hat{y}\in\mathcal{Y}} Q^i_{x,\hat{y}}\hat{y})^2 \\
&\quad + 2\sum_{x\in\mathcal{X}} \pi^i_x \cdot \sum_{y\in\mathcal{Y}} Q^i_{x,y} \cdot (\sum_{\hat{y}\in\mathcal{Y}} Q^i_{x,\hat{y}}\hat{y} - y)^2 \\
&\le 2\sum_{y\in\mathcal{Y}} y^2 \cdot (1 - S_{i,j}) + \beta_2 \ .
\end{aligned}
$$

□

To complete proof of Theorem 1

As we already bind $r_1(\cdot)$ with $\bar{r}^{\mathrm{IID}}_{k(t)}(\cdot)$, we only need to bound $\bar{r}^{\mathrm{IID}}_{k(t)}(\cdot)$ and consider the case with IID data and the expected prediction error at time t when combine with data from sources $k(t)$ for training. Denote $R_{|k(t)|t}(\mathcal{F})$ as the Rademacher complexity of space $\mathcal{F}$ with $|k(t)|t$ samples and f^* the optimal classifier trained on the set of data. Since we only have finite number of samples we first have the following lemma:

Lemma 8. *With probability being at least* $1 - \frac{2}{(|k(t)|t)^2}$,

$$
\bar{r}^{IID}_{k(t)}(f^*) \le \min_{f\in\mathcal{F}} \bar{r}^{IID}_{k(t)}(f) + 8y^* R_{|k(t)|t}(\mathcal{F}) + 8(y^*)^2 \sqrt{\frac{\log(|k(t)|t)}{|k(t)|t}} \ .
$$

The proof is standard following the VC theory and it can be derived from the results reported in [3]. Further we know from [3], for squared loss function we have $R_{|k(t)|t}(\mathcal{F}) \le 2\sqrt{\frac{2d}{|k(t)|t} \cdot \log(\frac{2e|k(t)|t}{d})}$. Therefore

$$
\begin{aligned}
\bar{r}^{\mathrm{IID}}_{k(t)}(f^*) &\le \min_{f\in\mathcal{F}} \bar{r}^{\mathrm{IID}}_{k(t)}(f) + 8(y^*)^2 \sqrt{\frac{\log(|k(t)|t)}{|k(t)|t}} + 16y^* \cdot \sqrt{\frac{2d}{|k(t)|t} \cdot \log(\frac{2e|k(t)|t}{d})} \\
&\approx \min_{f\in\mathcal{F}} r^{\mathrm{IID}}_{k(t)}(f) + 8y^*(2\cdot\sqrt{2d} + y^*) \cdot \sqrt{\frac{\log(|k(t)|t)}{|k(t)|t}} \ ,
\end{aligned}
$$

when t is sufficiently large (to thus ignore the $\log 2e$ in term $\sqrt{\frac{2d}{|k(t)|t} \cdot \log(\frac{2e|k(t)|t}{d})}$). Then

$$
\begin{aligned}
\bar{r}^{\mathrm{IID}}_1(f^*) \le 2 \cdot \min_{f\in\mathcal{F}} \bar{r}^{\mathrm{IID}}_{k(t)}(f) + 2\beta_2 + \frac{2\beta_1}{|k(t)|} \cdot \sum_{i\in k(t)} (1 - s_i) \\
+ 8y^*(2\cdot\sqrt{2d} + y^*) \cdot \sqrt{\frac{\log(|k(t)|t)}{|k(t)|t}} \ .
\end{aligned}
$$

The last two terms are clear. In particular the 3rd term is a constant brought in by the disparities between data sources while the 4th term is the bias with finite number of samplings. Consider the 1st term we have,

$$
\begin{aligned}
\min_{f\in\mathcal{F}} \bar{r}^{\mathrm{IID}}_{k(t)}(f) &\le \min_{f\in\mathcal{F}} \frac{\sum_{i\in k(t)} \alpha \cdot [r^{\mathrm{IID}}_i(f_1) + r^{\mathrm{IID}}_1(f)]}{|k(t)|} \\
&\le \frac{2\beta_1}{|k(t)|} \cdot \sum_{i\in k(t)} (1 - s_i) + 2\beta_2 + \cdot \min_{f\in\mathcal{F}} r^{\mathrm{IID}}_1(f) \ .
\end{aligned}
$$

Plug back the results we establish the theorem. □

Proof of Proposition 3

Suppose $i \in k^*(t)$ and there exists a $n < t$ such that $i \notin k^*(n)$. First consider the following fact: let $0 < \delta < 1$ we have

$$|\sqrt{\frac{\log \delta t}{\delta t}} - \sqrt{\frac{\log t}{t}}| = \frac{1}{\sqrt{\log t} + \sqrt{\frac{\log t + \delta}{\delta}}} \cdot \frac{|(1-1/\delta)\log t - \delta|}{\sqrt{t}} \ .$$

Easy to see the first term is strictly decreasing. For the second term since $\sqrt{t}$ is of a higher order compared with $\log t$ we expect this term to be decreasing when t passes certain threshold. Since $i \in k^*(t)$ and $i \notin k^*(n)$ and the fact we proved earlier that the optimal selection is always a continuous group we know $|k^*(n)| < |k^*(t)|$ and denote $\delta := |k^*(n)|/|k^*(t)|$. Therefore reducing $k^*(t)$ to $k^*(n)$ will return a better strategy for time t : compared with time n, the loss from the term $\sqrt{\frac{\log t}{t}}$ to $\sqrt{\frac{\log \delta \cdot t}{\delta \cdot t}}$ is smaller, while the gain in average similarity is the same. Similar arguments hold for the term $(\lambda_2^i)^t$, which is also strictly decreasing with t. Proved. □

Proof of Theorem 5

In order to prove the results, we analyze the error of mis-calculating $k^*(t)$.

Error in ordering data sources

We have two steps towards calculating $k^*(t)$ in our algorithm we first need to order data sources $\{1, 2, ..., K\}$ in their similarity to user 1 to invoke the linear search. For simplicity of following analyses we assume $s_1 > s_2 > ... > s_K$, and denote by $\Delta_{\min} = \min_{i,j} |s_i - s_j|$. The error of mis-ordering at time t is bounded by the following event

$$P(\text{mis-ordering at time } t) \leq P(\omega_m(t)) \ ,$$

where

$$\omega_m(t) = \{\exists i : |\tilde{s}_i - s_i| \geq \frac{\Delta_{\min}}{2}\} \ .$$

this is easy to verify : otherwise the inaccurate measurements are not enough to leverage a sub-optimal option.

Then

$$\begin{aligned}
P(\omega_m(t)) &\leq \sum_{i=1}^{K} P(|\tilde{s}_i - s_i| \geq \frac{\Delta_{\min}}{2}) \\
&= \sum_{i \leq K} P(\max_{x,y} ||\tilde{Q}^i_{x,y} - \tilde{Q}^1_{x,y}|^2 - |Q^i_{x,y} - Q^1_{x,y}|^2| \geq \frac{\Delta_{\min}}{2}) \\
&\leq \sum_{i \leq K} \sum_x \sum_y P(|\tilde{Q}^i_{x,y} - Q^i_{x,y}| + |\tilde{Q}^1_{x,y} - Q^1_{x,y}| \geq \frac{\Delta_{\min}}{4}) \ ,
\end{aligned}$$

where the first inequality is due to union bound and the last inequality comes from the following fact

$$\begin{aligned}
&\left| |\tilde{Q}^i_{x,y} - \tilde{Q}^1_{x,y}|^2 - |Q^i_{x,y} - Q^1_{x,y}|^2 \right| \\
&= \left| |\tilde{Q}^i_{x,y} - \tilde{Q}^1_{x,y}| + |Q^i_{x,y} - Q^1_{x,y}| \right| \cdot \left| |\tilde{Q}^i_{x,y} - \tilde{Q}^1_{x,y}| - |Q^i_{x,y} - Q^1_{x,y}| \right| \\
&\leq 2 \left| |\tilde{Q}^i_{x,y} - \tilde{Q}^1_{x,y}| - |Q^i_{x,y} - Q^1_{x,y}| \right| \ .
\end{aligned}$$

Consider $\left| |\tilde{Q}^i_{x,y} - \tilde{Q}^1_{x,y}| - |Q^i_{x,y} - Q^1_{x,y}| \right|$.

$$\begin{aligned}
&|\tilde{Q}^i_{x,y} - \tilde{Q}^1_{x,y}| - |Q^i_{x,y} - Q^1_{x,y}| \\
&= |\tilde{Q}^i_{x,y} - Q^i_{x,y} + Q^i_{x,y} - \tilde{Q}^1_{x,y}| - |Q^i_{x,y} - Q^1_{x,y}| \\
&\leq |\tilde{Q}^i_{x,y} - Q^i_{x,y}| + |Q^i_{x,y} - \tilde{Q}^1_{x,y}| - |Q^i_{x,y} - Q^1_{x,y}| \\
&\leq |\tilde{Q}^i_{x,y} - Q^i_{x,y}| + |Q^i_{x,y} - \tilde{Q}^1_{x,y} - Q^i_{x,y} + Q^1_{x,y}| \\
&= |\tilde{Q}^i_{x,y} - Q^i_{x,y}| + |Q^1_{x,y} - \tilde{Q}^1_{x,y}| \qquad (9)
\end{aligned}$$

and moreover we have

$$\begin{aligned}
&|\tilde{Q}^i_{x,y} - \tilde{Q}^1_{x,y}| - |Q^i_{x,y} - Q^1_{x,y}| \\
&= |\tilde{Q}^i_{x,y} - \tilde{Q}^1_{x,y}| - |Q^i_{x,y} - \tilde{Q}^1_{x,y} + \tilde{Q}^1_{x,y} - Q^1_{x,y}| \\
&\geq |\tilde{Q}^i_{x,y} - \tilde{Q}^1_{x,y}| - |Q^i_{x,y} - \tilde{Q}^1_{x,y}| - |\tilde{Q}^1_{x,y} - Q^1_{x,y}| \\
&\geq -|\tilde{Q}^i_{x,y} - \tilde{Q}^1_{x,y} - Q^i_{x,y} + \tilde{Q}^1_{x,y}| - |\tilde{Q}^1_{x,y} - Q^1_{x,y}| \\
&= -|\tilde{Q}^i_{x,y} - Q^i_{x,y}| - |Q^1_{x,y} - \tilde{Q}^1_{x,y}|
\end{aligned} \tag{10}$$

From above two inequality we know

$$\left| |\tilde{Q}^i_{x,y} - \tilde{Q}^1_{x,y}| - |Q^i_{x,y} - Q^1_{x,y}| \right| \leq |\tilde{Q}^i_{x,y} - Q^i_{x,y}| + |\tilde{Q}^1_{x,y} - Q^1_{x,y}| \tag{11}$$

Again via union bound we have

$$\begin{aligned}
&P(|\tilde{Q}^i_{x,y} - Q^i_{x,y}| + |\tilde{Q}^1_{x,y} - Q^1_{x,y}| \geq \frac{\Delta_{\min}}{4}) \\
&\leq P(|\tilde{Q}^i_{x,y} - Q^i_{x,y}| \geq \frac{\Delta_{\min}}{8}) + P(|\tilde{Q}^1_{x,y} - Q^1_{x,y}| \geq \frac{\Delta_{\min}}{8}) \, .
\end{aligned} \tag{12}$$

Next we prove for each $i \in \mathcal{D}$, $n_{i,x}(t) = O(t)$ w.h.p. We invoke the following result, Theorem 3.3 from [12].

Lemma 9. *For finite-state, irreducible Markov chain $X_i(t), t = 1, 2, \ldots$ with state space $\mathcal{X}^i$ and transition probability P^i, initial distribution q^i and stationary distribution π^i, denote $N^i_q = ||(\frac{q_x}{\pi_x})_x||_2$. Let $\hat{P}^i = P^{i,T} \cdot P^i$ be the multiplicative symmetrization of P^i where $P^{i,T}$ is the adjoint of P^i on $l_2(\pi)$. Let $\kappa = 1 - \lambda_2$ where λ_2 is the second largest eigenvalue of the matrix $\hat{P}^i$. ε is often referred to as the eigenvalue gap of $\hat{P}^i$. Let $f : \mathcal{X}^i \to R$ be such that $\sum_{y \in \mathcal{X}^i} \pi^i_y \cdot f(y) = 0$, $||f||_\infty \leq 1$ and $0 < ||f||_2^2 \leq 1$. For any positive integer n and $0 < \gamma \leq 1$ we have*

$$P(\frac{\sum_{t=1}^n f(X_i(t))}{n} \geq \gamma) \leq N_q \cdot e^{-\frac{n\gamma^2\kappa}{28}} \, . \tag{13}$$

Let $f(X_i(t)) = -\mathbf{1}_{X_i(t)=x} + \pi^i_x$. Not hard to verify such f satisfies all conditions required in Lemma 9. Then we have

$$P(n_{i,x}(t) \leq (\pi^i_x - \gamma)t) \leq N_q \cdot e^{-\frac{t\gamma^2\kappa}{28}} \, , \tag{14}$$

which holds specifically for a constant $\gamma < \pi^i_x$.

With above (by Chernoff-Hoeffding bound),

$$P(|\tilde{Q}^i_{x,y} - Q^i_{x,y}| > 2\frac{\Delta_{\min}}{8}) \leq 2e^{-2\frac{\Delta^2_{\min}}{16} \cdot O(t)} \, .$$

Thus

$$P(\omega_m(t)) \leq O(e^{-Ct}) \, , \tag{15}$$

for a certain constant C.

Error in finding the best crowd

Denote the estimated error bound as $\tilde{\mathcal{U}}(t)$:

$$\tilde{\mathcal{U}}_{k(t)}(t) := \mathcal{U}_{k(t)}(t; \{\tilde{s}_i\}_{i \in k(t)}) \, ,$$

Now consider we are with correct ordering of all sources. We then further consider the following two events at step t:

$$\omega_1(t) = \{\forall k, |\tilde{\mathcal{U}}_{[k]}(t) - \mathcal{U}_{[k]}(t)| \leq O(\sqrt{\frac{\log t}{t}})\},$$

$$\omega_2(t) = \{\exists k, |\tilde{\mathcal{U}}_{[k]}(t) - \mathcal{U}_{[k]}(t)| > O(\sqrt{\frac{\log t}{t}})\}.$$

$\omega_1(t)$ is the event that estimation for $\mathcal{U}_{[k]}(t)$ is bounded by $O(\sqrt{\frac{\log t}{t}})$ from its true value for all k; while $\omega_2(t)$ is its complement set, i.e., we consider two cases regrading the estimation accuracy of the upper bound of prediction performance. Clearly $\omega_1(t) \cap \omega_2(t) = \emptyset$ and $\omega_1(t) \cup \omega_2(t) = \Omega$. Then we have the following

$$r_1(f_{\tilde{k}^*(t)}(t)) = r_1(f_{\tilde{k}^*(t)}(t)|\omega_1(t))P(\omega_1(t)) + r_1(f_{\tilde{k}^*(t)}(t)|\omega_2(t))P(\omega_2(t)) \ .$$

We first bound $P(\omega_2(t))$. First via union bound we have

$$\begin{aligned} &P(\{\exists k, |\tilde{\mathcal{U}}_{[k]}(t) - \mathcal{U}_{[k]}(t)| > O(\sqrt{\frac{\log t}{t}})\}) \\ &\leq \sum_{k=1}^{K} P(|\tilde{\mathcal{U}}_{[k]}(t) - \mathcal{U}_{[k]}(t)| > O(\sqrt{\frac{\log t}{t}})) \ . \end{aligned}$$

Denote the mean of the top k similarities as $\bar{s}_{[k]} := \frac{\sum_{i=1}^{k} s_i}{k}$ and $\widetilde{\bar{s}_{[k]}}$ as its estimated version. Consider each term in the summation we have by Chernoff-Hoeffding inequality,

$$\begin{aligned} &P(|\tilde{\mathcal{U}}_{[k]}(t) - \mathcal{U}_{[k]}(t)| > O(\sqrt{\frac{\log t}{t}})) \\ &= P(|\widetilde{\bar{s}_{[k]}} - \bar{s}_{[k]}| > O(\sqrt{\frac{\log t}{t}})) \\ &\leq \sum_{i \in \mathcal{D}} P(|\tilde{s}_i - s_i| > O(\sqrt{\frac{\log t}{t}})) \\ &\leq \sum_{i \in \mathcal{D}} \left(P(|\tilde{Q}^i_{x,y} - Q^i_{x,y}| \geq O(\sqrt{\frac{\log t}{t}})) + P(|\tilde{Q}^1_{x,y} - Q^1_{x,y}| \geq O(\sqrt{\frac{\log t}{t}})) \right) , \end{aligned}$$

as we have similarly argued in bounding ordering error. For each of the term above we have (again by Chernoff bound),

$$P(|\tilde{Q}^i_{x,y} - Q^i_{x,y}| > O(\sqrt{\frac{\log t}{t}})) \leq 2e^{-2O(\frac{\log t}{t}) \cdot n_{i,x}(t)} = O(\frac{1}{t^2}) \ ,$$

with appropriately chosen constants.

When $\omega_1(t)$ happens we know that the regret from choosing the incorrect maximum $\mathcal{U}(t)$ is bounded at most by $|\tilde{\mathcal{U}}_{\tilde{k}^*(t)} - \mathcal{U}_{k^*(t)}| \leq O(\sqrt{\frac{\log t}{t}})$ since when a sub-optimal set is chosen, its regret is bounded away from its true value by at most $O(\sqrt{\frac{\log t}{t}})$ and so is the optimal set. We thus proved the theorem: to summarize with probability being at least

$$1 - O(e^{-Ct})(\text{mis-ordering}) - O(\frac{1}{t^2})(\omega_2(t)) = 1 - O(\frac{1}{t^2})$$

we have

$$r_1(f_{\tilde{k}^*(t)}(t)) = r_1(f_{\tilde{k}^*(t)}(t)|\omega_1(t), \text{correct ordering}) \leq \mathcal{U}_{k^*(t)} + O(\sqrt{\frac{\log t}{t}}) \ .$$

□

Proof of Theorem 6

Most of this Section's proof is similarly to the ones in proving Theorem 5, but with limited number of sampling.

Bounding $E[\sum_{n=1}^{t} \mathbf{1}_{O(n) \neq \emptyset}]$ and exploration errors $E[R_e(t)]/t$

We start with bound the exploration errors $E[R_e(t)]$. In order to do so, we first establish the bounded number of exploration phases. Specifically we prove $E[\sum_{n=1}^{t} \mathbf{1}_{O(n) \neq \emptyset}] \leq O(t^z)$ w.h.p.. First notice at time t we have for each state i we have most $D(t)$ number of samples from exploration. Denote $\tau_{i,x}(n)$ as the length of regeneration cycle for n-th samples of each state x of user i. Then we have

$$\sum_{n=1}^{t} \mathbf{1}_{O(n) \neq \emptyset} \leq \sum_{i \in \mathcal{D}} \sum_{x \in \mathcal{X}} \sum_{n=1}^{D(t)} \tau_{i,x}(n) \ . \tag{16}$$

Consider each sum $\sum_{n=1}^{D(t)} \tau_{i,x}(n)$. Notice $\{\tau_{i,x}(n)\}_n$ forms an IID process due to the renewal properties of Markovian processes. And it is known $E[\tau_{i,x}(n)] = \frac{1}{\pi_x^i}$. Then we have for any $0 < \gamma < \frac{1}{\pi_x^i}$ we have by Chernoff-Hoeffding inequality that

$$P(\frac{\sum_{n=1}^{D(t)} \tau_{i,x}(n)}{D(t)} - \frac{1}{\pi_x^i} < -\gamma) \leq e^{-2\gamma^2 D(t)} , \tag{17}$$

which finishes the proof.

Notice with training with data $k^*(t)$, compared to using only user 1's own data, the benefits come from a faster converging term $\sqrt{\frac{\log t}{t}}$, then

$$|\mathcal{U}_{[1]}(t) - \mathcal{U}_{k^*(t)}(t)| \leq O(\sqrt{\frac{\log t}{t}}) . \tag{18}$$

We then have

$$E[R_e(t)] \leq \sum_{n=1}^{t} \mathbf{I}_{O(n)\neq\emptyset} O(\sqrt{\frac{\log n}{n}}) . \tag{19}$$

Since the number of explorations have been bounded at the order of $O(t^z)$ and combine this with the fact that $\sqrt{\frac{\log t}{t}}$ is a decay function in t in general we have

$$\begin{aligned} E[R_e(t)] &\leq \sum_{n=1}^{O(t^z)} O(\sqrt{\frac{\log n}{n}}) \\ &\leq O(\sqrt{\log t^z} \cdot \sum_{n=1}^{O(t^z)} O(\sqrt{\frac{1}{n}}) \\ &\leq O(\sqrt{z \log t}) \cdot O((t^z)^{1-1/2}) \\ &= O(\sqrt{z \log t} \cdot t^{z/2}) , \end{aligned}$$

which gives us

$$\frac{E[R_e(t)]}{t} \leq O(\sqrt{z \log t} \cdot t^{z/2-1}) . \tag{20}$$

Bounding exploitation error

Now consider the exploitation errors. Again we first separate our discussions according two events.

$$\begin{aligned} \omega_1(t) &= \{\forall k, |\tilde{\mathcal{U}}_{[k]}(t) - \mathcal{U}_{[k]}(t)| \leq O(\sqrt{\frac{\log t}{t^z}})\}, \\ \omega_2(t) &= \{\exists k, |\tilde{\mathcal{U}}_{[k]}(t) - \mathcal{U}_{[k]}(t)| > O(\sqrt{\frac{\log t}{t^z}})\} , \end{aligned}$$

and

$$r_1(f_{k_1(t)}(t)) = r_1(f_{k_1(t)}(t)|\omega_1(t))P(\omega_1(t)) + r_1(f_{k_1(t)}(t)|\omega_2(t))P(\omega_2(t)) .$$

Similarly we could prove that with $O(\log t \cdot t^z)$ number of samples

$$\begin{aligned} P(\omega_2(t)) &\leq \sum_{1\leq k\leq K} P(|\tilde{\mathcal{U}}_{[k]}(t) - \mathcal{U}_{[k]}(t)| > O(\frac{1}{t^{z/2}})) \\ &\leq \sum_{i\in[k]} \sum_{x\in\mathcal{X}, y\in\mathcal{Y}s} P(|\tilde{Q}_{x,y}^i - Q_{x,y}^i| > O(\sqrt{\frac{\log t}{t^z}})) \\ &\leq 2e^{-2O(\frac{\log t}{t^z})\cdot D(t)} = O(\frac{1}{t^2}) , \end{aligned}$$

with appropriately selected constants.

Now we can focus on $r_1(f_{k_1(t)}(t)|\omega_1(t))$. When $\omega_1(t)$ happens we know that the regret from choosing the incorrect set of data sources is bounded at most by $|\tilde{\mathcal{U}}_{k(t)} - \tilde{\mathcal{U}}_{k^*(t)}| \leq O(\sqrt{\frac{\log t}{t^z}})$ since when a sub-optimal set is chosen, its regret is bounded away from its true value by at most $O(\sqrt{\frac{\log t}{t^z}})$ and so is the optimal set, i.e.,

$$|\mathcal{U}_{k_1(t)}(t) - \mathcal{U}_{k^*(t)}(t)| \leq O(\sqrt{\frac{\log t}{t^z}}) \, .$$

This observations leads to:

$$r_1(f_{k_1(t)}(t)|\omega_1(t)) \leq \mathcal{U}_{k^*(t)}(t) + O(\sqrt{\frac{\log t}{t^z}}) + |E[\hat{\mathcal{U}}_{k_1(t)}(t) - \mathcal{U}_{k_1(t)}(t)|\omega_1(t)]| \, , \tag{21}$$

where

$$E[\hat{\mathcal{U}}_{k_1(t)}(t)|\omega_1(t)] \leq E[\bar{\mathcal{U}}_{k_1(t)}(t)|\omega_1(t)] + E[e(t)||\omega_1(t)] \, , \tag{22}$$

and

$$\begin{aligned}\bar{\mathcal{U}}_{k_1(t)}(t) = 4 \min_{f \in \mathcal{F}} r_1^{\text{IID}}(f) + 6\beta_2 + 6\beta_1 \frac{\sum_{i \in k_1(t)} n_i(t)(1-s_i)}{N_{k_1(t)}(t)} \\ + \tilde{\rho}_{k_1(t)}(t) + 8y^*(2\sqrt{2d} + y^*) \cdot \sqrt{\frac{\log N_{k_1(t)}(t)}{N_{k_1(t)}(t)}} \, .\end{aligned} \tag{23}$$

Notice the subtle difference between $\hat{\mathcal{U}}_{k_1(t)}(t)$ and $\mathcal{U}_{k_1(t)}(t)$. $\hat{\mathcal{U}}_{k_1(t)}(t)$ is further bounded by two terms: one is $\bar{\mathcal{U}}_{k_1(t)}(t)$, the error bound with sub-sampled data and the other term $e(t)$ corresponds to the effects of dis-continuous samplings.

To make it more clear, we start the discussion by noticing that an incorrect calculation of $k_1(t)$ not only has effects on the prediction at time t, but also affects the learning process in all following steps due to this potential miss of collecting data. For details please refer to the difference between $\bar{\mathcal{U}}_{k(t)}(T)$ and $\mathcal{U}_{k(t)}(t)$: when a wrong decision is made at time t and data from the optimal sources have not been collected, the performance of the prediction for all following stages will suffer from sub-sampling. Denote $\bar{n}_i(t)$ as the number of missed data up to time t for $i \in k_1(t)$ and we first address two questions with sub-sampling for sequentially arriving Markovian data : 1). Under $\omega_1(t)$, is $\bar{n}_i(t)$ bounded above and how? (which affects $\bar{\mathcal{U}}_{k_1(t)}(t)$) 2). due to the dis-continuous sampling, there will be extra bias incurred for the distribution of sampled Markovian data. How to quantify this bias? (which affects $e(t)$).

In the rest of the proof, we show the following.

$$E[\bar{\mathcal{U}}_{k_1(t)}(t) - \mathcal{U}_{k_1(t)}(t)|\omega_1(t)] \leq O(\log t \cdot t^{-2/3}), \; E[e(t)||\omega_1(t)] \leq O(t^{-2/3}) \, . \tag{24}$$

Under $\omega_1(t)$, we further consider two events defined as follows.

$$\omega_3(t) = \{\forall i \in k_1(t) : \bar{n}_i(t) < t^{\theta})\} \, , \tag{25}$$

$$\omega_4(t) = \{\exists i \in k_1(t) : \bar{n}_i(t) \geq t^{\theta})\} \, , \tag{26}$$

for a tunable constant $0 < \theta < 1$. Again $\omega_3(t) \cap \omega_4(t) = \emptyset$ and $\omega_3(T) \cap \omega_4(t) = \Omega$. Again we have

$$E[\bar{\mathcal{U}}_{k_1(t)}(t)|\omega_1(t)] = E[\bar{\mathcal{U}}_{k_1(t)}(t)|\omega_1(t), \omega_3(t)]P(\omega_3(t)) + E[\bar{\mathcal{U}}_{k_1(t)}(t)|\omega_1(t), \omega_4(t)]P(\omega_4(t)).$$

We first prove the boundedness of $\omega_4(t)$. Since at any time t, for $i \in k^*(t)$, we know we have $i \in k^*(n), n < t$ except for a constant based on Proposition 3, this is true for all $i \in k^*(t)$ and is also true for $i \in k'(t)$ such that having estimated learning error bound $\tilde{\mathcal{U}}_{k'(t)}(t)$ within $[\tilde{\mathcal{U}}_{k^*(t)} - \sqrt{\frac{\log t}{t^z}}, \tilde{\mathcal{U}}_{k^*(t)} + \sqrt{\frac{\log t}{t^z}}]$. Since as can be similarly argued in Proposition 3, if $i \in k'(t)$ then $i \in k'(n)$ for $n \leq t$. Then due to the construction of $k_2(t)$ and under $\omega_1(t)$(with appropriately selected constant), we know this also holds for $k_1(t)$. Denote the constant as C. That suggests the fact as long as $\omega_2(n)$ is NOT true, a data sources in the optimal set $k_1(n)$ will not be missed, which suggests the following

$$\begin{aligned}E[\bar{n}_i(t)] &\leq E[\sum_{n=1}^{t} \mathbf{1}_{\omega_2(n)}] + C \\ &\leq \sum_{n=1}^{t} P(\omega_2(n) + C \\ &\leq \sum_{n=1}^{t} O(\frac{1}{n^2}) + C \\ &\leq C^{`} + C.\end{aligned}$$

Next we prove $E[\bar{n}_i^2(t)]$ is also bounded above.

$$\begin{aligned}E[\bar{n}_i^2(t)] &\leq E[(\sum_{n=1}^{t}\mathbf{1}_{\omega_2(n)}+C)^2]\\ &= E[(\sum_{n=1}^{t}\mathbf{1}_{\omega_2(n)})^2]+2CE[\sum_{n=1}^{t}\mathbf{1}_{\omega_2(n)}]+C^2\\ &\leq E[\sum_{n=1}^{t}\mathbf{1}^2_{\omega_2(n)}]+2CC^{'}+C^2\ .\end{aligned}$$

Now consider the square term. First notice

$$E[\mathbf{1}_{\omega_2(t_1)}\mathbf{1}_{\omega_2(t_2)}]\leq E[\mathbf{1}_{\omega_2(\max\{t_1,t_2\})}]\ .$$

We then have

$$\begin{aligned}E[(\sum_{n=1}^{t}\mathbf{1}_{\omega_2(n)})^2] &\leq 2E[\sum_{n=1}^{t}n\mathbf{1}_{\omega_2(n)}]\\ &\leq 2\sum_{n=1}^{t}nO(\frac{1}{n^2})\\ &\leq O(\log t)\ .\end{aligned}$$

Therefore we know $\text{var}(\bar{n}_i(t))\leq O(\log t)$. Then via Chernoff bound we know

$$P(n_i(t)\leq t-t^{\theta})=P(\bar{n}_i(t)\geq t^{\theta})\leq O(\frac{\log t}{t^{2\theta}})\ .$$

Now we analyze $|E[\hat{\mathcal{U}}_{k_1(t)}(t)-\mathcal{U}_{k_1(t)}(t)|\omega_1(t),\omega_3(t)]|$. As argued earlier we have

$$|E[\hat{\mathcal{U}}_{k_1(t)}(t)-\mathcal{U}_{k_1(t)}(t)|\omega_1(t),\omega_3(t)]|\leq|E[\bar{\mathcal{U}}_{k_1(t)}(t)-\mathcal{U}_{k_1(t)}(t)|\omega_1(t),\omega_3(t)]|+E[e(t)|\omega_1(t),\omega_3(t)]\ . \tag{27}$$

Notice

$$\begin{aligned}|E[\bar{\mathcal{U}}_{k_1(t)}(t)-\mathcal{U}_{k_1(t)}(t)|\omega_1(t),\omega_3(t)]| \leq E\Bigg[&\underbrace{6\beta_1\sum_{i\in k^*(t)}(1-s_i)|\frac{n_i(t)}{N_{k^*(t)}(t)}-\frac{1}{k}|}_{D_1(t)}+\underbrace{|\tilde{\rho}_{k^*(t)}(t)-\rho_{k^*(t)}(t)|}_{D_2(t)}\\ &+\underbrace{|8y^*(2\sqrt{2d}+y^*)\cdot|\sqrt{\frac{\log|k^*(t)|t}{|k^*(t)|t}}-\sqrt{\frac{\log N_{k^*(t)}(t)}{N_{k^*(t)}(t)}}|\Bigg|}_{D_3(t)}\omega_1(t),\omega_3(t)\Bigg]\ ,\end{aligned}$$

We have the following lemma.

Lemma 10. *We have*

- $E[D_1(t)|\omega_1(t),\omega_3(t)]\leq O(\frac{1}{t^{1-\theta}})$.
- $E[D_2(t)|\omega_1(t),\omega_3(t)]$ *decays exponentially fast.*
- $E[D_3(t)|\omega_1(t),\omega_3(t)]\leq O(\frac{\sqrt{\log t}}{t^{3/2-\theta}})$.
- $E[e(t)|\omega_1(t),\omega_3(t)]\leq O(\frac{1}{t^{1-\theta}})$.

Adding up the terms we have

$$|E[\hat{\mathcal{U}}_{k_1(t)}(t)-\mathcal{U}_{k_1(t)}(t)|\omega_1(t),\omega_3(t)]|\leq O(\frac{\log t}{t^{2\theta}})+O(\frac{1}{t^{1-\theta}})+O(\frac{\sqrt{\log t}}{t^{3/2-\theta}})\ ,$$

and the optimal upper bound occurs when $2\theta=1-\theta$ which gives us $\theta^*=1/3$ and the optimal bound follows as

$$|E[\hat{\mathcal{U}}_{k_1(t)}(t)-\mathcal{U}_{k_1(t)}(t)|\omega_1(t),\omega_3(t)]|\leq O(\frac{\log t}{t^{2/3}})\ .$$

□

Proof for Lemma 10

Bound on $E[D_1(t)|\omega_1(t),\omega_3(t)]$

Shorthand $|k^*(t)|$ as k. Take the difference between the coefficients for each term $1-s_1$ satisfies the following,

$$\frac{t}{N_{k^*(t)}(t)} - \frac{1}{k} = \frac{kt - N_{k^*(t)}(t)}{k \cdot N_{k^*(t)}(t)} \geq 0 \,.$$

For all others $i \neq 1$, we have

$$\frac{n_i(t)}{N_{k^*(t)}(t)} - \frac{1}{k} = \frac{kn_i(t) - N_{k^*(t)}(t)}{k \cdot N_{k^*(t)}(t)} \,.$$

Therefore we have

$$\begin{aligned} |\frac{\sum_i n_i(t)\cdot(1-s_i)}{N_{k^*(t)}(t)} - \frac{\sum_i (1-s_i)}{k}| &\leq \frac{kT - N_{k^*(t)}(t)}{k \cdot N_{k^*(t)}(t)} \cdot (1-s_1) \\ &+ \sum_{i\geq 2} |\frac{kn_i(t) - N_{k^*(t)}(t)}{k \cdot N_{k^*(t)}(t)}| \cdot (1-s_i) \,. \end{aligned} \tag{28}$$

Notice under $\omega_3(t)$, $N_{k^*(t)}(t) \geq kt - k\cdot t^\theta$ and we know

$$\begin{aligned} \frac{kt - N_{k^*(t)}(t)}{k \cdot N_{k^*(t)}(t)} \cdot (1-s_1) &\leq \frac{kt^z}{k(kt - k\cdot t^\theta)} \cdot (1-s_1) \\ &= \frac{t^\theta}{k(t-t^\theta)} \cdot (1-s_1) \,, \end{aligned} \tag{29}$$

and moreover

$$\begin{aligned} &|\frac{kn_i(t) - N_{k^*(t)}(t)}{k \cdot N_{k^*(t)}(t)}| \cdot (1-s_i) \\ &\leq \sum_j \frac{|n_i(t) - n_j(t)|}{k(kt - k\cdot t^\theta)} \cdot (1-s_i) \\ &\leq k \cdot \frac{t^\theta}{k(kt - k\cdot t^\theta)} \cdot (1-s_i) \\ &= \frac{t^\theta}{kt - k\cdot t^\theta} \cdot (1-s_i) = O(\frac{1}{t^{1-\theta}}) \end{aligned} \tag{30}$$

here we have used the fact that if $t - t^\theta \leq n_i(t) \leq t$ and $t - t^\theta \leq n_j(t) \leq T$ we must also have $|n_i(t) - n_j(t)| \leq t^\theta$. So is the expectation bounded. Proved . □

Bound on $E[D_2(t)|\omega_1(t),\omega_3(t)]$

This one is fairly simple to prove. Consider each of the difference term.

$$|(\lambda_2^i)^t - (\lambda_2^i)^{n_i(t)}| \leq |(\lambda_2^i)^{n_i(t)}| \,,$$

since $n_i(t) \leq t$. However

$$|(\lambda_2^i)^{n_i(t)}| \leq (\lambda_2^i)^{t-t^\theta} \,,$$

as $n_i(t) \geq t - t^\theta$. Proved. □

Bound on $E[D_3(t)|\omega_1(t),\omega_3(t)]$

Denote function $g(x) := \sqrt{\frac{\log x}{x}}$ and by mean-value theorem we have

$$|g(x-\delta) - g(x)| \leq |\delta| \cdot \max_{y\in[x-\delta,x]} \frac{\partial g(y)}{\partial y} \,.$$

Notice

$$\frac{\partial g(x)}{\partial x} = \frac{1}{2} \cdot \frac{1}{\sqrt{\frac{\log x}{x}}} \cdot |\frac{1-\log x}{x^2}| \ .$$

For $x \geq 3$ we have

$$\frac{\partial g(x)}{\partial x} \approx \frac{1}{2e} \cdot \frac{1}{\sqrt{\frac{\log x}{x}}} \cdot \frac{\log x}{x^2} = \frac{1}{2e} \cdot \sqrt{\frac{\log x}{x^3}} \ ,$$

Since $\sqrt{\frac{\log x}{x^3}}$ is a strictly decreasing function when $x \geq 3$, we have (under $\omega_3(t)$) the worst case bound is given by

$$\begin{aligned} &|\sqrt{\frac{\log(kt-(k-1)\cdot t^\theta)}{kt-(k-1)\cdot t^\theta}} - \sqrt{\frac{\log kt}{kt}}| \\ &\leq (k-1)\cdot t^\theta \cdot \frac{1}{2e} \cdot \sqrt{\frac{\log(kt-(k-1)\cdot t^\theta)}{(kt-(k-1)\cdot t^\theta)^3}} \ , \end{aligned}$$

which is decreasing sub-linearly as long as $\theta < 1$. Moreover when t is large enough we have

$$\sqrt{\frac{\log(kt-(k-1)\cdot t^\theta)}{(kt-(k-1)\cdot t^\theta)^3}} \leq \sqrt{\frac{\log t}{t^3}} \ ,$$

which leads us to the fact that

$$|\sqrt{\frac{\log(kt-(k-1)\cdot t^\theta)}{kt-(k-1)\cdot t^\theta}} - \sqrt{\frac{\log kt}{kt}}| \leq O(\frac{\sqrt{\log t}}{t^{3/2-\theta}}) \ .$$

So is the expectation bounded. Proved. □

Bound on $E[e(t)|\omega_1(t), \omega_3(t)]$

Due to discontinous sampling of Markovian data (the fact that $n_{i,x}(t) > 0$) the resultant data collection, in the form of its empirical distributions $\tilde{\pi}_x^i$ will be biased. To see this more clearly. Suppose $\tilde{f}$ is trained on a dataset $\tilde{D}$ and f is on D, where $\tilde{D}$ is a biased version of D. Then we have $E[e(t)|\omega_1(t), \omega_3(t)]$ bounded as follows (we omit the conditioning on $\omega_1(t), \omega_3(t)$ for brevity)

$$\begin{aligned} E_{z\sim D}[\mathcal{L}(\tilde{f},z)] - E_{z\sim D}[\mathcal{L}(f,z) &= E_{z\sim\tilde{D}}[\mathcal{L}(\tilde{f},z)] - E_{z\sim\tilde{D}}[\mathcal{L}(f,z)] \\ &+ (E_{z\sim D}[\mathcal{L}(\tilde{f},z)] - E_{z\sim\tilde{D}}[\mathcal{L}(\tilde{f},z)]) + (E_{z\sim D}[\mathcal{L}(f,z)] - E_{z\sim\tilde{D}}[\mathcal{L}(f,z)]) \end{aligned}$$

By definition of $\tilde{f}$ (also optimality) we know $E_{z\sim\tilde{D}}[\mathcal{L}(\tilde{f},z)] - E_{z\sim\tilde{D}}[\mathcal{L}(f,z)] \leq 0$. For $E_{z\sim D}[\mathcal{L}(\tilde{f},z)] - E_{z\sim\tilde{D}}[\mathcal{L}(\tilde{f},z)]$ we have

$$|E_{z\sim D}[\mathcal{L}(\tilde{f},z)] - E_{z\sim\tilde{D}}[\mathcal{L}(\tilde{f},z)]| \leq \max \mathcal{L} \cdot \sum_{x\in\mathcal{X}} E[|\tilde{\pi}_x^i - \pi_x^i|] \ .$$

Notice under $\omega_3(t)$,

$$E[|\tilde{\pi}_x^i - \pi_x^i|] \leq O(\frac{t^\theta}{t-t^\theta}) = O(\frac{1}{t^{1-\theta}}) \ ,$$

where the upper bound comes from the extreme cases all missed samples are from one specific state transition. Therefore we proved

$$|E_{z\sim D}[\mathcal{L}(\tilde{f},z)] - E_{z\sim\tilde{D}}[\mathcal{L}(\tilde{f},z)]| \leq O(\frac{1}{t^{1-\theta}}) \ ,$$

and similar analysis applies to $(E_{z\sim D}[\mathcal{L}(f,z)] - E_{z\sim\tilde{D}}[\mathcal{L}(f,z)])$. Proved. □

Proof of Theorem 7

We now analyze the difference in cost for requesting additional data. There are mainly two sources for this extra cost : 1). first of all, we know the there is unnecessary cost for exploration phases. This is the cost mainly for requesting enough samples to train or learn the similarity information between user 1 and any other users. 2). second is the unnecessary cost at exploitation phases when bad decisions are made (requesting data from a source that is outside the optimal set). More rigorously we have

$$\begin{aligned} E[R_c(t)] &= cE[\sum_{n=1}^{t}\sum_{i=1}^{K}\mathbf{1}_{i\notin k^*(n),i\in k_2(n)}] \\ &= cE[\sum_{n=1}^{t}\sum_{i=1}^{K}\mathbf{1}_{i\notin k^*(n),i\in k_2(n)}\cdot\mathbf{1}_{O(n)\neq\emptyset}] \\ &\quad + cE[\sum_{n=1}^{t}\sum_{i=1}^{K}\mathbf{1}_{i\notin k^*(n),i\in k_2(n)}\cdot\mathbf{1}^c_{O(n)\neq\emptyset}] \end{aligned}$$

Consider the first term above we have

$$E[\sum_{n=1}^{t}\sum_{i=1}^{K}\mathbf{1}_{i\notin k^*(n),i\in k_2(n)}\cdot\mathbf{1}_{O(n)\neq\emptyset}] \leq KE[\sum_{n=1}^{t}\sum_{i=1}^{K}\mathbf{1}^c_{O(n)\neq\emptyset}] \leq O(ct^z)\ .$$

The last inequality is due to the fact number of exploration rounds are bounded above by $O(t^z)$.

Now consider the second term. For exploitation phases, consider the possibility of requesting redundant samples. We again decompose our discussion into the cases corresponding to events $\omega_1(t)$ and $\omega_2(t)$ (as defined in the proof for Theorem 6). Then

$$\begin{aligned} &E[\sum_{n=1}^{t}\sum_{i=1}^{K}\mathbf{1}_{i\notin k^*(n),i\in k_2(n)}\cdot\mathbf{1}^c_{O(n)\neq\emptyset}] \\ &= \sum_{n=1}^{t}E[\sum_{i=1}^{K}\mathbf{1}_{i\notin k^*(n),i\in k_2(n)}\cdot\mathbf{1}^c_{O(n)\neq\emptyset}|\omega_1(n)]P(\omega_1(n))] \\ &+ \quad \sum_{n=1}^{t}E[\sum_{i=1}^{K}\mathbf{1}_{i\notin k^*(n),i\in k_2(n)}\cdot\mathbf{1}^c_{O(n)\neq\emptyset}|\omega_2(n)]P(\omega_2(n))]\ . \end{aligned}$$

As we showed probability for cases $\omega_2(t)$ is bounded above and the cost regret associated with the case is also bounded above:

$$\begin{aligned} &\sum_{n=1}^{t}E[\sum_{i=1}^{K}\mathbf{1}_{i\notin k^*(n),i\in k_2(n)}\cdot\mathbf{1}^c_{O(n)\neq\emptyset}|\omega_2(n)]P(\omega_2(n))] \\ &\leq \sum_{n=1}^{t}K\cdot O(\frac{1}{n^2}) = O(1)\ . \end{aligned}$$

Now consider the case with $\omega_1(t)$. Clearly at time t, if $k_2(t)\subseteq k^*(t)$ there would be no extra cost for redundant data. Consider the case $k^*(t)\subset k_2(t)$. Based on our sampling policy, with bounded probability as long as we have

$$\mathcal{U}_{k_2(t)} > \mathcal{U}_{k^*(t)} + O(\sqrt{\frac{\log t}{t^z}})\ , \tag{31}$$

there will be no error in requesting data from users in the set $k_2(t)$ by observing the following fact

$$\begin{aligned} \tilde{\mathcal{U}}_{k_2(t)} &> \mathcal{U}_{k_2(t)} - O(\sqrt{\frac{\log t}{t^z}}) > \mathcal{U}_{k^*(t)} + O(\sqrt{\frac{\log t}{t^z}}) - O(\sqrt{\frac{\log t}{t^z}}) \\ &= \mathcal{U}_{k^*(t)} + O(\sqrt{\frac{\log t}{t^z}}) \geq \tilde{\mathcal{U}}_{k^*(t)} + \sqrt{\frac{\log t}{t^z}}\ , \end{aligned}$$

with appropriately chosen constants. So is

$$\tilde{\mathcal{U}}^{tr}_{k_2(t)} > \tilde{\mathcal{U}}^{tr}_{k^*(t)} + \sqrt{\frac{\log t}{t^z}}\ .$$

Since $\frac{\sum_{i=1}^{k_2(t)}s_i}{|k_2(t)|} < \frac{\sum_{i=1}^{k^*(t)}s_i}{|k^*(t)|}$ as $k^*(t)\subset k_2(t)$, there exists a constant T_0 such that the Eqn.(31) holds (the gain in Term 4 becomes less than the loss in similarity Term 2.). Therefore the cost regret with over-sampling is bounded up by cKT_0 . Again summing up all above we have the bounds for $E[R_c(T)]$. Proved. □