ONLINE LEARNING WITH PROBABILISTIC FEEDBACK

Pouya M. Ghari *Yanning Shen**

University of California, Irvine
Department of Electrical Engineering and Computer Science

ABSTRACT

Online learning with expert advice is widely used in various machine learning tasks. It considers the problem where a learner chooses one from a set of experts to take advice and make a decision. In many learning problems, experts may be related, henceforth the learner can observe the losses associated with a subset of experts that are related to the chosen one. In this context, the relationship among experts can be captured by a feedback graph, which can be used to assist the learner's decision-making. However, in practice, the nominal feedback graph often entails uncertainties, which renders it impossible to reveal the actual relationship among experts. To cope with this challenge, the present work develops a novel online learning algorithm to deal with uncertainties while making use of the uncertain feedback graph. The proposed algorithm is proved to enjoy sublinear regret under mild conditions. Experiments on real datasets are presented to demonstrate the effectiveness of the novel algorithm.

Index Terms— Online Learning, Graphs, Expert Advice.

1. INTRODUCTION

Online learning with expert advice considers the case where there exists a learner and a set of experts, where the learner interacts with the experts to make a decision [1]. At each time instant, the learner chooses one of the experts and it takes the action advised by the chosen expert, then incurs the loss associated with the taken action. Such framework can be used to model different learning tasks such as online multi-kernel learning see e.g., [2, 3]. Conventional online learning literature mostly focuses on two settings, *full information* setting [4–6] or *bandit* setting [6–8]. In the full information setting, at each time instant, the learner can observe the loss associated with all experts. By contrast, in the bandit setting, the learner can only observe the loss associated with the chosen expert. However, in some applications such as the web advertising problem, where a user clicks on an ad and information about other related ads is revealed), the learner can make partial observations of losses associated with a subset of experts. In cases where querying for advice from expert incurs cost, the learner may choose to observe the loss of subset of experts, see e.g. [9, 10]. To cope with this scenario, *online learning with feedback graphs* was developed in [11], where partial observations of losses are modeled using a directed *feedback graph*. The full information and bandit settings are both special cases of online learning with either a fully connected feedback graph or a feedback graph with only self loops.

Existing works rely on the assumption the feedback graph in known *perfectly* before decision making [12–15], or after decision making [13, 16–19]. However, such information may not be available in practice. In addition, the feedback graph may contain uncertainty in practice. For example, consider an online advertisement based on a survey, where users are asked whether they are interested in certain product along with possible reasons (cost, color, etc). Note that the answer to certain product may also indicate the participant's potential interest in other products with similar cost or color. If the participant expresses interest in the product because of its affordable cost, this implies *potential* interest in other products with the same or lower price. In this case, the relationship among products can be modeled using a nominal feedback graph, where an edge exists between two nodes (products) if they share similar attributes (cost, color), implying that users *may be* interested in both products. Such a nominal feedback graph can be used to help decide which product to be advertised. However, the actual relationship among the user's interests in the products remains uncertain, which leads to uncertainty in the feedback graph. Considering the case where the exact feedback graph may not be available, [20] shows that not knowing the entire feedback graph can make the side observations useless. [21] studies the case where the exact feedback graph is unknown but is known to be generated from the Erdös-Rényi model. However, such assumption may not be valid in practice. In addition, both [20] and [21] assume that the loss associated with the chosen expert is observed.

The present paper extensively studies the case where the learner only has access to a feedback graph that may contain uncertainties, namely *nominal feedback graph*, and the learner may not be able to observe the loss associated with the chosen expert. The learner relies on this nominal feedback graph to choose among experts, and then incurs a loss associated with the chosen expert. At the same time, it observes the loss associated with a subset of experts resulting from the unknown actual feedback graph. Furthermore, different from [20] and [21], the present work does not assume it is guaranteed that the learner observes the loss associated with the chosen expert. This is true in, e.g., apple tasting problem [22]. The present work develops novel online learning algorithm to cope with uncertainties in the nominal feedback graph. Regret analysis is provided to prove that our novel algorithm can achieve sublinear regret under mild conditions. Experiments on a number of real datasets are presented to showcase the effectiveness of our novel algorithm.

2. PROBLEM STATEMENT

Consider the case where there exist K experts and the learner chooses to take the advice of one of the experts at each time instant t. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ represent the directed nominal feedback graph with a set of vertices $\mathcal{V}$, where the vertex $v_i \in \mathcal{V}$ represents the i-th expert, and there exist an edge from v_i to v_j (i.e. $(i, j) \in \mathcal{E}$), if the learner observes the loss associated with the j-th expert (i.e. $\ell_t(v_j)$) with probability p_{ij} while choosing the i-th expert. This paper considers the case where the nominal feedback graph is static such that it does not change over time. This can be the case in some practical

Work in this paper was partially supported by Microsoft Academic Grant Award for AI Research, PI: Y. Shen (yannings@uci.edu).

applications such as the online advertisement example provided in section 1. In particular, since attributes (cost, color, etc) of products do not change over time, the nominal feedback graph is static. Let $\mathcal{N}_i^{\text{in}}$ and $\mathcal{N}_i^{\text{out}}$ represent in-neighborhood and out-neighborhood of v_i in $\mathcal{G}$, respectively. Thus, $v_j \in \mathcal{N}_i^{\text{out}}$ if there is an edge from v_i to v_j (i.e. $(i,j) \in \mathcal{E}$). Similarly, $v_j \in \mathcal{N}_i^{\text{in}}$ if there is an edge from v_j to v_i (i.e. $(j,i) \in \mathcal{E}$). The nominal feedback graph $\mathcal{G}$ is revealed to the learner. The present paper considers non-stochastic adversarial online learning problems. At each time instant t, the environment privately selects a loss function $\ell_t(.)$ with $\ell_t(.) : \mathcal{V} \to [0,1]$. The learner then chooses one of the experts to take its advice. Then, the learner will incur the loss associated with the chosen expert. Let I_t denote the index of the chosen expert. Note that the learner observes $\ell_t(v_{I_t})$ with probability of $p_{I_t I_t}$, hence the loss remains unknown with the probability of $1 - p_{I_t I_t}$.

3. ONLINE LEARNING WITH PROBABILISTIC FEEDBACK

The present section studies the scenario where the nominal feedback graph $\mathcal{G}$ is revealed to the learner while the probabilities $\{p_{ij}\}$ associated with edges are not given. The present section employs geometric resampling to obtain loss estimates when there exist uncertainty. Resampling is a statistical inference technique that has been exploited in online learning [21], and particle filtering (see e.g. [23, 24]).

3.1. Geometric Resampling Approach

At time instant t, the learner assigns the weight $w_{i,t}$ to the i-th expert which indicates the confidence about the advice given by the i-th expert. Moreover, let $\mathbf{A}$ denote the adjacency matrix of the nominal feedback graph $\mathcal{G}$ with $A(i,j)$ denoting the (i,j)-th entry of A. Let X_{ij} be a Bernoulli random process with random variables $X_{ij}(t) = 1$ with probability p_{ij}. When the learner chooses the i-th expert at time t, the learner observes $\ell_t(v_j)$ only if $v_j \in \mathcal{N}_i^{\text{out}}$ and $X_{ij}(t) = 1$. If $t \leq KM$ the learner chooses the k-th expert at time instant t where $k = t - \lfloor t/K \rfloor K$. In this way, it is guaranteed that at least M samples of the mean ergodic random process X_{ij} are observed. Based on these observations, a loss estimate is obtained whose expected value is an approximation of the loss $\ell_t(v_i), \forall i \in [K]$ where $[K] := \{1, \ldots, K\}$. At $t > KM$, the learner draws one of the experts according to the PMF,

$$\pi_{i,t} = (1-\eta)\frac{w_{i,t}}{W_t} + \frac{\eta}{|\mathcal{D}|}\mathcal{I}(v_i \in \mathcal{D}), \ \ \forall i \in [K] \tag{1}$$

where $W_t = \sum_{i=1}^K w_{i,t}$, $\mathcal{D}$ represents a dominating set for $\mathcal{G}$, $\mathcal{I}(\cdot)$ denotes the indicator function and η is the learning rate. At each time instant $t > KM$, let $\tau_{ij,1}^{(t)}, \ldots, \tau_{ij,M}^{(t)}$ denote the last M time instants before t at which the learner observes samples of the random process X_{ij}. Let $Y_{ij,1}(t), \ldots, Y_{ij,M}(t)$ denote a random permutation of $X_{ij}(\tau_{ij,1}^{(t)}), \ldots, X_{ij}(\tau_{ij,M}^{(t)})$. At each time instant t, the learner draws with replacement M experts according to the PMF $\boldsymbol{\pi}_t$ in (1) in M independent trials. Let d_u denote the index of the selected expert at the u-th trial, and $P_{i,1}(t), \ldots, P_{i,M}(t)$ be a sequence of random variables associated with v_i at time instant t where $P_{d_u,u}(t) = 1$ and $P_{d'_u,u}(t) = 0$ if $d'_u \neq d_u$. Let

$$Z_{i,u}(t) = \sum_{\forall j : v_j \in \mathcal{N}_i^{\text{in}}} P_{j,u}(t) Y_{ji,u}(t) \tag{2}$$

for all $1 \leq u \leq M$. An under-estimate of loss can be obtained as

$$\tilde{\ell}_t(v_i) = Q_{i,t}\ell_t(v_i)\mathcal{I}(v_i \in \mathcal{S}_t). \tag{3}$$

where $Q_{i,t} := \min\{\{u \mid 1 \leq u \leq M, Z_{i,u}(t) = 1\}, M\}$, $\mathcal{S}_t$ represent the set of vertices associated with experts whose losses are observed by the learner at time instant t and the expected value of $\tilde{\ell}_t(v_i)$ can be written as $\mathbb{E}_t[\tilde{\ell}_t(v_i)] = \left(1 - (1-q_{i,t})^M\right)\ell_t(v_i)$, where $q_{i,t} = \sum_{\forall j : v_j \in \mathcal{N}_i^{\text{in}}} \pi_{j,t} p_{ji}$. See (12) – (15) for detailed derivation. Then, the weights $\{w_{i,t}\}_{i=1}^K$ are updated as

$$w_{i,t+1} = w_{i,t} \exp\left(-\eta\tilde{\ell}_t(v_i)\right), \ \ \forall i \in [K]. \tag{4}$$

The geometric resampling based online learning framework (Exp3-GR) is summarized in Algorithm 1.

4. REGRET ANALYSIS

In order to analyze the performance of Algorithm 1, we first preset two assumptions needed:
(a1) $0 \leq \ell_t(v_i) \leq 1, \forall t : t \in \{1, \ldots, T\}, \forall i : i \in \{1, \ldots, K\}$.
(a2) If $(i,j) \in \mathcal{E}$, the learner can observe the loss associated with the j-th expert with probability at least $\epsilon > 0$ when it chooses the i-th expert, and $(i,i) \in \mathcal{E}, \ \forall i$.

Theorem 1. *Under (a1) and (a2), the expected regret of Exp3-GR is bounded by*

$$\begin{aligned}
&\sum_{t=1}^T \mathbb{E}_t[\ell_t(v_{I_t})] - \min_{v_i \in \mathcal{V}} \sum_{t=1}^T \ell_t(v_i) \\
\leq &\frac{\ln K}{\eta} + (K-1)M + \sum_{t=KM+1}^T (1-q_{i,t})^M \\
&+ \eta(1-\eta)(T-KM) + \eta \sum_{t=KM+1}^T \sum_{i=1}^K \frac{\pi_{i,t}}{q_{i,t}}.
\end{aligned} \tag{5}$$

Proof. Since Exp3-GR chooses the experts one by one for the exploration at the first KM time instants, $\mathbb{E}_t[\ell_t(v_i)] = \ell_t(v_k)$ and under (a1), $\sum_{t=1}^{KM} \mathbb{E}_t[\ell_t(v_i)] - \sum_{t=1}^{KM} \ell_t(v_i) \leq (K-1)M$ hold true. In addition, for $t > KM$ we have

$$\frac{W_{t+1}}{W_t} = \sum_{i=1}^K \frac{w_{i,t+1}}{W_t} = \sum_{i=1}^K \frac{w_{i,t}}{W_t} \exp\left(-\eta\tilde{\ell}_t(v_i)\right). \tag{6}$$

According to (1) and the inequality $e^{-x} \leq 1 - x + \frac{1}{2}x^2, \forall x \geq 0$, the following inequality holds

$$\begin{aligned}
&\frac{W_{t+1}}{W_t} \\
\leq &\sum_{i=1}^K \frac{\pi_{i,t} - \frac{\eta}{|\mathcal{D}|}\mathcal{I}(v_i \in \mathcal{D})}{1-\eta}\left(1 - \eta\tilde{\ell}_t(v_i) + \frac{1}{2}(\eta\tilde{\ell}_t(v_i))^2\right).
\end{aligned} \tag{7}$$

Taking logarithm of both sides of (7) and using $1 + x \leq e^x$, we have

$$\begin{aligned}
&\ln \frac{W_{t+1}}{W_t} \\
\leq &\sum_{i=1}^K \frac{\pi_{i,t} - \frac{\eta}{|\mathcal{D}|}\mathcal{I}(v_i \in \mathcal{D})}{1-\eta}\left(-\eta\tilde{\ell}_t(v_i) + \frac{1}{2}(\eta\tilde{\ell}_t(v_i))^2\right).
\end{aligned} \tag{8}$$

Summing (8) over time obtains

$$\ln \frac{W_{T+1}}{W_1} \tag{9}$$

$$\leq \sum_{t=1}^{T}\sum_{i=1}^{K} \frac{\pi_{i,t} - \frac{\eta}{|\mathcal{D}|}\mathcal{I}(v_i \in \mathcal{D})}{1-\eta}\left(-\eta\tilde{\ell}_t(v_i) + \frac{1}{2}(\eta\tilde{\ell}_t(v_i))^2\right).$$

Furthermore, the left hand side of (9) can be bounded from below as

$$\ln \frac{W_{T+1}}{W_1} \geq \ln \frac{w_{i,T+1}}{W_1} = -\eta \sum_{t=1}^{T} \tilde{\ell}_t(v_i) - \ln K \tag{10}$$

where the equality holds due to the fact that $W_1 = \sum_{j=1}^{K} w_{j,1} = K$. Then, (9) and (10) lead to

$$\begin{aligned}
&\sum_{t=t'}^{T}\sum_{i=1}^{K} \pi_{i,t}\tilde{\ell}_t(v_i) - \sum_{t=t'}^{T}\tilde{\ell}_t(v_i) \\
\leq &\frac{\ln K}{\eta} + \sum_{t=t'}^{T}\sum_{i\in\mathcal{D}} \frac{\eta}{|\mathcal{D}|}\tilde{\ell}_t(v_i) \\
&+ \sum_{t=t'}^{T}\sum_{i=1}^{K} \frac{\eta}{2}(\pi_{i,t} - \frac{\eta}{|\mathcal{D}|}\mathcal{I}(v_i \in \mathcal{D}))\tilde{\ell}_t(v_i)^2
\end{aligned} \tag{11}$$

where $t' = KM + 1$. According to (3), expected value of loss estimate $\tilde{\ell}_t(v_i)$ can be expressed as

$$\mathbb{E}_t[\tilde{\ell}_t(v_i)] = \sum_{\forall j: v_j \in \mathcal{N}_{i,t}^{\text{in}}} \pi_{j,t}p_{ji}\mathbb{E}_t[Q_{i,t}]\ell_t(v_i) = q_{i,t}\mathbb{E}_t[Q_{i,t}]\ell_t(v_i) \tag{12a}$$

$$\mathbb{E}_t[\tilde{\ell}_t(v_i)^2] = \sum_{\forall j: v_j \in \mathcal{N}_{i,t}^{\text{in}}} \pi_{j,t}p_{ji}\mathbb{E}_t[Q_{i,t}^2]\ell_t(v_i)^2 = q_{i,t}\mathbb{E}_t[Q_{i,t}^2]\ell_t(v_i)^2. \tag{12b}$$

Note that the expected values depend on random variable $\{Z_{i,u}(t)\}_{u=1}^{M}$ in (2), where $P_{i,u}(t)$ and $Y_{ij,u}(t)$, $\forall i \in [K]$, $\forall (i,j) \in \mathcal{E}_t$ are independent Bernoulli random variables with parameters $\pi_{i,t}$ and p_{ij}, respectively. Therefore, $\{Z_{i,u}(t)\}_{u=1}^{M}$ are also Bernoulli random variables with expected value

$$\begin{aligned}
\mathbb{E}_t[Z_{i,u}(t)] &= \mathbb{E}_t\left[\sum_{\forall j: v_j \in \mathcal{N}_{i,t}^{\text{in}}} P_{j,u}(t)Y_{ji,u}(t)\right] \\
&= \sum_{\forall j: v_j \in \mathcal{N}_{i,t}^{\text{in}}} \mathbb{E}_t[P_{j,u}(t)]\mathbb{E}_t[Y_{ji,u}(t)] \\
&= \sum_{\forall j: v_j \in \mathcal{N}_{i,t}^{\text{in}}} \pi_{j,t}p_{ji} = q_{i,t}.
\end{aligned} \tag{13}$$

In other words, $Z_{i,u}(t)$ is a Bernoulli random variable whose value is 1 with probability $q_{i,t}$. The expected value of $Q_{i,t}$ and $Q_{i,t}^2$ can henceforth be written as

$$\begin{aligned}
\mathbb{E}_t[Q_{i,t}] &= \sum_{u=1}^{M} uq_{i,t}(1-q_{i,t})^{u-1} + M(1-q_{i,t})^M \\
&= \frac{1-(Mq_{i,t}+1)(1-q_{i,t})^M}{q_{i,t}} + M(1-q_{i,t})^M \\
&= \frac{1-(1-q_{i,t})^M}{q_{i,t}}
\end{aligned} \tag{14a}$$

$$\begin{aligned}
\mathbb{E}_t[Q_{i,t}^2] &= \sum_{u=1}^{M} u^2q_{i,t}(1-q_{i,t})^{u-1} + M^2(1-q_{i,t})^M \\
&= \frac{2-2(1-q_{i,t}^{M+2})}{q_{i,t}^2} - \frac{1+(2M+3)(1-q_{i,t})^{M+1}}{q_{i,t}} \\
&\quad - (M+1)^2(1-q_{i,t})^M + M^2(1-q_{i,t})^M \\
&= \frac{2-2(1-q_{i,t}^{M+2})}{q_{i,t}^2} - \frac{1+(2M+3)(1-q_{i,t})^{M+1}}{q_{i,t}} \\
&\quad - (2M+1)(1-q_{i,t})^M.
\end{aligned} \tag{14b}$$

Combining (12) with (14), we arrive at

$$\begin{aligned}
\mathbb{E}_t[\tilde{\ell}_t(v_i)] &= q_{i,t}\frac{1-(1-q_{i,t})^M}{q_{i,t}}\ell_t(v_i) \\
&= \left(1-(1-q_{i,t})^M\right)\ell_t(v_i) \leq \ell_t(v_i)
\end{aligned} \tag{15a}$$

$$\begin{aligned}
\mathbb{E}_t[\tilde{\ell}_t(v_i)^2] = &\left(\frac{2-2(1-q_{i,t}^{M+2})}{q_{i,t}} - 1\right)\ell_t(v_i)^2 \\
&+ (2M+3)(1-q_{i,t})^{M+1}\ell_t(v_i)^2 \\
&- q_{i,t}(2M+1)(1-q_{i,t})^M\ell_t(v_i)^2 \leq \frac{2}{q_{i,t}}.
\end{aligned} \tag{15b}$$

Combining (11) and (15), it can be concluded that

$$\begin{aligned}
&\sum_{t=t'}^{T}\sum_{i=1}^{K} \pi_{i,t}\ell_t(v_i) - \sum_{t=t'}^{T}\ell_t(v_i) - \sum_{t=t'}^{T}\sum_{i=1}^{K}\pi_{i,t}(1-q_{i,t})^M\ell_t(v_i) \\
\leq &\frac{\ln K}{\eta} + \sum_{t=t'}^{T}\sum_{i=1}^{K} \frac{\eta}{2}(\pi_{i,t} - \frac{\eta}{|\mathcal{D}|}\mathcal{I}(v_i \in \mathcal{D}))\frac{2}{q_{i,t}} \\
&+ \sum_{t=t'}^{T}\sum_{i\in\mathcal{D}} \frac{\eta}{|\mathcal{D}|}\ell_t(v_i).
\end{aligned} \tag{16}$$

According to (a1) $\ell_t(v_i) \leq 1$ and using the fact that $\frac{2}{q_{i,t}} \geq 2$, (16) can be further bounded by

$$\begin{aligned}
&\sum_{t=t'}^{T}\sum_{i=1}^{K} \pi_{i,t}\ell_t(v_i) - \sum_{t=t'}^{T}\ell_t(v_i) \\
\leq &\frac{\ln K}{\eta} + \sum_{t=t'}^{T}(1-q_{i,t})^M + \sum_{t=t'}^{T}\sum_{i\in\mathcal{D}}\frac{\eta-\eta^2}{|\mathcal{D}|} + \sum_{t=t'}^{T}\sum_{i=1}^{K}\eta\frac{\pi_{i,t}}{q_{i,t}} \\
= &\frac{\ln K}{\eta} + \sum_{t=t'}^{T}(1-q_{i,t})^M + \eta(1-\eta)(T-KM) + \eta\sum_{t=t'}^{T}\sum_{i=1}^{K}\frac{\pi_{i,t}}{q_{i,t}}.
\end{aligned} \tag{17}$$

Note that when $t > t'$, we have $\mathbb{E}_t[\ell_t(v_{I_t})] = \sum_{i=1}^{K}\pi_{i,t}\ell_t(v_i)$. Combining $\sum_{t=1}^{KM}\mathbb{E}_t[\ell_t(v_i)] - \sum_{t=1}^{KM}\ell_t(v_i) \leq (K-1)M$ with (17) leads to (5) which completes the proof. □

Building upon Theorem 1, the following Corollary presents regret bound for Exp3-GR when the dominating set $\mathcal{D}$ is found via greedy set cover algorithm (see e.g. [25]).

Corollary 1.1. *Assume that greedy set cover algorithm is employed to find a dominating set of the nominal feedback graph $\mathcal{G}$. If $M \geq \frac{|\mathcal{D}|\ln T}{2\eta\epsilon}$, under (a1) and (a2), Exp3-GR satisfies*

$$\begin{aligned}
&\sum_{t=1}^{T}\mathbb{E}_t[\ell_t(v_{I_t})] - \min_{v_i\in\mathcal{V}}\sum_{t=1}^{T}\ell_t(v_i) \\
\leq &\mathcal{O}\left(\frac{\alpha(\mathcal{G})}{\epsilon}\ln(KT)\sqrt{KT\ln K}\right)
\end{aligned} \tag{18}$$

where $\alpha(\mathcal{G})$ denotes the independence number of the graph $\mathcal{G}$.

Algorithm 1 Exp3-GR: Exp3 with geometric resampling

Input: learning rate $\eta > 0$, the minimum number of observations M, $\mathcal{G} = (\mathcal{V}, \mathcal{E})$.
Initialize: $w_{i,1} = 1, \forall i \in [K]$.
for $t = 1, \ldots, T$ **do**
  if $t \leq KM$ **then**
    Set $k = t - \lfloor \frac{t}{K} \rfloor K$ and draw the expert node v_k.
  else
    Select one expert according to PMF π_t in (1).
    Observe $\{\ell_t(v_i) : v_i \in \mathcal{S}_t\}$, compute $\tilde{\ell}_t(v_i), \forall i$ via (3).
    Update $w_{i,t+1}, \forall i \in [K]$ via (4).
  end if
end for

Proof. According to (a2), if $(i, j) \in \mathcal{E}$, the learner observes the loss of the j-th expert when it chooses the i-th expert with probability at least ϵ. Recalling (1) it can be inferred that $\pi_{i,t} > \eta/|\mathcal{D}|, \forall i \in \mathcal{D}$. Combining $q_{i,t} = \sum_{\forall j: v_j \in \mathcal{N}_i^{\text{in}}} \pi_{j,t} p_{ji}$ with the fact that for each $v_i \in \mathcal{V}$ there is at least one edge from $\mathcal{D}$ to $v_i, \forall i \in [K]$, $q_{i,t}$ can be bounded below as $q_{i,t} > \frac{\eta\epsilon}{|\mathcal{D}|}$. Combining the condition $M \geq \frac{|\mathcal{D}| \ln T}{2\eta\epsilon}$ with $q_{i,t} > \frac{\eta\epsilon}{|\mathcal{D}|}$, we have $Mq_{i,t} \geq \frac{1}{2} \ln T$ which leads to $e^{-Mq_{i,t}} \leq \frac{1}{\sqrt{T}}$. Thus, using the fact $1 + x \leq e^x$, we have

$$(1 - q_{i,t})^M \leq e^{-Mq_{i,t}} \leq \frac{1}{\sqrt{T}}. \tag{19}$$

Hence, the third term in (5), i.e., $\sum_{t=t'}^{T} (1 - q_{i,t})^M$ can be bounded by $\mathcal{O}(\sqrt{T})$. Furthermore, consider the case where $\eta = \mathcal{O}(\sqrt{\frac{K \ln K}{T}})$, and greedy set cover algorithm is used to determine the dominating set, we have $|\mathcal{D}| = \mathcal{O}(\alpha(\mathcal{G}) \ln K)$ (see e.g. [13]). Therefore, $M = \mathcal{O}(\frac{\alpha(\mathcal{G})}{\epsilon\sqrt{K}} \ln T \sqrt{T \ln K}) \geq \frac{|\mathcal{D}| \ln T}{2\eta\epsilon}$. Hence, the expected regret of Exp3-GR satisfies (18), and the Corollary 1.1 is proved. □

Comparison with [21]. Note that while Exp3-GR and Exp3-Res proposed in [21] both employ the geometric resampling technique, there exist two major differences: i) Exp3-Res assumes the actual feedback graph is generated from Erdös-Rényi model, and the probabilities of the presence of edges are equal across all edges, while Exp3-GR considers the unequally probable case; and ii) unlike Exp3-Res, Exp3-GR does not assume that the learner is guaranteed to observe the loss associated with the chosen expert.

5. EXPERIMENTS

Performance of the proposed algorithm Exp3-GR is compared with online learning algorithms Exp3 [7], Exp3-Res [21] and Exp3-DOM [13]. Exp3 considers bandit setting, and Exp3-Res assumes Erdös-Rényi model for the feedback graph. Furthermore, Exp3-DOM treats the nominal feedback $\mathcal{G}$ as the actual one without considering uncertainties. Performance is tested for regression task on several real datasets obtained from the UCI Machine Learning Repository [26]:
Air Quality: contains $9,358$ responses from sensors each with 13 features. The goal is to predict polluting chemical concentration [27].
CCPP: has $9,568$ samples, with 4 features collected from a combined cycle power plant, used to predict hourly energy output [28].
Tom's Hardware: contains $10,000$ samples from a technology forum with 96 features. The goal is to predict the average number of display about a certain topic on Tom's hardware [29].

Table 1: MSE and standard deviation ($\times 10^{-3}$) on datasets.

	Air Quality	CCPP	Tom's
Exp3	8.12 ± 0.48	20.49 ± 0.23	5.63 ± 0.42
Exp3-Res	11.68 ± 0.35	10.12 ± 0.24	6.40 ± 0.36
Exp3-DOM	5.60 ± 0.36	11.22 ± 0.21	4.33 ± 0.34
Exp3-GR	$\mathbf{4.68} \pm 0.17$	$\mathbf{8.42} \pm 0.23$	$\mathbf{3.48} \pm 0.32$

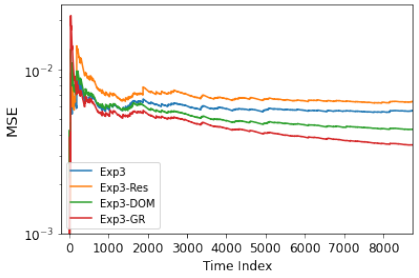

Fig. 1: MSE performance on Tom's Hardware dataset.

In all experiments, 9 experts are trained using 10% of each dataset. Among them, 8 are trained via kernel ridge regression, with 5 using RBF kernels with bandwidth of $10^{-2}, 10^{-1}, 1, 10, 100$, 3 using Laplacian kernels with bandwidth $10^{-2}, 1, 100$, and one expert is obtained via linear regression. The nominal graph $\mathcal{G}$ is fully connected. Performance of algorithms are evaluated based on mean square error (MSE) over 20 independent runs, which is defined as $\text{MSE} := \frac{1}{20} \sum_{n=1}^{20} \frac{1}{t} \sum_{\tau=1}^{t} (\hat{y}_{\tau,n} - y_\tau)^2$ where $\hat{y}_{\tau,n}$ and y_τ are the prediction of the chosen expert at n-th run and the true label of the datum at time τ, respectively. The learning rate η is set to $\frac{1}{\sqrt{t}}$ for all algorithms except for Exp3-Res which uses the suggested learning rate by [21]. Parameter M is set as 25.

We test the case where p_{ij} drawn from uniform distribution $\mathcal{U}[0.25, 0.5]$. Table 1 lists the MSE of all algorithms along with standard deviation of MSE for all datasets. It can be observed that Exp3-GR can achieve lower MSE compared with Exp3 which shows the effectiveness of using information provided by the uncertain graph. Lower MSE of Exp3-GR compared to Exp3-DOM implies that considering and uncertain graph $\mathcal{G}$ as certain is not ideal. Moreover, it can be observed Exp3-GR outperforms Exp3-Res when the actual feedback graph is not generated by Erdös-Rényi model. Figure 1 illustrates the MSE performance of algorithms on Tom's Hardware dataset over time. It can be readily observed that our proposed algorithm obtains lower MSE over time than Exp3-DOM and Exp3-Res which do not consider the uncertainty in the feedback graph.

6. CONCLUSION

The present paper studied the problem of online learning with *uncertain* feedback graphs, where potential uncertainties in the feedback graphs were modeled using probabilistic models. The proposed algorithm Exp3-GR was developed to exploit information revealed by the nominal feedback graph to help the learner with decision making. It is proved that Exp3-GR can achieve sublinear regret bound. Experiments on a number of real datasets were carried out to demonstrate that our novel algorithm can effectively address uncertainties in the feedback graph, and help enhance the learning ability of the learner.

7. REFERENCES

[1] Nicolò Cesa-Bianchi and Gabor Lugosi, *Prediction, Learning, and Games*, Cambridge University Press, USA, 2006.

[2] Yanning Shen, Tianyi Chen, and Georgios B. Giannakis, "Random feature-based online multi-kernel learning in environments with unknown dynamics," *Journal of Machine Learning Research*, vol. 20, no. 1, pp. 773–808, Jan 2019.

[3] Pouya M Ghari and Yanning Shen, "Online multi-kernel learning with graph-structured feedback," in *Proceedings of the International Conference on Machine Learning*, Jul. 2020, vol. 119, pp. 3474–3483.

[4] Nick Littlestone and Manfred K. Warmuth, "The weighted majority algorithm," *Information and Computation*, vol. 108, no. 2, pp. 212 – 261, 1994.

[5] Nicolò Cesa-Bianchi, Yoav Freund, David Haussler, David P. Helmbold, Robert E. Schapire, and Manfred K. Warmuth, "How to use expert advice," *Journal of the ACM*, vol. 44, no. 3, pp. 427–485, May 1997.

[6] Alon Resler and Yishay Mansour, "Adversarial online learning with noise," in *Proceedings of International Conference on Machine Learning*, Jun 2019, pp. 5429–5437.

[7] Peter Auer, Nicolò Cesa-Bianchi, Yoav Freund, and Robert E. Schapire, "The nonstochastic multiarmed bandit problem," *SIAM Journal on Computing*, vol. 32, no. 1, pp. 48–77, Jan 2003.

[8] Shangdong Yang and Yang Gao, "An optimal algorithm for the stochastic bandits while knowing the near-optimal mean reward," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 5, pp. 2285–2291, Jun. 2021.

[9] Kareem Amin, Satyen Kale, Gerald Tesauro, and Deepak Turaga, "Budgeted prediction with expert advice," in *AAAI Conference on Artificial Intelligence*, Austin, Texas, USA, Feb 2015.

[10] Pouya M. Ghari and Yanning Shen, "Graph-aided online learning with expert advice," in *Asilomar Conference on Signals, Systems, and Computers*, 2020, pp. 470–474.

[11] Shie Mannor and Ohad Shamir, "From bandits to experts: On the value of side-observations," in *Proc. of International Conference on Neural Information Processing Systems*, 2011, pp. 684–692.

[12] Noga Alon, Nicolò Cesa-Bianchi, Ofer Dekel, and Tomer Koren, "Online learning with feedback graphs: Beyond bandits," in *Proceedings of Conference on Learning Theory*, Paris, France, Jul 2015, vol. 40, pp. 23–35.

[13] Noga Alon, Nicolò Cesa-Bianchi, Claudio Gentile, Shie Mannor, Yishay Mansour, and Ohad Shamir, "Nonstochastic multi-armed bandits with graph-structured feedback," *SIAM Journal on Computing*, vol. 46, no. 6, pp. 1785–1826, 2017.

[14] Fang Liu, Swapna Buccapatnam, and Ness B. Shroff, "Information directed sampling for stochastic bandits with graph feedback," in *Proceedings of AAAI Conference on Artificial Intelligence*, Feb 2018.

[15] Corinna Cortes, Giulia Desalvo, Claudio Gentile, Mehryar Mohri, and Scott Yang, "Online learning with sleeping experts and feedback graphs," in *Proceedings of International Conference on Machine Learning*, Jun 2019, pp. 1370–1378.

[16] Tomáš Kocák, Gergely Neu, Michal Valko, and Rémi Munos, "Efficient learning by implicit exploration in bandit problems with side observations," in *Proceedings of International Conference on Neural Information Processing Systems*, Dec 2014, p. 613–621.

[17] Tomáš Kocák, Gergely Neu, and Michal Valko, "Online learning with noisy side observations," in *Proceedings of International Conference on Artificial Intelligence and Statistics*, Cadiz, Spain, May 2016, pp. 1186–1194.

[18] Anshuka Rangi and Massimo Franceschetti, "Online learning with feedback graphs and switching costs," in *Proceedings of International Conference on Artificial Intelligence and Statistics*, Apr 2019, pp. 2435–2444.

[19] Corrina Cortes, Giulia DeSalvo, Claudio Gentile, Mehryar Mohri, and Ningshan Zhang, "Online learning with dependent stochastic feedback graphs," in *Proceedings of International Conference on Machine Learning*, Jul 2020.

[20] Alon Cohen, Tamir Hazan, and Tomer Koren, "Online learning with feedback graphs without the graphs," in *Proceedings of International Conference on Machine Learning*, Jun 2016, p. 811–819.

[21] Tomáš Kocák, Gergely Neu, and Michal Valko, "Online learning with Erdös-Rényi side-observation graphs," in *Proceedings of Conference on Uncertainty in Artificial Intelligence*, Jun 2016, p. 339–346.

[22] David P. Helmbold, Nicholas Littlestone, and Philip M. Long, "Apple tasting," *Information and Computation*, vol. 161, no. 2, pp. 85–139, Sep 2000.

[23] Tiancheng Li, Tariq Pervez Sattar, and Shudong Sun, "Deterministic resampling: Unbiased sampling to avoid sample impoverishment in particle filters," *Signal Processing*, vol. 92, no. 7, pp. 1637–1645, 2012.

[24] Luca Martino and Víctor Elvira, "Compressed monte carlo with application in particle filtering," *Information Sciences*, vol. 553, pp. 331–352, 2021.

[25] Vasek Chvatal, "A greedy heuristic for the set-covering problem," *Mathematics of Operations Research*, vol. 4, no. 3, pp. 233–235, Aug 1979.

[26] Dheeru Dua and Casey Graff, "UCI machine learning repository," 2017.

[27] Saverio De Vito, Ettore Massera, Marco Piga, Luca Martinotto, and Girolamo Di Francia, "On field calibration of an electronic nose for benzene estimation in an urban pollution monitoring scenario," *Sensors and Actuators B: Chemical*, vol. 129, no. 2, pp. 750 – 757, 2008.

[28] Pınar Tüfekci, "Prediction of full load electrical power output of a base load operated combined cycle power plant using machine learning methods," *International Journal of Electrical Power and Energy Systems*, vol. 60, pp. 126 – 140, 2014.

[29] François Kawala, Ahlame Douzal-Chouakria, Eric Gaussier, and Eustache Dimert, "Prédictions d'activité dans les réseaux sociaux en ligne," in *4ième conférence sur les modèles et l'analyse des réseaux : Approches mathématiques et informatiques*, France, Oct. 2013, p. 16.