Online Learning Using Only Peer Prediction

Yang Liu
Computer Science and Engineering
University of California, Santa Cruz
yangliu@ucsc.edu

David P. Helmbold
Computer Science and Engineering
University of California, Santa Cruz
dph@soe.ucsc.edu

Abstract

This paper considers a variant of the classical online learning problem with expert predictions. Our model's differences and challenges are due to lacking any direct feedback on the loss each expert incurs at each time step t. We propose an approach that uses peer prediction and identify conditions where it succeeds. Our techniques revolve around a carefully designed peer score function $s()$ that scores experts' predictions based on the peer consensus. We show a sufficient condition, that we call *peer calibration*, under which standard online learning algorithms using loss feedback computed by the carefully crafted $s()$ have bounded regret with respect to the unrevealed ground truth values. We then demonstrate how suitable $s()$ functions can be derived for different assumptions and models.

1 Introduction

Consider the following online expert selection problem: at discretized time steps $t = 1, 2, ..., T$, each of N experts will form a forecast about a binary event $y_t \in \{1 \text{ (happening)}, 0 \text{ (not happening)}\}$. Let's denote expert i's prediction of how likely $y_t = 1$ will happen at time t as $p_i(t) \in [0, 1]$. In the classical online learning setting, after each round t (at time t^+), y_t is observed and each expert incurs a loss $\ell_{i,t} := \ell(p_i(t), y_t)$ according to a given loss function ℓ, which can be the squared loss, a 0-1 loss, or some other loss function. The best expert is defined as the one whose predictions minimize the total losses in hindsight: $a^* = \text{argmin}_i \sum_{t=1}^{T} \ell_{i,t}$. At each round t, the algorithm selects an expert (often using randomization) and follows its prediction, denote the selected expert as $a(t)$. To lighten the notation, we denote the prediction made according to selection $a(t)$, i.e. $p_{a(t)}(t)$, as $p_a(t)$. The algorithm's performance is typically evaluated using the following definition of regret:

$$R_T := \mathbb{E}\left[\sum_{t=1}^{T} \ell_{a(t),t}\right] - \sum_{t=1}^{T} \ell_{a^*,t} \tag{1}$$

where the expectation is with respect to the algorithm's internal randomization, and the goal is to guarantee small regret R_T (e.g. sub-linear in T).

In several natural applications of online learning, neither the ground-truth y_t values nor the true losses may be immediately available. One example is hiring junior faculty candidates by committee in a large department. Which faculty candidates will develop into superstars will only become apparent later in their careers, and many offers must be issued before this information becomes available. Our setting involves taking into account the opinions of experts (committee members) based on the particulars of the applicants at the time of hiring. Similarly school admission and other selection committees are also applications of our setting. Our goal is to identify and follow the best of these experts using a peer prediction method, where we will purely rely on predictions made by peer experts to identify proxies of the true losses. This setting also finds applications in other domains:

- Crowdsourcing: follow the best labelers, or learn how to best aggregate their advice, without ever knowing the ground truth labels.
- Long-term forecasting: use predictions from experts made long before the outcomes are realized, update forecasters' weights in advance, and make better predictions. This could correspond to the experts making all of their predictions at time 0.
- Limited access to ground truth: even when there is limited access to some ground truth values, peer prediction allows more efficient use of this limited information.

We study the situation where all (or sometimes most) of the y_t's are unavailable, and thus the $\ell_{i,t}$'s cannot be directly computed. The goal is still to have small regret R_T with

Proceedings of the 23rdInternational Conference on Artificial Intelligence and Statistics (AISTATS) 2020, Palermo, Italy. PMLR: Volume 108.

respect to the (now unseen) y_t's as defined in Eqn. (1). Our model is even more extreme than typical bandit problems: we do not even get the $\ell_{i,t}$ losses for the chosen experts.

With this paucity of feedback we must relax the adversarial setting typical in online learning models through some additional assumptions. Instead of the unavailable true losses, we construct *peer-score* functions, using peer prediction, to estimate the goodness of the experts' predictions. A natural requirement is that the consensus of the expert's predictions is somehow correlated with the true outcomes. We enforce this by requiring that both the original loss function and the peer-score function be "calibrated" by compatible divergence functions[1]. Note that even though we require the consensus to be correlated with the ground truth, this correlation can be a weak one. Our work focuses on selecting the best experts instead of performing the optimal aggregation - in practice, a small committee of the best experts can often outperform the crowd consensus [Tetlock and Gardner, 2016, Goldstein et al., 2014, Liu and Liu, 2015].

Our peer-score functions do not simply take the majority prediction as a proxy label: they explicitly adjust for biases in the majority opinion. This enables us to bound the regret when standard online algorithms are run using these peer scores as proxy losses.

Contributions and Outline: Our contributions include formalizing a peer prediction framework to study online learning problems without ground truth feedback. This framework is developed in Section 2, and involves relating the peer scores to the ground-truth losses through their calibrating functions. A second contribution is formalizing conditions on the peer scoring that guarantees any good online algorithm using the peer scoring will (w.h.p.) have good regret with respect to the unseen labels (Theorem 3). In addition, we derive suitable peer scoring functions for the square-loss with a methodology that generalizes to other calibrated and bounded loss functions in Section 3. This methodology assumes that the peer reference answers are related to the true labels through a known i.i.d. noise rate. Our third contribution is relaxing this assumption, providing bounds for known asymmetric error rates and when the noise rate is unknown, but a converging estimate of it is available (also in Section 3). We then show how such a converging estimate of the i.i.d. error rate can be efficiently produced from limited access to ground truth (4.1) or even using just the expert's predictions (4.2). Finally, we examine time-varying error rates in Section 4.3 and show how a competitive-style regret bound can be derived for that case. Our results can also be viewed as an effort to achieve self-supervision in online learning. All proofs can be found in the Appendix of our full version [Liu and Helmbold, 2019].

[1]Divergence functions are like distance functions but the triangle inequality may not hold, for instance in Bregman divergences.

Related work: As a well-established research area, it is impossible to do a thorough survey on online learning in limited space, and we refer readers to [Cesa-Bianchi and Lugosi, 2006] for a textbook treatment. Learning results can be categorized based on the types of feedback the problem admits, including: full feedback [Littlestone and Warmuth, 1994, Cesa-Bianchi et al., 1997, Arora et al., 2012], bandit feedback [Auer et al., 2002], partial feedback [Mannor and Shamir, 2011], graph feedback [Alon et al., 2015], among many others. Our results complement the online learning literature via introducing a solution framework that has no feedback but uses assumptions on peer predictions. The idea of using peer predictions has appeared in the peer prediction literature [Prelec, 2004, Miller et al., 2005, Witkowski and Parkes, 2012, Radanovic and Faltings, 2013, Dasgupta and Ghosh, 2013, Shnayder et al., 2016, Kong and Schoenebeck, 2019, Radanovic et al., 2016]. Peer prediction functions have the nice property that experts' losses will be minimized if the event happens with exactly their reported/forecasted probabilities. Our work is also relevant to the literature of learning with noisy data [Angluin and Laird, 1988, Cesa-Bianchi et al., 2011, Natarajan et al., 2013, van Rooyen and Williamson, 2015, Scott, 2015, Resler and Mansour, 2019] (with [Resler and Mansour, 2019] focusing particularly on the online learning settings as we do). The ideas are also tied to establishing calibrated surrogate losses that are robust to label noise. However, knowledge of the noise rates are often assumed to be known. We provide alternatives when such priori knowledge is absent.

Some of our example applications resemble delayed feedback settings, which have been studied previously (e.g. [Mesterharm, 2005, Joulani et al., 2013, Thune et al., 2019]). Although our paper makes stronger assumptions on the experts' predictions, the resulting bounds hold even if the feedback never arrives.

2 Peer Calibration and Main Result

After stating the prediction model, we define calibrating functions $f()$ and $g()$ for the original loss function $\ell()$ and the peer-scores $s()$, respectively. We then define the compatibility of $f()$ and $g()$ needed for our main result, and state our main theorem bounding the regret when appropriate peer-scores are used.

2.1 Preliminaries

Prediction model At each t the following happens:

- Nature selects an unknown outcome distribution p_t.
- Outcome y_t for the occurrence of event t is drawn with $y_t \sim p_t$.
- Each expert i predicts a probability $p_i(t)$ that event t occurs, possibly based on the context of the current and

previous events.
- The algorithm selects, perhaps with the aid of randomization, an expert $a(t)$ and predicts with $p_{a(t)}(t)$, based only on the experts' current and past predictions.

Although one can consider the p_t and $p_i(t)$ values as generated adversarially, the purpose of the paper is to examine what reasonable assumptions on the $p_i(t)$ values lead to successful learning with peer feedback.

As mentioned earlier, our goal is to minimize

$$R_T = \mathbb{E}\left[\sum_{t=1}^{T} \ell_{a(t),t}\right] - \sum_{t=1}^{T} \ell_{a^*,t}.$$

We will also use $L_i := \sum_{t=1}^{T} \ell_{i,t}$ for the total loss of expert i with respect to the ground truth y_t. Since we appeal to martingale inequalities, the p_t's must depend only on the previous trials.

Peer prediction Instead of using y_t which remains largely unavailable, the algorithm uses a reference answer $\hat{y}_t$ to evaluate each expert i's prediction. In short, $\hat{y}_t \in \{0, 1\}$ is some aggregation of the experts' predictions: $\hat{y}_t := \mathcal{A}(\{p_i(t)\}_{i=1}^N)$, where $\mathcal{A}(\cdot)$ maps the predictions of all experts to a single estimated label. For instance, $\mathcal{A}(\cdot)$ can be taken as the majority votes of the thresholded experts' predictions, or the "most likely" y-value found by comparing $\prod_{i=1}^N p_i(t)$ with $\prod_{i=1}^N (1 - p_i(t))$.

We will call $\hat{y}_t$ a peer *reference answer*. Then a *peer-score* function $s_{i,t} := s(p_i(t), \hat{y}_t)$ is used as a proxy for the loss of expert i's prediction. We aim to study what $s()$, when combined with standard online learning algorithms, guarantees a small regret R_T (with respect to the unseen y_t). Of course, when s or $\hat{y}_t$ is not properly designed, the peer-scores may not characterize the true performance of each expert. For instance, simply checking each prediction against the majority opinion of the set of experts may not properly identify the best expert – rather it will elect the ones who predict the majority opinion more. We will see later that suitable $s()$'s are more subtle.

2.2 Loss calibration

Definition 1. *A loss function ℓ is f-calibrated if*

$$\mathbb{E}_{y\sim p}[\ell(p', y)] - \mathbb{E}_{y\sim p}[\ell(p, y)] = f(p', p),$$

where $f()$ is a (non-negative) divergence function that measures the difference between p and p'.

If the loss is f-calibrated, then the second term $\mathbb{E}_{y\sim p}[\ell(p, y)]$ is the *minimum* expected loss that can be achieved, and it corresponds to the loss of a genie who predicts with the true distribution of y. We now give an example of an f-calibrated loss.

Lemma 1. *Squared loss $\ell(p_a(t), y) = (y - p_a(t))^2$ is calibrated with $f(p_a(t), p_t)) = (p_t - p_a(t))^2$.*

Throughout this paper, we will use squared loss as the running example, but our results generalize to other bounded proper losses, thanks to the Savage representation [Gneiting and Raftery, 2007] (see Appendix of our full version [Liu and Helmbold, 2019]). If ℓ is f-calibrated, we have the following:

$$\sum_{t=1}^{T} \mathbb{E}_{y_t\sim p_t}[\ell_{i,t}] - \sum_{t=1}^{T} \mathbb{E}_{y_t\sim p_t}[\ell(p_t, y_t)] = \sum_{t=1}^{T} f(p_i(t), p_t)$$

The second term, $\sum_{t=1}^{T} \mathbb{E}_{y_t\sim p_t}[\ell(p_t, y_t)]$ corresponds to the best possible forecaster that predicts with the distributions used to draw the outcomes y_t. Let a_f^* be the best expert w.r.t. $f()$: $a_f^* = \text{argmin}_i \sum_{t=1}^{T} f(p_i(t), p_t)$.

We'd like to argue that the best expert a^* in hindsight should roughly (and with high probability) minimize $\sum_{t=1}^{T} f(p_i(t), p_t)$, due to the convergence of $\sum_{t=1}^{T} \ell_{i,t}$ and $\sum_{t=1}^{T} \ell(p_t, y_t)$. Define $\mathcal{H}_t$ as the information set of relevant history up to time t, including all earlier $y_{t'}$'s, and $p_i(t')$'s, $t' \le t$. We will use the following martingale lemma:

Lemma 2. *Let $q(1), q(2), \ldots$ be a sequence of prediction distributions where each $q(t)$ depends only on $\mathcal{H}_{t-1}$ (and is thus conditionally independent of y_t), then $\ell_t := \sum_{\tau=1}^{t} \ell(q(\tau), y_\tau) - \sum_{\tau=1}^{t} \ell(p_\tau, y_\tau) - \sum_{\tau=1}^{t} f(q(\tau), p_\tau)$ formulates a martingale.*

The above lemma, together with the convergence properties of martingales, implies that, with high probability, the expert with the minimum sum of f scores also has low loss with respect to the true labels, so $L_{a_f^*} \approx L_{a^*}$. More precisely, the Hoeffding-Azuma inequality for martingales gives the following bound for any $\mathcal{E}_{mart} > 0$:

$$\mathbb{P}\left(\left|\sum_{\tau=1}^{t} \ell(q(\tau), y_\tau) - \sum_{\tau=1}^{t} \ell(p_\tau, y_\tau) - \sum_{\tau=1}^{t} f(q(\tau), p_\tau)\right| \ge \mathcal{E}_{mart}\right) \le 2\exp\left(-\frac{\mathcal{E}_{mart}^2}{8t}\right) \quad (2)$$

Lemma 3. *With prob. at least $1 - 2N \cdot \exp\left(-\frac{\mathcal{E}_{mart}^2}{32T}\right)$, we have $L_{a_f^*} \le L_{a^*} + \mathcal{E}_{mart}$.*

Recall that p is the probability that $y = 1$, and let $\hat{p}$ be the probability that the reference feedback $\hat{y} = 1$. We define calibration for the peer-score function as follows.

Definition 2. *A peer-score function $s()$ is g-calibrated if*

$$\mathbb{E}_{\hat{y}\sim\hat{p}}[s(p', \hat{y})] - \mathbb{E}_{\hat{y}\sim\hat{p}}[s(p, \hat{y})] = g(p', p) \quad (3)$$

where $g()$ is a divergence function measuring the difference between p and p' in the context of $\hat{p}$.

Since $\hat{p}$ appears on the left-hand-side, $g()$ will in general depend on $\hat{p}$ and it could be treated as an additional argument. However, we assume that $\hat{p}$ is the same function of p over all rounds, and thus are able to suppress this dependency. This is the case if, for example, each $\hat{y}_t$ is an i.i.d. η-perturbation of y_t so $\mathbb{P}(\hat{y}_t \neq y_t) = \eta$. Later in the paper we will consider alternative ways of generating the reference labels, but the analysis will implicitly use a function $g()$ whose $\hat{p}_t$ probabilities are a fixed function of p_t.

Let a^*_g be the best expert with respect to g and the p_t values: $a^*_g = \text{argmin}_i \sum_{t=1}^T g(p_i(t), p_t)$, and let a^*_{peer} be the best expert with respect to $s()$: $a^*_{peer} = \text{argmin}_i \sum_{t=1}^T s_{i,t}$. Consider running a "no regret" online learning algorithm over the experts using $s(p_i(t), \hat{y}_t)$ for the expert's losses. The guarantee of the online learning algorithm bounds the following regret [Cesa-Bianchi and Lugosi, 2006]:

$$R_T^{peer} := \mathbb{E}\left[\sum_{t=1}^T s_{a(t),t}\right] - \sum_{t=1}^T s_{a^*_{peer},t} \tag{4}$$

Our goal is to use this bound on R_T^{peer} to obtain bounds on R_T. As before, the Hoeffding-Azuma inequality for martingales easily gives the following bound for any $r > 0$,

$$\text{where } \sigma_g \geq |s(q(\tau), \hat{y}_\tau) - s(p_\tau, \hat{y}_\tau) - g(q(\tau), p_\tau)|$$

is a scale parameter bounding the magnitude of the random variables:

$$\mathbb{P}\left(\left|\sum_{\tau=1}^t s(q(\tau), \hat{y}_\tau) - \sum_{\tau=1}^t s(p_\tau, \hat{y}_\tau) - \sum_{\tau=1}^t g(q(\tau), p_\tau)\right| \geq r\right) \leq 2\exp\left(-\frac{r^2}{2\sigma_g^2 \cdot t}\right) \tag{5}$$

It is important to realize that although the true loss ℓ is needed (counterfactually) to evaluate for the ultimate regret, and the peer-score function $s()$ is needed to run the algorithm, the corresponding calibrating functions $f()$ and $g()$ are used only for the analysis.

We now come to the key definition of the paper. This definition establishes a connection between the true losses $\ell()$ and the peer-scores $s()$ through a relationship between their calibrating functions $f()$ and $g()$. Very informally, it says that if the algorithm's predictions and the predictions of the best expert with respect to $g()$ have related g-divergences, then the algorithm's predictions and the predictions of the best expert with respect to $f()$ have somewhat similar f-divergences. This is what will allow us to move from peer-score regrets to regrets on the true losses. It may be more surprising that peer scoring functions with the needed property can be constructed for natural situations than that this connection leads to good regret bounds.

Definition 3. *We call g "ψ-compatible with f" if there exists an invertible, increasing, and convex function ψ with $\psi(0) = 0$ such that for all p_t*

$$f(p_a(t),p_t) - f(p_{a^*_f}(t), p_t) \leq \psi^{-1}\Big(g(p_a(t), p_t) - g(p_{a^*_g}(t), p_t)\Big)$$

This definition is essentially the ψ-transform in supervised learning [Bartlett et al., 2006]. Compatibility gives a very strong relationship between f and g. In particular, If f and g are ψ-compatible, then immediately:

Fact 1 ψ^{-1} is concave and increasing, and $\psi^{-1}(0) = 0$.

This peer calibration leads us to the following propositions (proven in the Appendix of [Liu and Helmbold, 2019]):

Proposition 1. *If g is ψ-compatible with f, then:*

$$\sum_{t=1}^T f(p_a(t), p_t) - \sum_{t=1}^T f(p_{a^*_f}(t), p_t) \leq T \cdot \psi^{-1}\left(\frac{\sum_{t=1}^T g(p_a(t), p_t) - \sum_{t=1}^T g(p_{a^*_g}(t), p_t)}{T}\right)$$

Proposition 2. *If g is ψ-compatible with f, then: there exist a^*_g, a^*_f such that $a^*_g = a^*_f$.*

2.3 Peer calibration is sufficient

We are now ready to sketch the proof of our main theorem: that learning from the peer-score $s()$ losses leads to low-regret with respect to the $\ell()$ losses on the unseen ground-truth y_t values. The proof proceeds by first observing that the peer-scored loss of the algorithm is at most the peer-scored loss of a^*_{peer} plus the algorithm's expected regret bound, which we write as $\mathcal{E}_{online}(T, N) \in O(\sqrt{T \ln N})$.

We use the martingale relationship between the $s()$ losses and its $g()$ calibration and the Hoeffding-Azuma inequality to show that the $s()$ losses for a^*_{peer} and the predictions used by the algorithm are closely related to their calibrating $g()$ values. We denote the tolerable gap with $\mathcal{E}_{mart}(\delta, \sigma_g, T) = \sqrt{2\sigma_g^2 \cdot T \cdot \ln \frac{2}{\delta}}$ (recall that σ_g is the scale parameter for martingale sequence), this guarantees that each is within the gap with probability $1 - \delta$.

The optimalities of a^*_{peer} and a^*_g for $s(\cdot)$ and $g(\cdot)$ respectively imply that the total sum of $g()$ values for the algorithm's predictions are within $\mathcal{E}_{online}(T, N) + 2\mathcal{E}_{mart}(\delta, \sigma_g, T)$ of the total for the optimizing a^*_g (with probability at least $1 - 2\delta$). The compatibility between $f()$ and $g()$ ensures that $a^*_f = a^*_g$, so the algorithm is also likely to incur similar $g(\cdot)$ as a^*_f. The ψ-compatibility also allows us to use ψ^{-1} to convert average per-trial closeness wrt $g()$ into closeness wrt $f()$.

Another pair of martingale inequalities show the $\ell()$ actual losses with respect to the ground-truth y_t's are closely related to the calibrated $f()$ functions for a^*_f and the $a(t)$ predictions used by the algorithm. Gaps of $\mathcal{E}_{mart}(\delta, 2, T) =$

$\sqrt{2 \cdot 2^2 \cdot T \cdot \ln \frac{2}{\delta}}$ are needed to show that each is within the gap with probability $1-\delta$. Adding these gaps to the regret bound (and subtracting another 2δ from the confidence) gives the following theorem:

Theorem 3. *If g is ψ-compatible with f, then with probability at least $1-4\delta$,*

$$R_T \leq T \cdot \psi^{-1}\left(\frac{2\mathcal{E}_{mart}(\delta, \sigma_g, T) + \mathcal{E}_{online}(T,N)}{T}\right) + 2\mathcal{E}_{mart}(\delta, 2, T) \quad (6)$$

3 Application to Square Loss

When the loss and peer-score functions are calibrated with compatible functions, a small regret with respect to the unseen y outcomes results using the peer-score for the experts' losses. In this section, we derive a suitable peer-score $s()$ and compatible calibrating $g()$ for the square loss.

We start by assuming each reference $\hat{y}_t$ is a perturbed version of y_t with a symmetric (label independent) and homogeneous (time independent) perturbation probability η: $\mathbb{P}(\hat{y}_t \neq y_t) = \eta$, with $\eta < 0.5$, i.e. $\hat{y}$ is better than random guessing. Although this homogeneous error rate assumption looks restrictive, it is weaker than the common one in the inference literature in crowdsourcing where all agents' error rates are assumed to be homogeneous. In practice, an aggregated reference answer is relatively more stable across tasks, especially when the population is large.

We initially assume η is known, but then extend the analysis to the non-symmetric case and when only an approximation to η is available. Further extensions are in the following section.

3.1 A peer prediction function and its regret

Take ℓ as the squared loss: $\ell(p_a(t), y) = (y - p_a(t))^2$. From Lemma 1, $f(p_a(t), p_t) = (p_t - p_a(t))^2$ calibrates $\ell()$, therefore:

$$\mathbb{E}_{y_t \sim p_t}\left[\sum_{t=1}^{T} \ell_{i,t}\right] - \mathbb{E}_{y_t \sim p_t}\left[\sum_{t=1}^{T} \ell(p_t, y_t)\right] = \sum_{t=1}^{T} f(p_i(t), p_t) = \sum_{t=1}^{T} (p_t - p_i(t))^2$$

Denote the true probability of $\hat{y}_t = 1$ with $\hat{p}_t$. Simple algebra shows that

$$\begin{aligned} &\hat{p}_t := \mathbb{P}(\hat{y}_t = 1) \\ &= \mathbb{P}(\hat{y}_t = 1 | y_t = 1)\mathbb{P}(y_t = 1) + \mathbb{P}(\hat{y}_t = 1 | y_t = 0)\mathbb{P}(y_t = 0) \\ &= (1-\eta) \cdot p_t + \eta \cdot (1 - p_t) = (1 - 2\eta) \cdot p_t + \eta. \end{aligned} \quad (7)$$

This observation enables us to prove the following lemma with a bit of simple algebra. First we define:

$$F(\eta, p_t) := -\eta(1-\eta)(1-2p_t)^2 + 2\eta \cdot p_t^2 - 2\eta \cdot p_t + \eta$$

which is independent of i.

Lemma 4. *For expert $i = 1, ..., N$ and time $1 \leq t \leq T$:*

$$\begin{aligned} &\mathbb{E}_{\hat{y}_t \sim \hat{p}_t}\left[\ell(p_i(t), \hat{y}_t)\right] - \mathbb{E}_{\hat{y}_t \sim \hat{p}_t}\left[\ell(p_t, \hat{y}_t)\right] \\ &= (1-2\eta) f(p_i(t), p_t) - 2\eta \cdot p_i(t)(1 - p_i(t)) - F(\eta, p_t). \end{aligned}$$

The above lemma inspires us to design the following peer-score function $s(\cdot)$ by first cancelling the $p_i(t)(1 - p_i(t))$ terms in $\ell(p_i(t), \hat{y}_t)$ and then observing that $(1 - 2\eta) f(p_i(t), p_t) - F(\eta, p_t)$ is compatible with f since $F(\eta, p_t)$ is invariant across all experts.

Theorem 4. *If the peer-score function and $g(p_i(t), p_t)$ are:*

$$s_{i,t} := \ell(p_i(t), \hat{y}_t) + 2\eta \cdot p_i(t)(1 - p_i(t)), \quad (8)$$

$$g\left(p_i(t), p_t\right) := (1-2\eta)(p_t - p_i(t))^2 - F(\eta, p_t) - 2\eta p_t(1 - p_t), \quad (9)$$

then $s()$ is $g()$-calibrated and g is $\psi^{-1}(x) = x/(1-2\eta)$-compatible with $f()$.

Therefore Theorem 3 gives the following regret bound, which holds with probability $1 - 4\delta$:

$$\begin{aligned} R_T \leq & T \cdot \psi^{-1}\left(\frac{2\mathcal{E}_{mart}(\delta, \sigma_g, T) + \mathcal{E}_{online}(T,N)}{T}\right) \\ &+ 2\mathcal{E}_{mart}(\delta, 2, T) \\ = & \frac{2\mathcal{E}_{mart}(\delta, \sigma_g, T) + \mathcal{E}_{online}(T,N)}{1-2\eta} + 2\mathcal{E}_{mart}(\delta, 2, T) \end{aligned}$$

where $\sigma_g = \max\{4 + \max F(\eta, p_t), 2 - \min F(\eta, p_t)\}$. A couple of remarks follow:

- The above bound assumes $\eta < 0.5$ and diverges as the $\hat{y}$ become uninformative ($\eta \to 1/2$).
- Theorem 4's peer-score construction can be generalized to other calibrated loss functions ℓ using the Savage representation of proper scoring rules [Gneiting and Raftery, 2007] (Appendix of [Liu and Helmbold, 2019]).

3.2 Asymmetric error rate

We now relax the assumption of symmetric label noise: for known η_0 and η_1, let $\mathbb{P}(\hat{y}_t = 1 | y_t = 0) = \eta_0$, $\mathbb{P}(\hat{y}_t = 0 | y_t = 1) = \eta_1$, with $\eta_0 + \eta_1 < 1$ (better than random guessing) Liu and Chen [2017]. A more general approach relates to learning with noisy data [Natarajan et al., 2013, Scott, 2015, Menon et al., 2015, van Rooyen and Williamson, 2015], where the goal is to design a surrogate loss function that calibrates the true losses in the presence of label biases. For instance, one such s is

$$s(p_a(t), \hat{y}_t) = (1 - \eta_{1-\hat{y}_t})\ell(p_a(t), \hat{y}_t) - \eta_{\hat{y}_t}\ell(p_a(t), 1 - \hat{y}_t) \quad (10)$$

Then we have

Lemma 5 (Natarajan et al. [2013])**.** *For each time t, $\mathbb{E}[s(p_a(t), \hat{y}_t)] = (1 - \eta_0 - \eta_1) \cdot \mathbb{E}[\ell(p_a(t), y_t)]$.*

Following above lemma immediately we will have

Proposition 5. *$s()$ defined in Eqn.(10) is g-calibrated where $g() := (1-\eta_0-\eta_1)f()$ is ψ-compatible with f for $\psi^{-1}(x) = x/(1-\eta_0-\eta_1)$.*

Therefore we establish the following regret bound from Theorem 3

$$\frac{2\mathcal{E}_{mart}(\delta,\sigma_g,T)+\mathcal{E}_{online}(T,N)}{1-\eta_0-\eta_1}+2\mathcal{E}_{mart}(\delta,2,T).$$

Estimating the two error rates η_0 and η_1 is generally a harder task than estimating a single error rate, especially when the errors may vary over time (a challenge addressed in Section 4 and 4.3).

Mapping to a class-independent error rate setting In light of above discussion, we propose an approach to map the asymmetric error rate case to a symmetric one. At each time t the trial is "flipped" with probability $1/2$. When a trial is "flipped" we use outcome $\tilde{y}_t := 1-y_t$ and flipped predictions $\tilde{p}_i(t) := 1-p_i(t)$, so $\hat{y}_t$ is also flipped. After flipping, $\hat{y}_t$ has the nice property:

Lemma 6. *$\hat{y}_t$ has class-independent error rates w.r.t. $\tilde{y}_t$.*

This result allows us to focus on the class-independent error rate setting.

3.3 Using estimated noise rates

In practice, the error rate of $\hat{y}$ is unknown a priori. Before considering the learning of error rates, we generalize Theorem 3 and show how using an estimate $\hat{\eta}_t$ for $\eta_t = \mathbb{P}(\hat{y}_t \neq y_t)$ affects the regret bounds. Suppose the peer-score becomes (adapted from Eqn. (8))

$$s_{i,t} := \ell(p_i(t),\hat{y}_t) + 2\hat{\eta}_t \cdot p_i(t)(1-p_i(t)),$$

with $\hat{\eta}_t$ replacing η, and we have a bound $|\hat{\eta}_t - \eta| \leq \epsilon_t$ then we get the following.

Theorem 6. *Suppose noisy estimates $\hat{\eta}_t$ replace the true noise rate η in Eqn. (9) where each $|\hat{\eta}_t - \eta| \leq \epsilon_t$, and the algorithm uses the resulting peer-scores. Then Theorems 3 and 4 imply, with probability at least $1-4\delta$*

$$R_T \leq \frac{2\mathcal{E}_{mart}(\delta,\sigma_g,T)+\mathcal{E}_{online}(T,N)+\sum_{t=1}^{T}\epsilon_t}{1-2\eta} + 2\mathcal{E}_{mart}(\delta,2,T)$$

where $\sigma_g = \max\{4+\max F(\eta,p_t), 2-\min F(\eta,p_t)\}$.

4 Approximating the error rates

Here we extend the analysis to when the error rates of the reference answers are *unknown* (Section 4.1 and 4.2) and *heterogeneous* across time (Section 4.3), expanding the applicability of our results.

4.1 Limited access to ground truth

Suppose the error rate $\eta = \mathbb{P}(\hat{y}_t \neq y_t)$ of the reference answer is homogeneous but is unknown a priori. We start with an easier setting where we occasionally get the y_t ground truth feedback with some known probability. We show that this limited access to ground truth can be better utilized to estimate the $\hat{y}_t$ error rate, rather than directly estimating each of the losses.

Suppose, at each time t, the ground truth label becomes available with probability p^*. We apply standard importance weighting to estimate the error rate η as follows:

$$\hat{\mathbb{1}}(\hat{y}_t, y_t) = \begin{cases} \frac{\mathbb{1}(\hat{y}_t = y_t)}{p^*}, & \text{if ground truth becomes available} \\ 0, & \text{otherwise} \end{cases}$$

then we estimate η as follows at step t: $\hat{\eta}_t := \frac{\sum_{n=1}^{t}\hat{\mathbb{1}}(\hat{y}_t,y_t)}{t}$. The expectation $\mathbb{E}[\hat{\eta}_t] = \eta$, next we show this estimation costs another $O(\frac{\sqrt{T\cdot\ln\frac{2}{\delta}}}{p^*})$ regret term in $\psi^{-1}(\frac{\cdot}{T})$ with probability at least $1-\delta$ (using Theorem 6).

The martingale nature of the y_ts imply $\hat{\mathbb{1}}(\hat{y}_t, y_t)$ also forms a martingale sequence. By the "maximal" version of Hoeffding-Azuma inequality we know

$$\mathbb{P}\left(\max_{t\leq T}\left|\sum_{n=1}^{t}\hat{\mathbb{1}}(\hat{y}_t,y_t)-\eta\cdot t\right|>\epsilon\right) \leq 2\exp\left(\frac{-2\epsilon^2}{t\cdot(\frac{1}{p^*})^2}\right)$$

Setting $\epsilon = \sqrt{\frac{t}{2(p^*)^2}\ln\frac{2}{\delta}}$, we have with probability at most δ that: $\left|\sum_{n=1}^{t}\hat{\mathbb{1}}(\hat{y}_t,y_t)-\eta\cdot t\right| > \sqrt{\frac{t}{2(p^*)^2}\ln\frac{2}{\delta}}$. Therefore

$$|\hat{\eta}_t - \eta| = \left|\frac{\sum_{n=1}^{t}\hat{\mathbb{1}}(\hat{y}_t,y_t)}{t} - \frac{\eta\cdot t}{t}\right| \leq \frac{\sqrt{\ln\frac{2}{\delta}}}{p^*\sqrt{2t}},\ \forall t \qquad (11)$$

with probability at least $1-\delta$. According to Theorem 6, this will introduce another regret term:

$$\sum_{t=1}^{T}|\hat{\eta}_t-\eta| = \sum_{t=1}^{T}\frac{\sqrt{\ln\frac{2}{\delta}}}{p^*\sqrt{2t}} = O\left(\frac{\sqrt{T\cdot\ln\frac{2}{\delta}}}{p^*}\right)$$

Estimating a single error rate allows the $\frac{1}{p^*}$ term to be independent of the number of experts, as opposed to the typical $\frac{\sqrt{T\ln(N/\delta)}}{p^*}$ regret [Cesa-Bianchi and Lugosi, 2006].

4.2 No access to ground truth

The task of estimating the error rate η is much harder when there is no ground truth information available. We propose the following method to estimate it:

- Randomly partition the experts into two groups, namely groups A, B. Denote the aggregated reference answers within each group as $\hat{y}_{A,t}$ and $\hat{y}_{B,t}$ respectively.

- Denote the error rates for $\hat{y}_{A,t}$ and $\hat{y}_{B,t}$ as η_A, η_B respectively. Assume $\eta_A, \eta_B < 0.5$, the error rates stay constant over time, and they are conditionally independent given the ground truth y_t: $\mathbb{P}(\hat{y}_{A,t}, \hat{y}_{B,t}|y_t) = \mathbb{P}(\hat{y}_{A,t}|y_t)\mathbb{P}(\hat{y}_{B,t}|y_t)$.

We leverage the comparison between the two groups. Define $c_{1,t}, c_{2,t}, c_{3,t}$ as the following (unknown) parameters estimatable without y_ts:

$$c_{1,t} = \frac{\sum_{\tau=1}^{t} \mathbb{P}(\hat{y}_{A,\tau} = 1)}{t}, \quad c_{2,t} = \frac{\sum_{\tau=1}^{t} \mathbb{P}(\hat{y}_{B,\tau} = 1)}{t},$$
$$c_{3,t} = \frac{\sum_{\tau=1}^{t} \mathbb{P}(\hat{y}_{A,\tau} = \hat{y}_{B,\tau} = 1)}{t}$$

We have the following theorem:

Theorem 7. *Rates $\eta_A, \eta_B < 1/2$ are uniquely characterized by the following three equations:*

$$P_{0,t} \cdot \eta_A + (1 - P_{0,t})(1 - \eta_A) = c_{1,t},$$
$$P_{0,t} \cdot \eta_B + (1 - P_{0,t})(1 - \eta_B) = c_{2,t}$$
$$P_{0,t} \cdot \eta_A \cdot \eta_B + (1 - P_{0,t})(1 - \eta_A)(1 - \eta_B) = c_{3,t},$$

where $P_{0,t} = \frac{\sum_{\tau=1}^{t} \mathbb{1}(y_\tau = 0)}{t}$, η_A, and η_B are the unknowns.

Parameters $c_{1,t}, c_{2,t}, c_{3,t}$ can be empirically estimated along the way, providing estimates for η_A, η_B via solving the equations. Then we can set $\hat{y}_t$ as either $\hat{y}_{A,t}$ or $\hat{y}_{B,t}$, and use the estimated $\hat{\eta}_A, \hat{\eta}_B$ correspondingly.

Denote the estimation of η_A, η_B at time t as $\hat{\eta}_{A,t}, \hat{\eta}_{B,t}$ respectively using estimates of $c_{1,t}, c_{2,t}, c_{3,t}$. A finer degree analysis also gives us:

Theorem 8. *At t, w.p. $\geq 1 - 3\delta$, $|\hat{\eta}_{A,t} - \eta_A| \leq O(\sqrt{\frac{\ln \frac{2}{\delta}}{2t}})$, $|\hat{\eta}_{B,t} - \eta_B| \leq O(\sqrt{\frac{\ln \frac{2}{\delta}}{2t}})$, when P_0 is bounded away from 0.5.*[2]

This leads to a $O(\sqrt{\frac{\ln(6/\delta)}{2t}})$ regret for η_A, η_B with probability $\geq 1 - \delta$, which incurs an additional $\sum_{t=1}^{T} O(\sqrt{\frac{\ln(6/\delta)}{2t}}) = O(\sqrt{T \cdot \ln \frac{6}{\delta}})$ regret (Theorem 6).

4.3 Heterogeneous error rates

Now we consider a setting where the error rates, now denoted $\eta_t < 0.5$, change. The challenge is the previous techniques lead to minimizing a term like (according to Lemma 4 and Theorem 4):

$$\sum_{t=1}^{T} (1 - 2\eta_t) f(p_a(t), p_t) \sim \sum_{t=1}^{T} (1 - 2\eta_t)(\ell_{a,t} - \ell(p_t, y_t))$$

[2] When P_0 is close to 0.5, the first and second equations presented in the estimation equations in Theorem 7 can uniquely determine η_A, η_B separately.

instead of the constant $1 - 2\eta$ coefficient, which enables compatible calibration. Our previous error estimation procedure estimates the average error rate instead of treating each η_t separately.

Inspired by the uniform noise case, if the η_ts can be made similar enough, then peer calibration techniques can give bounds even in the heterogeneous case. We use the following flipping based mechanism to reduce the heterogeneity: randomly flip the peer reference answer with probability $\hat{p}$:

$$\tilde{y}_t := \begin{cases} \hat{y}_t, & \text{w.p. } 1 - \hat{p} \\ 1 - \hat{y}_t, & \text{w.p. } \hat{p} \end{cases}$$

and use this newly flipped $\tilde{y}_t$ as our peer reference outcome. With this flipping, the error rate $\tilde{\eta}_t$ for reference answer $\tilde{y}_t$ becomes: $\tilde{\eta}_t = \eta_t(1 - \hat{p}) + (1 - \eta_t)\hat{p}$. This implies that for any two times t_1, t_2 we have $|\tilde{\eta}_{t_1} - \tilde{\eta}_{t_2}| := (1 - 2\hat{p})|\eta_{t_1} - \eta_{t_2}|$. Let $\tilde{\eta}$ be the average $\frac{\sum_{t=1}^{T} \tilde{\eta}_t}{T}$, implying $|\tilde{\eta}_t - \tilde{\eta}| \leq (1 - 2\hat{p}) \max_{t_1,t_2} |\eta_{t_1} - \eta_{t_2}|$. As $\hat{p} \to 0.5$, the slack in this inequality becomes arbitrarily small, and the different error rates at different t become similar (homogeneous). Thus a properly chosen $\hat{p}$ can make $|\eta_t - \tilde{\eta}|$ small enough to exploit the similarity between the $f()$ and $g()$ functions almost as if they were compatible.

With this flipping, we can estimate η_t as the average error rate up to time t using methods from Sections 4.1 and 4.2 for use in the peer-scores, denoting as $\hat{\eta}_t$. And then let

$$s_{i,t} = \ell(p_i(t), \tilde{y}_t) + \hat{\eta}_t \cdot p_i(t) \cdot (1 - p_i(t)).$$

We now focus on binary expert predictions where $p_i(t) \in \{0, 1\}$. Note all our previous results hold for the binary prediction case as $p_i(t)$s can be interpreted as with probability 0 or 1. For the competitive ratio $c_{\mathsf{comp}}(\alpha) := \alpha\big(\frac{1}{1 - 2\max_t \tilde{\eta}_t} + 1\big)$, we have:

Theorem 9. *For any $\alpha = 2 + \epsilon$ ($\epsilon > 0$), there exists a $0 < \hat{p} < 1/2$ (bounded away from 0.5) such that, with probability at least $1 - \delta - \delta_g$, the above process's regret R_T is bounded as follows:*

$$R_T \leq \frac{\mathcal{E}_{mart}(\frac{\delta}{2N}, 2, T) + \mathcal{E}_{mart}(\frac{\delta}{2N}, \sigma_g, T) + \mathcal{E}_{online}(T, N)}{1 - 2\max_t \tilde{\eta}_t} + c_{comp}(\alpha) \cdot L_{a^*}.$$

Thus we achieve a competitive ratio w.r.t. the optimal loss, up to an additional sub-linear term. Note $\max_t \tilde{\eta}_t$ is bounded away from 0.5 if both $\hat{p}$ and η_ts are.

Without the estimation of error rates? A recent work [Liu and Guo, 2019] provides instructions on constructing peer-calibrated (in a similar notion) score/loss functions (peer loss) when facing noisy supervisions for a supervised learning setting, but without the need of specifying/estimating the noise rates. Peer loss is similarly inspired by peer prediction functions. In the future we will

explore how the results in [Liu and Guo, 2019] can be reproduced in our online learning setting.

5 Concluding remarks

In this paper, we developed a framework for online learning problems where peer assessment is the only feedback. We derived appropriate peer-score functions that can be used as proxies for the experts' losses and showed they result in low-regret algorithms. With this lower level of feedback, additional assumptions are needed. To a certain degree, our solution provides a solution template for self-supervised online learning under different assumptions.

Appendix

All missing details can be found in [Liu and Helmbold, 2019]. Below we sketch the main proof of Theorem 3.

Step 1 Using Martingale inequality we know

$$\sum_{t=1}^{T} s_{a(t),t} - \sum_{t=1}^{T} s_{p_t,t} \to \sum_{t=1}^{T} g(p_a(t), p_t),$$
$$\sum_{t=1}^{T} \ell_{a(t),t} - \sum_{t=1}^{T} \ell_{p_t,t} \to \sum_{t=1}^{T} f(p_a(t), p_t)$$

The above approximation incurs at most $\mathcal{E}_{mart}(\delta, \sigma_g, T)$ (for s) and $\mathcal{E}_{mart}(\delta, 2, T)$ (for ℓ) error with probability at least $1 - 4\delta$. In particular, from Eqn. (5), with probability at least $1 - 2\delta$, the following holds:

$$\left|\sum_{t=1}^{T} s_{a(t),t} - \sum_{t=1}^{T} s_{p_t,t} - \sum_{t=1}^{T} g(p_a(t), p_t)\right| \le \mathcal{E}_{mart}(\delta, \sigma_g, T)$$
$$\left|\sum_{t=1}^{T} s_{a^*_{peer},t} - \sum_{t=1}^{T} s_{p_t,t} \sum_{t=1}^{T} g(p_{a^*_{peer}}(t), p_t)\right| \le \mathcal{E}_{mart}(\delta, \sigma_g, T)$$

Similarly with probability at least $1 - 2\delta$,

$$\left|\sum_{t=1}^{T} \ell_{a(t),t} - \sum_{t=1}^{T} \ell_{p_t,t} - \sum_{t=1}^{T} f(p_a(t), p_t)\right| \le \mathcal{E}_{mart}(\delta, 2, T)$$
$$\left|\sum_{t=1}^{T} \ell_{a^*_{peer},t} - \sum_{t=1}^{T} \ell_{p_t,t} - \sum_{t=1}^{T} f(p_{a^*_{peer}}(t), p_t)\right| \le \mathcal{E}_{mart}(\delta, 2, T)$$

Step 2 Using facts in Step 1, the following holds

$$\begin{aligned}
&\sum_{t=1}^{T} g(p_a(t), p_t) - \sum_{t=1}^{T} g(p_{a^*_g}(t), p_t) \\
&\le \sum_{t=1}^{T} g(p_a(t), p_t) - \sum_{t=1}^{T} s_{a(t),t} + \sum_{t=1}^{T} s_{p_t,t} + \sum_{t=1}^{T} s_{a^*_g,t} \\
&- \sum_{t=1}^{T} s_{p_t,t} - \sum_{t=1}^{T} g(p_{a^*_g}(t), p_t) + \sum_{t=1}^{T} s_{a(t),t} - \sum_{t=1}^{T} s_{a^*_{peer},t} \\
&\le 2\mathcal{E}_{mart}(\delta, \sigma_g, T) + \sum_{t=1}^{T} s_{a(t),t} - \sum_{t=1}^{T} s_{a^*_{peer},t}
\end{aligned}$$

The first inequality is because $\sum_{t=1}^{T} s_{a^*_{peer},t} \le \sum_{t=1}^{T} s_{a^*_g,t}$ (optimality of a^*_{peer}).

Step 3 By Proposition 1 we know

$$\begin{aligned}
&\sum_{t=1}^{T} f(p_a(t), p_t) - \sum_{t=1}^{T} f(p_{a^*_f}(t), p_t) \\
&\le T \cdot \psi^{-1}\left(\frac{\sum_{t=1}^{T} g(p_a(t), p_t) - \sum_{t=1}^{T} g(p_{a^*_g}(t), p_t)}{T}\right) \\
&\le T \cdot \psi^{-1}\left(\frac{2\mathcal{E}_{mart}(\delta, \sigma_g, T) + \sum_{t=1}^{T} s_{a(t),t} - \sum_{t=1}^{T} s_{a^*_{peer},t}}{T}\right)
\end{aligned}$$

Step 4 Then $\sum_{t=1}^{T} \ell_{a(t),t} - \sum_{t=1}^{T} \ell_{a^*,t}$ becomes

$$\begin{aligned}
&\sum_{t=1}^{T} \ell_{a(t),t} - \sum_{t=1}^{T} \ell_{a^*,t} \\
&= (\sum_{t=1}^{T} \ell_{a(t),t} - \sum_{t=1}^{T} \ell_{p_t,t}) - (\sum_{t=1}^{T} \ell_{a^*,t} - \sum_{t=1}^{T} \ell_{p_t,t}) \\
&\le \sum_{t=1}^{T} f(p_a(t), p_t) - \sum_{t=1}^{T} f(p_{a^*}(t), p_t) + 2\mathcal{E}_{mart}(\delta, 2, T) \\
&\le \sum_{t=1}^{T} f(p_a(t), p_t) - \sum_{t=1}^{T} f(p_{a^*_f}(t), p_t) + 2\mathcal{E}_{mart}(\delta, 2, T) \\
&\le T\psi^{-1}\left(\frac{2\mathcal{E}_{mart}(\delta, \sigma_g, T) + \sum_{t=1}^{T} s_{a(t),t} - \sum_{t=1}^{T} s_{a^*_{peer},t}}{T}\right) \\
&\quad + 2\mathcal{E}_{mart}(\delta, 2, T)
\end{aligned}$$

Step 5 From the guarantee of running an online learning algorithm, we have

$$\mathbb{E}\left[\sum_{t=1}^{T} s_{a(t),t}\right] - \sum_{t=1}^{T} s_{a^*_{peer},t} \le \mathcal{E}_{online}(T, N) \qquad (12)$$

Further

$$\begin{aligned}
&\mathbb{E}\left[\sum_{t=1}^{T} \ell_{a(t),t}\right] - \sum_{t=1}^{T} \ell_{a^*,t} \\
&\le T\mathbb{E}\left[\psi^{-1}\left(\frac{2\mathcal{E}_{mart}(\delta, \sigma_g, T) + \sum_{t=1}^{T} s_{a(t),t} - \sum_{t=1}^{T} s_{a^*_{peer},t}}{T}\right)\right] \\
&\quad + 2\mathcal{E}_{mart}(\delta, 2, T) \\
&\le T\psi^{-1}\left(\frac{2\mathcal{E}_{mart}(\delta, \sigma_g, T) + \mathbb{E}\left[\sum_{t=1}^{T} s_{a(t),t} - \sum_{t=1}^{T} s_{a^*_{peer},t}\right]}{T}\right) \\
&\quad + 2\mathcal{E}_{mart}(\delta, 2, T) \qquad \text{(Concavity of } \psi^{-1}(\cdot)\text{)} \\
&\le T\psi^{-1}\left(\frac{2\mathcal{E}_{mart}(\delta, \sigma_g, T) + \mathcal{E}_{online}(T, N)}{T}\right) + 2\mathcal{E}_{mart}(\delta, 2, T).
\end{aligned}$$

This completes the proof.

Acknowledgement

Yang Liu would like to thank Yiling Chen for helpful and inspiring early discussions on this problem. This work is partially funded by the Defense Advanced Research Projects Agency (DARPA) and Space and Naval Warfare Systems Center Pacific (SSC Pacific) under Contract No. N66001-19-C-4014. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of DARPA, SSC Pacific or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein.

References

Noga Alon, Nicolo Cesa-Bianchi, Ofer Dekel, and Tomer Koren. Online learning with feedback graphs: Beyond bandits. In *Annual Conference on Learning Theory*, volume 40. Microtome Publishing, 2015.

Dana Angluin and Philip Laird. Learning from noisy examples. *Machine Learning*, 2(4):343–370, 1988.

Sanjeev Arora, Elad Hazan, and Satyen Kale. The multiplicative weights update method: a meta-algorithm and applications. *Theory of Computing*, 8(1):121–164, 2012.

Peter Auer, Nicolo Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine learning*, 47(2-3):235–256, 2002.

Peter L Bartlett, Michael I Jordan, and Jon D McAuliffe. Convexity, classification, and risk bounds. *Journal of the American Statistical Association*, 101(473):138–156, 2006.

Nicólo 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*, 44(3):427–485, 1997.

Nicolo Cesa-Bianchi and Gábor Lugosi. *Prediction, learning, and games*. Cambridge university press, 2006.

Nicolo Cesa-Bianchi, Shai Shalev-Shwartz, and Ohad Shamir. Online learning of noisy data. *IEEE Transactions on Information Theory*, 57(12):7907–7931, 2011.

Anirban Dasgupta and Arpita Ghosh. Crowdsourced judgement elicitation with endogenous proficiency. In *Proceedings of the 22nd international conference on World Wide Web*, pages 319–330, 2013.

Tilmann Gneiting and Adrian E. Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007.

Daniel G Goldstein, Randolph Preston McAfee, and Siddharth Suri. The wisdom of smaller, smarter crowds. In *Proceedings of the fifteenth ACM conference on Economics and computation*, pages 471–488. ACM, 2014.

Pooria Joulani, Andras Gyorgy, and Csaba Szepesvári. Online learning under delayed feedback. In *International Conference on Machine Learning*, 2013.

Yuqing Kong and Grant Schoenebeck. An information theoretic framework for designing information elicitation mechanisms that reward truth-telling. *ACM Transactions on Economics and Computation (TEAC)*, 7(1):1–33, 2019.

Nick Littlestone and Manfred K Warmuth. The weighted majority algorithm. *Information and computation*, 108 (2):212–261, 1994.

Yang Liu and Yiling Chen. Machine-learning aided peer prediction. In *Proceedings of the 2017 ACM Conference on Economics and Computation*, pages 63–80, 2017.

Yang Liu and Hongyi Guo. Peer loss functions: Learning from noisy labels without knowing noise rates. arXiv, 2019. URL `https://arxiv.org/abs/1910.03231`.

Yang Liu and David P. Helmbold. Online learning using only peer prediction. arXiv, 2019. URL `https://arxiv.org/abs/1910.04382`.

Yang Liu and Mingyan Liu. An online learning approach to improving the quality of crowd-sourcing. In *Proceedings of the 2015 ACM SIGMETRICS*, pages 217–230, New York, NY, USA, 2015. ACM.

Shie Mannor and Ohad Shamir. From bandits to experts: On the value of side-observations. In *Advances in Neural Information Processing Systems*, pages 684–692, 2011.

Aditya Menon, Brendan Van Rooyen, Cheng Soon Ong, and Bob Williamson. Learning from corrupted binary labels via class-probability estimation. In *International Conference on Machine Learning*, pages 125–134, 2015.

Chris Mesterharm. On-line learning with delayed label feedback. In *International Conference on Algorithmic Learning Theory*, pages 399–413. Springer, 2005.

Nolan Miller, Paul Resnick, and Richard Zeckhauser. Eliciting informative feedback: The peer-prediction method. *Management Science*, 51(9):1359 –1373, 2005.

Nagarajan Natarajan, Inderjit S Dhillon, Pradeep K Ravikumar, and Ambuj Tewari. Learning with noisy labels. In *Advances in neural information processing systems*, pages 1196–1204, 2013.

Dražen Prelec. A bayesian truth serum for subjective data. *Science*, 306(5695):462–466, 2004.

G. Radanovic and B. Faltings. A robust bayesian truth serum for non-binary signals. In *Proceedings of the 27th AAAI Conference on Artificial Intelligence*, 2013.

Goran Radanovic, Boi Faltings, and Radu Jurca. Incentives for effort in crowdsourcing using the peer truth serum.

ACM Transactions on Intelligent Systems and Technology (TIST), 7(4):48, 2016.

Alon Resler and Yishay Mansour. Adversarial online learning with noise. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 5429–5437, Long Beach, California, USA, 09–15 Jun 2019. PMLR. URL `http://proceedings.mlr.press/v97/resler19a.html`.

Clayton Scott. A rate of convergence for mixture proportion estimation, with application to learning from noisy labels. In *AISTATS*, 2015.

V. Shnayder, A. Agarwal, R. Frongillo, and D. C. Parkes. Informed Truthfulness in Multi-Task Peer Prediction. *ACM EC*, March 2016.

Philip E Tetlock and Dan Gardner. *Superforecasting: The art and science of prediction*. Random House, 2016.

Tobias Sommer Thune, Nicolò Cesa-Bianchi, and Yevgeny Seldin. Nonstochastic multiarmed bandits with unrestricted delays. *arXiv preprint arXiv:1906.00670*, 2019.

Brendan van Rooyen and Robert C Williamson. Learning in the presence of corruption. *arXiv preprint arXiv:1504.00091*, 2015.

J. Witkowski and D. Parkes. A robust bayesian truth serum for small populations. In *Proceedings of the 26th AAAI Conference on Artificial Intelligence*, AAAI '12, 2012.