Doubly Active Learning: when Active Learning meets Active Crowdsourcing

Yang Liu[#] and Yining Wang[*]
yangl@seas.harvard.edu, SEAS, Harvard University
[*]yiningwa@cs.cmu.edu, MLD, Carnegie Mellon University

Abstract

In this paper we study an active learning setting where training samples are adaptively selected to be labeled, and the learner can only query a set of crowdsourcing workers with unknown expertise level for label information. We aim to perform active sample assignment and active worker selection jointly. Though the idea is straightforward, we face several technical challenges towards reaching our goal. First, without knowing worker's expertise level (and hence the error rate of the solicited data), design of active learning procedures can be challenging. Secondly, due to the lack of ground-truth labels, we need a learning framework that helps differentiate workers. Finally, interleaving active sample selection and active worker selection requires non-trivial effort. We propose an Upper Confidence Bound (UCB)-alike algorithm `Crowd_UCB` for selecting the best worker among a candidate pool, without knowledge of the ground-truth. Then we connect `Crowd_UCB` with a parameter free active learning algorithm. We provide performance guarantees for the proposed algorithms.

Introduction

With increasing demand for data-efficient learning methods, active learning rises as a promising option for spending label budget more efficiently. The idea is fairly simple: by carefully collecting labels of data points near the decision boundary (possibly in a sequential manner), fewer samples are required to learn a model up to a certain precision. Past works have developed extensive results on improving classification performance via active learning (Balcan *et al.* 2006; Balcan *et al.* 2007; Balcan and Long 2013; Wang 2011), where labels for only a set of selected data samples are collected.

When there is no oracle being available for querying labels, such labeling tasks are often outsourced to *crowdsourcing workers* (Abraham *et al.* 2013; Ho and Vaughan 2012; Liu and Liu 2015). Interestingly the notion of active selection has also been studied recently in the context of worker selection in crowdsourcing, where workers may contribute label information with different qualities. Similar to the idea of active learning setting where data points are carefully screened and selected for labeling, in active crowdsourcing an algorithm selects different workers to label data samples. The hope is to collect more consistent labels by exploiting the heterogeneity of a pool of workers.

In this paper, we consider performing active data sampling and worker selection jointly. In practice, actively crowdsourcing tasks within an active learning context helps save labeling budget, compared to an uniform data assignment mechanism (via avoiding labeling non-useful samples). We also provide an analytical framework to study active learning where label queries do not necessarily go to an oracle or a fixed noisy worker (with either known or unknown expertise levels). Though the idea is natural, its implementation is highly non-trivial.

On one hand, the presence of several workers with diverse labeling capabilities significantly affects the active data selection procedure, as naive data selection methods do not take into account the heterogeneity of workers' expertise levels. To make matters even worse, many existing active learning algorithms would fail without knowing properties of the noisy labeling oracles (Balcan *et al.* 2007; Balcan and Long 2013), which adds difficulties to this "doubly-active" selection problem. On the other hand, due to the lack of ground-truth (labels), how to explore or learn the workers' expertise level is not clear either. Some ideas of using redundant assignments (Karger *et al.* 2013; Karger *et al.* 2011; Abraham *et al.* 2013; Liu and Liu 2015) were proposed and studied. But such redundancy will also degrade the learning performance when facing fixed budget. For example, when each sample is assigned twice, only half of the budget will be made effective in the final learning outcome.

In this paper, we propose a novel framework for the *doubly active learning* problem. We consider the specific problem of learning a multi-dimensional homogeneous [1] binary linear classifier under Tsybakov noise conditions by querying a pool of workers (oracles) with varied, yet unknown levels of accuracy. We propose an Upper Confidence Bound (UCB)-alike algorithm `Crowd_UCB` for selecting the best worker among a candidate pool, without knowledge of the ground-truth. Then we connect `Crowd_UCB` with a parameter free margin based active learning algorithm. The proposed algorithm enjoys rigorous guarantees for its perfor-

[1]A binary linear classifier f is *homogeneous* if its decision hyperplane passes through the origin (Dasgupta 2005).

mance. In addition, for a flipping error worker model, the performance of our proposed algorithm approximates the best *active learning* performance with the *best* worker in the worker pool. We believe our model is generally applicable to many other possible worker models.

Throughout the paper, we assume no prior knowledge on the quality of each individual worker, which would instead be estimated on the fly. Proofs of some technical lemmas are placed in the supplementary material.

Related works

Active learning: As a fast progressing area, it is nearly impossible to survey all relevant works in active learning. In the classical active learning setting a learner receives a pool of *unlabeled* data points and decides, in a feedback-driven manner, a small portion of data points whose labels are requested (Hsu 2010). In most of existing active learning work only one worker (labeling oracle) is present and the only worker answers all label requests made. One notable exception is (Zhang and Chaudhuri 2015) where both a strong labeler and a weak labeler are considered at the same time. In (Zhang and Chaudhuri 2015) the qualities of both labelers are known a priori and label requests directed to the weak labeler are unlimited, both are different from our models. Being aware of noise in imperfect labelers, a relabeling based active learning algorithm was proposed in (Lin and Weld 2016) to gradually remove uncertainties in labels. Most relevant to our work is (Yan *et al.* 2011), where the idea of active learning with crowd is well executed; yet no analytical framework nor theoretical guarantees was given. Instead we provide such a framework and sample complexity results.

Active worker selection: The notion of active worker selection has been adopted primarily in a crowdsourcing setting where the worker's expertise in labeling data differs from each other and is unknown a-priori. In (Abraham *et al.* 2013), an adaptive bandit survey problem is studied and a stopping time for task assignment when facing crowds of workers is determined. (Ho and Vaughan 2012) considered an online task assignment problem in crowdsourcing market. (Liu and Liu 2015) proposed an online algorithm that learns to select the best combination of workers using redundant assignments. Both (Tran-Thanh *et al.* 2014) and (Ho *et al.* 2013) studied active worker selection under different settings (a non-active classification setting and a budget constrained setting respectively), assuming access to feedbacks on workers' qualities, following each selection. In line of game theory approaches, (Bhat *et al.* 2014) studies the issues of cost-quality balance when workers/labelers are strategic. In (Liu and Chen 2017), authors proposed a sequential peer prediction approach to incentivize high quality workers. As far as we know, none of the active worker selection works has considered an active learning setting, which turns out to be a technically challenging task as we will elaborate later.

Proactive learning: The proactive learning setting is designed specifically to relax certain unrealistic assumptions (e.g., on noise in labels) made by most active learning research (Donmez 2010). One particularly interesting case is when multiple workers (labeling oracles) are available with varied unknown qualities. Prior research on proactive learning primarily focuses on heuristics that improve performance of practical systems (Donmez and Carbonell 2008; Donmez 2010), while theoretical analysis is few. (Yang and Carbonell 2009) derived theoretically sound algorithms for the setting where workers have different bounded labeling error. However, bounded noise is a rather strong assumption and is usually not satisfied in practice.

In all, despite solid theoretical understandings on both of the active procedures separately, there is a lack of theoretical foundation for this doubly active learning question, which is one of our major motivations for carrying out this study.

Problem formulation

We assume data points and their labels $(x, y) \in \mathcal{X} \times \mathcal{Y}$ are drawn from an underlying distribution $\mathcal{P}$ jointly over $\mathcal{X} \subseteq \mathbb{R}^d$ and $\mathcal{Y} \subseteq \mathbb{R}$. The objective is to find a classifier $f : \mathcal{X} \to \mathcal{Y}$ with small generalization error, defined as

$$\text{err}(f) := \mathbb{E}_{(x,y)\sim\mathcal{P}}[l(f(x), y)].$$

We restrict ourselves to binary classification, in which $\mathcal{Y} = \{-1, +1\}$ and

$$l(f(x), y) = \mathbb{1}(y \cdot f(x) > 0)$$

is the 0/1 loss. The Bayes optimal classifier is then given by

$$f^*(x) = \text{argmax}_{y\in\{+1,-1\}}\Pr[Y = y|X = x].$$

We further assume that f^* is linear; that is, $f^*(x) = \text{sgn}(w^* \cdot x)$ for some $w^* \in \mathbb{R}^d$, $\|w\|_2 = 1$. It is well-known that the Bayes optimal classifier f^* minimizes the generalization 0/1 loss err(f). In addition, for any classifier f we define its *excess error* as err(f) − err(f^*). By definition of f^*, the excess error is always non-negative.

Tsybakov noise condition To facilitate our active learning analysis, we adopt a Tsybakov low-noise condition (TNC) on the noise part of the label distribution $Y|X$. Similar conditions were widely adopted in previous work (Balcan and Long 2013; Castro and Nowak 2008; Ramdas and Singh 2013; Wang and Singh 2016) to demonstrate the power of adaptation in classification problems. Let

$$\eta(x) = \Pr[Y = +1|X = x]$$

denote the conditional label distribution and $\phi(x, w) = \frac{\pi}{2} - \arccos\theta(x, w) \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ be the angular distance between a data point x and the hyperplane orthogonal to a linear classifier $w \in \mathbb{R}^d$. TNC then asserts that

$$\mu \cdot |\phi(x, w^*)|^{\frac{\alpha}{1-\alpha}} \le |\eta(x) - 1/2|\,, \quad (1)$$

where $0 < \mu < \infty$ and $\alpha \in (0, 1)$ are parameters that characterize behavior of the noise. Intuitively, larger α implies slower decay of noise for data points near the hyperplane of the optimal classifier w^* and hence worse rates for learning w^*.

The data request procedure

In our doubly-active learning model, a learner draws a sequence of samples $\{x_t\}$ from the marginal distribution $\mathcal{P}_X$ of $\mathcal{P}$ on $\mathcal{X}$ and decides, in a sequential and feedback-driven manner, whether to request label for a particular sample. If such a request is made, the learner further decides on a particular worker the request is sent to from a crowd of candidate workers that the learner has no prior knowledge of. In summary we have the following two active selection problems:

Active data selection: at time t, the learner obtains an unlabeled data point $x_t \sim \mathcal{P}_X$ and decides whether to query its label y_t.

Active worker selection: if a labeling request is made for x_t, the learner further decides on a worker (or a set of workers) to obtain a label y_t of x_t.

We impose a budget limit $T > 0$ on the total number of label requests made. We do not assume that each worker only stays active for one query, nor that they will stay on-call for all T queries. For practical concerns, we separate T queries into E (as a function of T, to be specified later) treatments, with each of the treatments consisting of exactly $T^* := T/E$ queries. Each treatment k will be assigned to a set of N candidate workers denoting by $\mathcal{U}_k = \{k_1, k_2, \cdots, k_N\}, k = 1, ..., E$, each with diverse expertise level and accuracy in labeling the data points. Here for simplicity we keep the number of solicited workers to be a fixed number N. In practice this number can vary across stages. For readers that are familiar with active learning literature, above "separation into treatments" also fits into the framework of margin based active learning methods, where the learning carries over separate stages, with each stage having a pre-specified length (budget, or number of queries) (Balcan and Long 2013; Wang and Singh 2016).

Worker error models

We consider a "flipping error" model, which has been widely adopted in crowdsourcing research (Abraham *et al.* 2013; Karger *et al.* 2011; Karger *et al.* 2013; Liu and Liu 2015): each worker $i \in \mathcal{U}_k$, once assigned a sample x, observes a signal $Y^{(i)}$ of the ground-truth label Y. The probability of the observation being correct is

$$p_i := \Pr[Y^{(i)} = Y | X = x],$$

which could differ for each worker. Moreover we assume the following conditional independence

$$\Pr[Y^{(i)} = Y | Y, X = x] = \Pr[Y^{(i)} = Y | Y].$$

To simplify matters, we consider the case where $p_i > 0.5$ for all $i \in \mathcal{U}$, meaning that workers perform better than taking a random guess.

First we prove that this simple (while quite common) flipping error model naturally leads to shared optimal linear Bayes estimator f^* and TNC parameter α their label error distributions among workers. The following lemma solidates above claim rigorously:

Lemma 1. *Suppose the underlying conditional label distribution* $\Pr_{Y|X}$ *satisfies TNC as in Eqn. (1) with respect to linear Bayes classifier* w^* *and parameters* $(2p_i - 1)\mu, \alpha$. *Then the Bayes classifier of conditional label distribution* $\Pr_{Y^{(i)}|X}$ *is still linear with the same* w^* *and furthermore,* $\Pr_{Y^{(i)}|X}$ *satisfies the following TNC condition:*

$$(2p_i - 1)\mu \cdot |\phi(x, w^*)|^{\alpha/(1-\alpha)} \leq |\eta_i(x) - 1/2|,$$

where $\eta_i(x) = \Pr[Y^{(i)} = +1 | X = x]$ *is the conditional labeling distribution of worker* i.

A meta-algorithm for jointly active data selection/worker selection

In order to combine active data selection and active worker selection, we first introduce two main recipes for our solution: a parameter free margin based active learning algorithm, and a UCB algorithm tailored for a budget constrained crowdsourcing setting. Together they consist our meta algorithm. The parameter free active learning is mainly to deal with the uncertainty in the underlying noise of reported data, due to the uncertainty in selecting workers, as well as in workers' expertise level. The "crowd" UCB algorithm is a variant of classical UCB, so to combat the issue that when there is no direct observation/ground-truth.

`WS16`: Noise-adaptive active learning

We are going to adopt the method presented in (Wang and Singh 2016) to serve as our background active learning algorithm, which is parameter-free. The main idea of this algorithm is to split the query budget into $E = \Theta(\log T)$ iterations and at each iteration, constrained empirical risk minimization is performed on data points near the decision boundary to push the estimated classifier $\hat{w}_k$ closer to the Bayes classifier w^*. Through an inspiring "tipping point" analysis it can be shown that Algorithm 1 with the same parameterization is capable of adapting to different TNC noise parameters. Suppose the query distribution satisfied TNC condition with parameter (μ, α) defined in (1) and the algorithm knows $\alpha_{\min}$, a lower bound of α. We will refer to this algorithm as `WS16` in this paper.

The above algorithm is useful for our crowdsourcing setting, where none of workers' expertise levels, or equivalently the noise levels in contributed data, is required to be known. The following performance guarantee has been proved in (Wang and Singh 2016) (Theorem 1):

Theorem 1. *When* $\mathcal{P}_X$ *is log-concave, with TNC parameter* μ, α *and budget* T, *with probability at least* $1 - \delta$ *we have,*

$$\textit{err}(\hat{f}_E) - \textit{err}(f^*) = \tilde{\mathcal{O}}\left(\mu^{-\frac{1-\alpha}{\alpha}} \cdot (\frac{d + \log(1/\delta)}{T})^{\frac{1}{2\alpha}}\right). \quad (2)$$

`Crowd_UCB`: UCB for crowdsourcing feedback

As we discussed earlier, we will treat each stage of our active learning algorithm as one treatment (each $k = 1, ..., E$ as in (`WS16`)) and we will be assigning each treatment k to a crowd of workers $\{k_1, ..., k_N\}$ selectively. Note within this setting the pool of workers is changing over stages. For each treatment, the goal is to find the best worker, which we will refer as the best "option" in a general context.

Algorithm 1 (`WS16`) Noise-adaptive margin-based active learning (Wang and Singh 2016)

```
Parameters: dimension d, query budget T, failure probability δ, r = e^{−(1−α_min)/α_min}.
Initialization: E = ½ log T, T* = T/E, β_0 = π, random ŵ_0 with ‖ŵ_0‖_2 = 1.
for k = 1 to E do
    W = ∅. Set b_{k−1} = (2β_{k−1}/√d) √(E(1 + log(1/r))) if k > 1 and b_{k−1} = +∞ if k = 1.
    while |W| < T* do
        Obtain sample x from P_X.
        If |ŵ_{k−1} · x| > b_{k−1} reject; otherwise request the label y of x and add (x, y) to W.
    end while
    Find ŵ_k:
        ŵ_k ∈ argmin_{w:θ(w,ŵ_{k−1})≤β_{k−1}} Σ_{(x,y)∈W} 1(yw · x < 0).
    Update β_k = rβ_{k−1}.
end for
Output: the final estimated classifier f̂_E.
```

Ideally for each stage k, we would like to run a bandit type algorithm (e.g. UCB (Auer *et al.* 2002)) for selecting workers while treating each of them as an arm, in hope that the selection will converge to selecting the most competitive workers. A salient challenge here is that for each assignment, we do not directly observe workers' labeling quality, as whether a specific task has been correctly labeled or not is not clear (missing feedback for explorations). To tackle this issue, we are going to use *redundant assignments* (Abraham *et al.* 2013; Liu and Liu 2015) towards addressing this issue. Nevertheless, a direct deployment of such techniques will significantly reduce the effectiveness of query budgets. This observation inspires our algorithm `Crowd_UCB`. The basic idea is that we are going to learn a stopping time for UCB algorithm, that we only try to solicit feedback using redundant assignment before this point. Suppose to elicit "feedback" on each arm selection, we run a subroutine `Feed_Elicit` to obtain an informative feedback – this will be made clear within specific algorithm settings.

Consider each stage of our active learning algorithm (`WS16`)).[2] Denote X_i as the random variable for generating ith arm's reward statistics (e.g., worker's labeling accuracy, or a function of his accuracy). Suppose $\mathbb{E}[X_1] > \mathbb{E}[X_2] > ... > \mathbb{E}[X_N]$, and denote by $\Delta_i := \mathbb{E}[X_1] - \mathbb{E}[X_i], \forall i > 1$. Denote by $\tilde{X}_i(t)$ the sample mean estimation for option i based on observed feedbacks. Denote the number of selection of each option i up to time t as $N_i(t)$. Define a confidence function $U(t, \delta)$ for any $\epsilon \in (0, 1)$ and $\delta \in (0, \log(1+\epsilon)/e)$ as follows ((Jamieson and Nowak 2014)):

$$U(t,\delta) := (1+\sqrt{\epsilon})\sqrt{\frac{(1+\epsilon)\log(\frac{\log((1+\epsilon)t)}{\delta})}{2t}}. \quad (3)$$

Worker selection follows an index policy:

$$i^*(t) \in \text{argmax}_i I_i(t) := \tilde{X}_i(t) + \texttt{bias}_i(t), \quad (4)$$

where $\texttt{bias}_i(t)$ is defined as

$$\texttt{bias}_i(t) = (1+\beta)U(N_i(t), \delta/N),$$

for some positive constant $\beta > 0$. Denote

$$\kappa := (\frac{2+\beta}{\beta})^2\left(1 + \frac{\log(2\log((\frac{2+\beta}{\beta})^2 N/\delta))}{\log N/\delta}\right)$$

and define the following stopping criteria for `Crowd_UCB` ((Jamieson and Nowak 2014)), and denote by t^* the stopping time:

- At time t, denote the arm with highest number of selection as i^*. Claim *YES* (Stop) if

$$N_{i^*}(t) \geq \kappa \sum_{i \neq i^*} N_i(t).$$

Claim i^* as the best option, and select i^* from t on.

The algorithm is summarized in Algorithm 2. The following can be proved for `Crowd_UCB`:

Lemma 2. *Let $X_1 - \mathbb{E}[X_1], X_2 - \mathbb{E}[X_2], ...$ be i.i.d. sub-Gaussian random variable with scale parameter $\sigma \leq 1/2$. W.p. at least $1-\delta$, when stop (t^*), the best option is identified. Further:*

$$N_i(t^*) \leq 1 + \frac{2\gamma}{\Delta_i^2}\log\left(\frac{2\log(\gamma(1+\epsilon)\Delta_i^{-2})}{\delta/N}\right), \ \forall i > 1.$$

for some positive constant γ. Moreover we have

$$t^* \leq (1+\alpha)(n-1)(1 + \frac{2\gamma}{\Delta_i^2}\log\left(\frac{2\log(\gamma(1+\epsilon)\Delta_i^{-2})}{\delta/N}\right)).$$

Algorithm 2 (`Crowd_UCB`)

Input T^*, L. Run UCB over $t = 1, 2, ..., T^*$.

- At each t, check the *stopping criteria*.
- If *NO*, make selection a_t (to send query to):

$$i^*(t) = \text{argmax}_i I_i(t) := \tilde{X}_i(t) + \texttt{bias}_i(t).$$

(`Crowd_UCB`)

- Run `Feed_Elicit` to obtain informative feedback from the crowdsourcing setting.
- If *YES*, for the rest of stage t, keep selecting the claimed best option (worker), without running `Feed_Elicit` nor updating indices.

Denote the number of re-assignments for each sample (in order to run `Feed_Elicit`) at each step as K. Then the

[2] The bandit algorithm runs similarly for each stage k, so we will omit the subscript.

total number of wasted budget is bounded as follows (due to redundant assignments):

$$(K-1)N_{i^*}(t^*) + K\sum_{i\neq i^*} N_i(t^*) = [\kappa(K-1)+K]\sum_{i\neq i^*} N_i(t^*)$$
$$\leq [\kappa(K-1)+K]\bigg(1+\frac{2\gamma}{\Delta_i^2}\log\big(\frac{2\log(\gamma(1+\epsilon)\Delta_i^{-2})}{\delta/N}\big)\bigg).$$

Denote above quantity as $\bigtriangledown T(K,N,\delta,\Delta)$ (We use Δ to denote the vector of Δ_is), which is roughly on the order of $\mathcal{O}(\sum_{i>1}\frac{1}{\Delta_i^2}\log(1/\delta))$.

A meta algorithm

With above preparation, we present Algorithm 3, a meta algorithm `DAL` combing `WS16` and `Crowd_UCB`. Algorithm 3 is separated into stages as similarly presented in active learning studies. On each stage, the selection of samples is controlled by a shrinking margin (Wang and Singh 2016). For worker selection, we use `Crowd_UCB`. Note to start the algorithm we need to compute $\hat{T}$ which needs the knowledge of $\bigtriangledown T$ and further $\Delta_i, i>1$s. In fact we only need to know a lower bound on such Δs. Intuitively $\hat{T}$ is the lower bound on the number of effective data samples we collected from the best option (or workers). We have the following Lemma to bound it under (`DAL`).

Algorithm 3 (`DAL`) A meta-algorithm framework

Parameters: dimension d, query budget T, failure probability δ, $r = e^{-(1-\alpha_{\min})/\alpha_{\min}}$.
Find largest $\hat{T}$ s.t. $\hat{T} + \bigtriangledown T(K,N,\delta,\Delta)\cdot\frac{1}{2}\log\hat{T} \leq T$.
Initialization: $E = \frac{1}{2}\log\hat{T}$, $T^* = T/E$, $\beta_0 = \pi$, random $\hat{w}_0$ with $\|\hat{w}_0\|_2 = 1$.
Configure parameters and run `WS16` with $(E, \hat{T}, \beta, w_0)$.
for $k = 1$ to E **do**
  Run `WS16` for sample selection.
  Run `Crowd_UCB` with `Feed_Elicit` for T^* stages for worker selection, when a sample has been decided to query.
  At the end of each stage, only add the data collected from the identified best option (worker) for training and updating in `WS16`.
end for
Output: the final estimated classifier $\hat{f}_E$.

Lemma 3. *With (`DAL`), we have the following:* $\hat{T} \geq T - \bigtriangledown T\cdot\frac{1}{2}\log T - (\Delta T/2+1)$.

Proof. To see this, first as $\hat{T}\leq T$ we know $\log\hat{T}/2 \leq \log T/2$. Suppose $\hat{T} < T - \Delta T\cdot\log T/2 - (\Delta T+1)$. Then we have

$$\begin{aligned}&\hat{T}+1+\Delta T\cdot\log(\hat{T}+1)/2\\ \leq&\hat{T}+1+\Delta T\cdot\log\hat{T}/2+\Delta T/(2\hat{T})\\ \leq&\hat{T}+1+\Delta T\cdot\log T/2+\Delta T/2\\ <&T.\end{aligned}$$

For the first inequality we used the fact that $\log(1+y)\leq y, \forall\ 0\leq y\leq 1$. But the above contradicts the optimality of $\hat{T}$ ($\hat{T}+1$ also satisfies the condition). □

Though the solution framework is generally applicable, for different worker models, the design, analysis, index update for `Crowd_UCB`, as well as the `Feed_Elicit` subroutine differs from one to another.

Best worker selection

Goal: Since we have E stages of active learning (treatment), we have E crowds of workers. Suppose for each $k = 1,2,...,E$ we have $p_{1,k} > p_{2,k} > ... > p_{N,k}$. Under flipping errors, following Theorem 1 and Lemma 1, we know when sending queries to a worker with labeling accuracy p (consistently), the upper bound of error in the model outputted by the algorithm is inversely proportional to $2p-1$. Since $\min_k p_{1,k} > \min_k p_{2,k} > ... > \min_k p_{N,k},\ k=1,2,...,E$ we know selecting workers with the highest $p_{i,k}$ for each stage k minimizes the bound. The objective in this setting is then to design estimator $\hat{f}_E$ that matches the performance of the one with querying the best worker at each round.

Algorithm description

To complete Algorithm 3, we first propose a matching based method for feedback elicitation in Algorithm 4.

Algorithm 4 (`Feed_Elicit`)

Treat each worker as an arm.
Before *stop*, suppose at time t task x_t is assigned to worker i. Besides worker i, randomly select another worker to assign the same task. We name this worker as the reference worker, and we denote this worker for i by $r_i(t)$.
We query both workers for the label of x_t, and compare their answers $(Y^{(i)}(n), Y^{(r_i(n))}(n))$ (n-th time worker i being selected)
Define $\tilde{X}_i(t)$ in the index of `Crowd_UCB` as follows:

$$\tilde{X}_i(t) := \frac{\sum_{n=1}^{N_i(t)}\mathbb{1}(Y^{(i)}(n) = Y^{(r_i(n))}(n))}{N_i(t)}. \quad (5)$$

Our intuition for above method is as follows. Denote

$$q_{i,k} := \Pr[Y^{(i)}(n) = Y^{(r_i(n))}(n)|Y],$$

the probability of observing a match. This quantity closely relates to $p_{i,k}$:

$$q_{i,k} = \underbrace{p_{i,k}\cdot\frac{\sum_{j\neq i}p_{j,k}}{N-1}}_{\text{agree on correct answer}} + \underbrace{(1-p_{i,k})(1-\frac{\sum_{j\neq i}p_{j,k}}{N-1})}_{\text{agree on wrong answer}}.$$

Denote by $\bar{p}_k := \sum_i p_{i,k}/N$, we prove that

Lemma 4. *When* $N \geq 2/(2\bar{p}_k - 1)$,

$$q_{i,k} > q_{j,k} \Leftrightarrow p_{i,k} > p_{j,k}.$$

Proof. We prove for any k, so we omit the subscript in k. Note the following holds

$$q_i = p_i \frac{\sum_{j\neq i} p_j}{N-1} + (1-p_i)(1-\frac{\sum_{j\neq i} p_j}{N-1})$$
$$= p_i \frac{N\bar{p}-p_i}{N-1} + (1-p_i)(1-\frac{N\bar{p}-p_i}{N-1})$$
$$= -\frac{2}{N-1}p_i^2 + (\frac{2N\bar{p}}{N-1} - 1 + \frac{1}{N-1})p_i + 1 - \frac{N}{N-1}\bar{p}\,.$$

When $N \geq 2/(2\bar{p}-1)$, the optimizer (for above quadratic equation in p_i) happens at:

$$-\frac{\frac{2N\bar{p}}{N-1} - 1 + \frac{1}{N-1}}{2(-\frac{2}{N-1})}$$
$$= \frac{1}{2} + \frac{N}{4}(2\bar{p}-1)$$
$$\geq \frac{1}{2} + \frac{1}{2} = 1.$$

So q_i is increasing at the region of $0 \leq p_i \leq 1$. □

Hence in the selection process we can use $q_{i,k}$ to serve as a surrogate for $p_{i,k}$, i.e., selecting the worker with highest $q_{i,k}$ is equivalent with selecting the one with the best $p_{i,k}$. Note in this case the number of re-assignments for each data point at each step is $K = 2$. The algorithm is summarized in Algorithm 5.

Algorithm 5 (DAL_1)

1: Run WS16 with $(E, \hat{T}, \beta, w_0)$. At stage $k = 1, ..., E$
   - At time $t = 1, ..., T^*$, once a sample is sent to query, follow Crowd_UCB for worker selection.
   - Feed_Elicit: If not stopped, re-assign the task to a randomly selected another worker, and check whether the two answers match with each other.
   - $t := t + 1$. Update $I_i(t)$(Eqn.(4)), if not stopped.

2: $k := k + 1$, update training data set and margin.

3: Output the final classifier $\hat{f}_E$.

Sample complexity results

In order to apply Lemma 2 to our analysis, we need to check the sub-gaussian assumption. First we show that

Lemma 5. *Any bounded zero mean $\mathbb{E}[X] = 0$ random variable $a \leq X(\omega) \leq b$ is a sub-Gaussian random variable with scaling parameter $\sigma \leq (b-a)/2\sqrt{2}$.*

When $b - a \leq \sqrt{2}$ we know we have $\sigma \leq 1/2$, i.e., the condition in Lemma 2 will hold. When we take

$$X_i := \mathbb{1}(Y^{(i)} = Y^{(r_i)}),$$

notice $X_i - \mathbb{E}[X_i]$ is a zero random variable, and further

$$\max(X_i(\omega) - \mathbb{E}[X_i]) - \min(X_i(\omega) - \mathbb{E}[X_i]) = (1-q_i) - (0-q_i) = 1$$

So indeed this random variable is sub-Gaussian with $\sigma < 1/2$ according to Lemma 5. Then we can safely apply Lemma 2 to bound the number of wasted budget on selecting sub-optimal workers and redundant assignments. Denote $\Delta_i := \min_k q_{1,k} - q_{i,k}$ and $p^* := \min_k p_{1,k}$, we have:

Theorem 2. *With (DAL_1), w.p. $\geq 1-\delta$,*

$$err(\hat{f}_E) - err(f^*) = \tilde{\mathcal{O}}\Big(((2p^*-1)\mu)^{-\frac{1-\alpha}{\alpha}}(\frac{d+\log(2/\delta)}{\hat{T}})^{\frac{1}{2\alpha}}\Big),$$

$$\hat{T} \geq T - \triangledown T(K=2, N, \delta/2E, \Delta)\cdot(\log T/2+1) - 1.$$

Proof. First the number of samples $N_1(t)$ for claimed best option is bounded as (with high probability $\geq 1-\delta/2E$, using Lemma 2 and Lemma 3)

$$N_1(t) \geq \frac{T}{E} - \triangledown T \geq \frac{\hat{T}}{E},$$

with probability $\geq 1-\delta/2E$. Then via union bound (E number of events with probability $\delta/2E$), with probability $1-\delta/2$, the number of samples collected from the best worker per stage is at least $\hat{T}/E$, and the total number of samples collected from all stages is at least $\hat{T}$. Using Theorem 1 we know (using union bound to composite two $\delta/2$ probability events) w.p. $\geq 1-\delta$ that

$$\text{err}(\hat{f}_E) - \text{err}(f^*) = \tilde{\mathcal{O}}\Big(((2p^*-1)\mu)^{-\frac{1-\alpha}{\alpha}}(\frac{d+\log(2/\delta)}{\hat{T}})^{\frac{1}{2\alpha}}\Big).$$

□

Best combination of workers

In this section we explore how to find a best combination of workers, instead of targeting the single best ones.

Feedback elicitation

We would like to find a set of workers S that maximizes the probability of obtaining a correct majority voting answer over them, which writes as follows

$$p_S := \Pr\Big[\mathbb{1}(\sum_{i\in S} Y^{(i)})/|S| \geq 0.5) = Y\Big].$$

It is shown in (Liu and Liu 2015) that p_S can be written in a closed-form of $p_{i,k}, i \in S$:

$$p_S = \underbrace{\sum_{S':S'\subseteq S, |S'|\geq \lceil\frac{|S|+1}{2}\rceil} \prod_{i\in S'} p_{i,k} \cdot \prod_{j\in S\setminus S'}(1-p_{j,k})}_{\text{Majority wins}} + \underbrace{\frac{\sum_{S':S'\subseteq S, |S'|=\frac{|S|}{2}} \prod_{i\in S'} p_{i,k} \cdot \prod_{j\in S\setminus S'}(1-p_{j,k})}{2}}_{\text{Ties broken equally likely}}.$$

Not hard to see, we need to learn $p_{i,k}$ accurately in order to do so.

Our method is again built on checking how often workers' answers match each other. The difference is that we will allow self-comparison in that a sample point x can be reassigned to the same worker and obtain two independent copies of answers. Then the probability of observing a matching in such a case becomes

$$q_{i,k} = p_{i,k}\frac{\sum_{j=1}^{N} p_{j,k}}{N} + (1-p_{i,k})(1-\frac{\sum_{j=1}^{N} p_{j,k}}{N}) \\ := p_{i,k}\bar{p}_k + (1-p_{i,k})(1-\bar{p}_k), \quad (6)$$

where $\bar{p}_k$ denotes the average labeling accuracy. This assumption is reasonable if we can regard each worker's answer as a collected opinion from a set of workers, instead of from an individual worker (a similar notion and assumption adopted in (Abraham *et al.* 2013)). In fact, this phenomena, named "crowd within", has been testified in real human experiments (Vul and Pashler 2008). Sum over all $i \in \mathcal{U}_k$ in Eqn.(6) and re-arrange we have

$$\sum_i q_{i,k} = (2\bar{p}_k - 1)N\bar{p}_k + N(1-\bar{p}_k),$$

which gives us

$$2\bar{p}_k^2 - 2\bar{p}_k + 1 = \sum_i q_{i,k}/N.$$

$\bar{p}_k$ can then be solved from above equation, if knowing $\bar{q}_k := \sum_i q_{i,k}/N$ (through estimating the matching probability). Further it has a unique solution in the regime $[1/2, 1]$, where $\bar{p}_k$ lies in (as we have assumed $p_{i,k} > 0.5, \forall i, k$). Particularly the solution has the following format:

$$\bar{p}_k := \frac{1}{2} + \frac{\sqrt{1-2(1-\bar{q}_k)}}{2}.$$

The above solution is well defined as

$$\bar{q}_k = 2\bar{p}_k^2 - 2\bar{p}_k + 1 \geq 1/2.$$

Plug back $\bar{p}_k$ to equation Eqn. (6) we can solve for each $p_{i,k}$, as a function of $q_{i,k}, \bar{p}_k$:

$$p_{i,k} = \frac{q_{i,k} + \bar{p}_k - 1}{2\bar{p}_k - 1}.$$

With estimates of $p_{i,k}$s, we are able to compute p_S with fine accuracy. This will be our `Feed_Elicit` step. The above method also inspires us to take each set S as a worker, instead of each individuals. Then by Lemma 1, we know the generated labels from majority voting (can also be regarded as a flipping error model) also satisfy TNC condition. For demonstration purpose, we consider a special case $p_{i,k=1} = p_{i,k=2} = ... = p_{i,k=E}, \forall i = 1, 2, ..., N$. That is in practice, at each stage we are able to recruit workers with roughly the same expertise level. For the rest of the subsection we will drop the index on k.

Sample complexity results

Based on Lemma 2, when set S is selected for labeling each sample, redefine $T := T/|S|$ and $E = \log T/2$ we have the following results ($\alpha > 1/2$): w.p. $\geq 1-\delta$:

$$\text{err}(\hat{f}_E) - \text{err}(f^*) \\ \leq \mathcal{O}\Big((\frac{\mu}{2p_S - 1})^{\frac{1-\alpha}{\alpha}} (\frac{d + \log(1/\delta)}{T/|S|})^{\frac{1}{2\alpha}} \Big) \\ \leq \mathcal{O}\Big((\frac{|S|\mu}{2p_S - 1})^{\frac{1-\alpha}{\alpha}} (\frac{d + \log(1/\delta)}{T})^{\frac{1}{2\alpha}} \Big),$$

Knowing $p_{i,k}$s enables us to find the subset of workers who jointly minimizes the above bound

$$S^* = \text{argmin}_{S \subseteq \mathcal{U}} \frac{|S|}{2p_S - 1}.$$

This metric $\frac{|S|}{2p_S-1}$ captures the trade-off between accuracy and budget loss in redundant assignment. Denote by Δ_S the gap between a sub-optimal set S and the optimal S^* in above metric:

$$\Delta_S := \frac{2p_S - 1}{|S|} - \frac{2p_{S^*} - 1}{|S^*|}, \forall S \neq S^*,$$

and define the sample mean term in `Crowd_UCB` as follows:

$$\tilde{X}_S(t) := \frac{2\tilde{p}_S(\{\tilde{p}_{i,k}(t)\}_{i\in S}) - 1}{|S|},$$

where $\tilde{p}_S(\{\tilde{p}_{i,k}(t)\}_{i\in S})$ is the estimated p_S – note we flip denominator with numerator to revert the goal from finding the minima to finding the maxima. Also note that now we have more than N arms in `Crowd_UCB`. In fact the number of arms corresponds to the number of worker combination S – denote this number as N_S. First we establish a confidence bound on $\tilde{X}_S(t)$:

Lemma 6. *With probability at least $1-\delta$, there exists a constant $C > 0$ s.t.,*

$$|\tilde{X}_S(t) - \frac{2p_S - 1}{|S|}| \leq C \cdot U(N_i(t), \delta),$$

where $U(N_i(t), \delta)$ is as defined in Eqn. (3).

Before the algorithm claims a *stop*, when S is selected, we need to update the matching probability for each $i \in S$. This incurs an exploration complexity $K = 2|S| \leq 2N$. The algorithm looks very similar to Algorithm 5, except for the `Feed_Elicit` step. We state the results, without re-stating the details.

Theorem 3. *With adapting (`DAL_1`) to above worker combination setting, with probability at least $1-\delta$ that,*

$$err(\hat{f}_E) - err(f^*) = \tilde{\mathcal{O}}\Big(((2p_{S^*} - 1)\mu)^{-\frac{1-\alpha}{\alpha}} (\frac{d + \log(2/\delta)}{\hat{T}/|S^*|})^{\frac{1}{2\alpha}} \Big),$$

$$\hat{T} \geq T - \triangledown T(K = 2N, N_S, \delta/2E, \Delta/C)(\log T/2 + 1) - 1.$$

Conclusion

In this paper we propose a solution for a doubly active learning problem that performs active learning and active crowd worker selection (active crowdsourcing) jointly. We propose an analytical framework for analyzing this doubly active learning question. Algorithms under flipping worker noise models are given, and we provide sample complexity results for them to establish their performance guarantees. Future work includes relaxation of the log-concave $\mathcal{P}_X$ assumption and development of provably correct computational efficient algorithms for the doubly-active learning problem.

References

[Abraham *et al.* 2013] Ittai Abraham, Omar Alonso, Vasilis Kandylas, and Aleksandrs Slivkins. Adaptive crowdsourcing algorithms for the bandit survey problem. In *COLT*, 2013.

[Auer *et al.* 2002] Peter Auer, Nicolò Cesa-Bianchi, and Paul Fischer. Finite-time Analysis of the Multiarmed Bandit Problem. *Mach. Learn.*, 47:235–256, May 2002.

[Balcan and Long 2013] Maria-Florina Balcan and Philip Long. Active and passive learning of linear separators under log-concave distributions. In *COLT*, 2013.

[Balcan *et al.* 2006] Maria-Florina Balcan, Alina Beygelzimer, and John Langford. Agnostic active learning. In *ICML*, 2006.

[Balcan *et al.* 2007] Maria-Florina Balcan, Andrei Broder, and Tong Zhang. Margin based active learning. In *COLT*, 2007.

[Bhat *et al.* 2014] Satyanath Bhat, Swaprava Nath, Sujit Gujar, Onno Zoeter, Y Narahari, and Chris Dance. A mechanism to optimally balance cost and quality of labeling tasks outsourced to strategic agents. In *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*, pages 917–924. International Foundation for Autonomous Agents and Multiagent Systems, 2014.

[Castro and Nowak 2008] Rui M Castro and Robert D Nowak. Minimax bounds for active learning. *Information Theory, IEEE Transactions on*, 54(5):2339–2353, 2008.

[Cesa-Bianchi and Lugosi 2006] Nicolo Cesa-Bianchi and Gábor Lugosi. Prediction, learning, and games. 2006.

[Dasgupta 2005] Sanjoy Dasgupta. Coarse sample complexity bounds for active learning. In *NIPS*, 2005.

[Donmez and Carbonell 2008] Pinar Donmez and Jaime G Carbonell. Proactive learning: cost-sensitive active learning with multiple imperfect oracles. In *CIKM*, 2008.

[Donmez 2010] Pinar Donmez. Proactive learning: towards learning with multiple imperfect predictors. Technical Report CMU-LTI-10-002, Carnegie Mellon University, 2010.

[Ho and Vaughan 2012] Chien-Ju Ho and Jennifer Wortman Vaughan. Online task assignment in crowdsourcing markets. In *AAAI'12*, pages –1–1, 2012.

[Ho *et al.* 2013] Chien-Ju Ho, Shahin Jabbari, and Jennifer W Vaughan. Adaptive task assignment for crowdsourced classification. In *Proceedings of the 30th International Conference on Machine Learning (ICML-13)*, pages 534–542, 2013.

[Hsu 2010] Daniel Hsu. *Algorithms for active learning*. PhD thesis, University of California, San Diego, 2010.

[Jamieson and Nowak 2014] Kevin Jamieson and Robert Nowak. Best-arm identification algorithms for multi-armed bandits in the fixed confidence setting. In *Information Sciences and Systems (CISS), 2014 48th Annual Conference on*, pages 1–6. IEEE, 2014.

[Jamieson *et al.*] Kevin G Jamieson, Matthew Malloy, Robert D Nowak, and Sébastien Bubeck. lil'ucb: An optimal exploration algorithm for multi-armed bandits.

[Karger *et al.* 2011] David R Karger, Sewoong Oh, and Devavrat Shah. Iterative learning for reliable crowdsourcing systems. In *Advances in neural information processing systems*, pages 1953–1961, 2011.

[Karger *et al.* 2013] David R Karger, Sewoong Oh, and Devavrat Shah. Efficient crowdsourcing for multi-class labeling. In *ACM SIGMETRICS Performance Evaluation Review*, volume 41, pages 81–92. ACM, 2013.

[Lin and Weld 2016] Christopher H Lin and Daniel S Weld. Re-active learning: Active learning with relabeling. In *AAAI*, 2016.

[Liu and Chen 2017] Yang Liu and Yiling Chen. Sequential peer prediction: Learning to elicit effort using posted prices. In *AAAI*, 2017.

[Liu and Liu 2015] Yang Liu and Mingyan Liu. An online learning approach to improving the quality of crowdsourcing. In *SIGMETRICS*, 2015.

[Ramdas and Singh 2013] Aaditya Ramdas and Aarti Singh. Algorithmic connections between active learning and stochastic convex optimization. In *ALT*, 2013.

[Tran-Thanh *et al.* 2014] Long Tran-Thanh, Sebastian Stein, Alex Rogers, and Nicholas R Jennings. Efficient crowdsourcing of unknown experts using bounded multi-armed bandits. *Artificial Intelligence*, 214:89–111, 2014.

[Vul and Pashler 2008] Edward Vul and Harold Pashler. Measuring the crowd within probabilistic representations within individuals. *Psychological Science*, 19(7):645–647, 2008.

[Wang and Singh 2016] Yining Wang and Aarti Singh. Noise-adaptive margin-based active learning and lower bounds under tsybakov noise condition. In *AAAI*, 2016.

[Wang 2011] Liwei Wang. Smoothness, disagreement coefficient, and the label complexity of agnostic active learning. *Journal of Machine Learning Research*, 12:2269–2292, 2011.

[Yan *et al.* 2011] Yan Yan, Glenn M Fung, Rómer Rosales, and Jennifer G Dy. Active learning from crowds. In *Proceedings of the 28th international conference on machine learning (ICML-11)*, pages 1161–1168, 2011.

[Yang and Carbonell 2009] Liu Yang and Jaime Carbonell. Cost complexity of proactive learning via a reduction to realizable active learning. Technical Report CMU-ML-09-113, Carnegie Mellon University, 2009.

[Zhang and Chaudhuri 2015] Chicheng Zhang and Kamalika Chaudhuri. Active learning from weak and strong labelers. In *NIPS*, 2015.

Supplementary material

Proof for Lemma 1

Proof. First we show minimizing $\Pr[f_i(X) \neq Y]$ is equivalent with minimizing $\Pr[f_i(X) \neq Y_i]$:

$$\begin{aligned}\Pr[f_i(X) \neq Y] =& \Pr[Y \neq Y_i, Y \neq f_i(X)] + \Pr[Y = Y_i, Y \neq f_i(X)] \\ =& \Pr]Y \neq f_i(X)|Y \neq Y_i] \cdot \Pr[Y \neq Y_i] \\ &+ \Pr[Y \neq f_i(X)|Y = Y_i] \cdot \Pr[Y = Y_i] \\ =& \Pr[Y_i = f_i(X)|Y \neq Y_i](1-p_i) + \Pr[Y_i \neq f_i(X)|Y = Y_i]p_i \ .\end{aligned}$$

Since we have assumed conditional independence of the two labeling error, i.e., the following two events: $\{Y_i = f_i(X)\}, \{Y = Y_i\}$, we have

$$\Pr[f_i(X) \neq Y] = (2p_i - 1)\Pr[Y_i \neq f_i(X)] + (1 - p_i) \ .$$

Suppose w^* minimizes LHS, and $p_i > 0.5$, we know w^* also minimizes $\Pr]Y_i \neq f_i(X)]$ which establishes the claimed equivalence. Consider $\eta_i(x)$,

$$\begin{aligned}|\eta_i(x) - 1/2| = |&= \Pr[Y^{(i)} = +1|X = x] - 1/2| \\ =& |\Pr[Y^{(i)} = +1, Y = +1|X = x] \\ &+ \Pr[Y^{(i)} = +1, Y = -1|X = x] - 1/2| \\ =& |\Pr[Y^{(i)} = +1|Y = 1, X = x)\Pr[Y = +1|X = x] \\ &+ \Pr[Y^{(i)} = +1|Y = -1, X = x)\Pr[Y = -1|X = x) - 1/2| \\ =& |\Pr[Y^{(i)} = +1|Y = +1)\eta(x) \\ &+ \Pr[Y^{(i)} = +1|Y = -1)(1 - \eta(x)) - 1/2| \ ,\end{aligned}$$

where the last equality we used the conditional independency that $\Pr_{Y^{(i)}|Y,X} = \Pr_{Y^{(i)}|Y}$, and the definition of $\eta(x)$. Then

$$\begin{aligned}|\eta_i(x) - 1/2| &= |(2p_i - 1)\eta(x) + 1 - p_i - 1/2| \\ &= (2p_i - 1)|\eta(x) - 1/2| \\ &\geq (2p_i - 1)\mu \cdot |\phi(x, w^*)|^{\alpha/(1-\alpha)} \\ &= (2p_i - 1)|\mu \cdot |\phi(x, w_i^*)|^{\alpha/(1-\alpha)}.\end{aligned}$$

This establishes the TNC condition on $\eta_i(x)$. □

Proof for Lemma 2 (Jamieson *et al.*)

Proof. To prove this lemma, we first introduce a high probability results for MAB with fixed confidence interval. The following lemma is established in (Jamieson *et al.*):

Lemma 7. *Let $X_1, X_2, \ldots$ be i.i.d. sub-Gaussian random variable with scale parameter $\sigma \leq 1/2$ and mean $\mu_i \in \mathbb{R}$. For any $\epsilon \in (0, 1)$ and $\delta \in (0, \log(1+\epsilon)/e)$ one has with probability at least $1 - \frac{2+\epsilon}{\epsilon/2}(\frac{\delta}{\log(1+\epsilon)})^{1+\epsilon}$ that*

$$|\frac{\sum_{s=1}^t X_s}{t} - \mu_i| \leq U(t, \delta), \ \forall t \geq 1,$$

$$\textit{where } U(t,\delta) := (1 + \sqrt{\epsilon})\sqrt{\frac{(1+\epsilon)\log(\frac{\log((1+\epsilon)t)}{\delta})}{2t}}.$$

Define the bias term in UCB index as follows:

$$\texttt{bias}_i(t) = (1 + \beta)U(N_i(t), \delta/N),$$

for some positive constant $\beta > 0$. Then define

$$\kappa := (\frac{2+\beta}{\beta})^2\left(1 + \frac{\log(2\log((\frac{2+\beta}{\beta})^2 N/\delta))}{\log N/\delta}\right).$$

Define the following stopping criteria for `Crowd_UCB`:

- At time t, denote the arm with highest number of selection as i^*. Claim *YES* if

$$N_{i^*}(t) \geq \kappa \sum_{i \neq i^*} N_i(t).$$

Claim i^* as the best option, and select i^* starting from t.

With above UCB algorithm with stopping, it is proved in (Jamieson and Nowak 2014) that

Lemma 8. *When stop, with probability at least* $1-\delta$*,*

$$N_i(t) \leq 1 + \frac{2\gamma}{\Delta_i^2} \log\left(\frac{2\log(\gamma(1+\epsilon)\Delta_i^{-2})}{\delta/N}\right).$$

Based on above lemma we know after

$$t \geq (1+\kappa)(n-1)(1 + \frac{2\gamma}{\Delta_i^2} \log\left(\frac{2\log(\gamma(1+\epsilon)\Delta_i^{-2})}{\delta/N}\right)),$$

we will be selecting the best option with probability $1-\delta$. □

Proof for Lemma 5

Proof. This can be established following the results from Lemma 2.2, Chapter 2 of (Cesa-Bianchi and Lugosi 2006):

$$\log \mathbb{E}[e^{tX}] \leq t\mathbb{E}[X] + \frac{t^2(a-b)^2}{8}$$
$$\Rightarrow \mathbb{E}[e^{tX}] \leq e^{t\mathbb{E}[X] + \frac{t^2(a-b)^2}{8}} = e^{\frac{t^2(a-b)^2}{8}}.$$

□

Proof for Lemma 6

Proof. To see why this is true, we can first estimate q_i, suppose within confidence bound $U(N_i(t), \delta)$ with probability at least $1-\delta$. Then we know

$$|\tilde{q} - \bar{q}| \leq U(N_i(t), \delta).$$

Further we show that there exists a constant L_1 s.t.,

$$|\bar{p}^{est} - \bar{p}| \leq L_1 U(N_i(t), \delta).$$

To see why this is true, first we can bound the support region when estimating $\bar{q}$. Suppose $\bar{p} \geq 1/2 + \epsilon, \epsilon > 0$ we have

$$\bar{q} = 2\bar{p}^2 - 2\bar{p} + 1 \geq 2(1/2+\epsilon)^2 - 2(1/2+\epsilon) + 1 = 2\epsilon^2 + 1/2\ .$$

First estimate $\bar{q}$ and then obtain an estimation of $\bar{p}$. The estimation error is bounded as follows

$$\begin{aligned}|\bar{p}^{est} - \bar{p}| &= |\frac{\sqrt{1-2(1-\tilde{\bar{q}})}}{2} - \frac{\sqrt{1-2(1-\bar{q})}}{2}| \\ &\leq \max_{p \geq 2\epsilon^2 + 1/2} \frac{1}{2\sqrt{1-2(1-p)}} |\tilde{\bar{q}} - \bar{q}| \\ &= \frac{1}{4\epsilon} |\tilde{\bar{q}} - \bar{q}|\ ,\end{aligned}$$

where the first inequality uses mean value theorem. Using

$$p_i = \frac{q_i + \bar{p} - 1}{2\bar{p} - 1},$$

we can prove that p_i is Lipschitz in both $q_i, \bar{p}$:

$$\tilde{p}_i = \frac{\tilde{q}_i - 1 + \bar{p}^{est}}{2\bar{p}^{est} - 1}\ .$$

Now bound the estimation error for above equation:

$$\begin{aligned}|\tilde{p}_i - p_i| &= |\frac{\tilde{q}_i - 1 + \bar{p}^{est}}{2\bar{p}^{est} - 1} - \frac{\tilde{q}_i - 1 + \bar{p}}{2\bar{p} - 1}| \\ &\le |\frac{\tilde{q}_i - q_i + q_i - 1 + \bar{p}^{est}}{2\bar{p}^{est} - 1} - \frac{q_i - 1 + \bar{p}}{2\bar{p} - 1}| \\ &\le |\frac{\tilde{P}^i_{\text{match}} - P^i_{\text{match}}}{2\bar{p}^{est} - 1}| + \frac{|P^i_{\text{match}} - 1/2|}{2}|\frac{1}{\bar{p}^{est} - 1/2} - \frac{1}{\bar{p} - 1/2}| \\ &\le |\frac{\tilde{q}_i - q_i}{2\epsilon}| + \frac{|q_i - 1/2|}{2}\frac{1}{\epsilon^2}|\bar{p}^{est} - \bar{p}| \\ &\le (1/2\epsilon + |q_i - 1/2|/8\epsilon^3)|\tilde{q}_i - q_i| \\ &\le (1/2\epsilon + 1/16\epsilon^3)|\tilde{q}_i - q_i| \;.\end{aligned}$$

Then

$$\begin{aligned}|\tilde{p}_{i,k} - p_{i,k}| &\le L_2 U(N_i(t), \delta) + L_3 \cdot L_1 U(N_i(t), \delta) \\ &= (L_2 + L_1 L_3) \cdot U(N_i(t), \delta).\end{aligned}$$

According to Lemma 7 in (Liu and Liu 2015), we know p_S is Lipschitz in $\tilde{p}_{i,k}$s ($i \in S$), so

$$|\tilde{p}_S - p_S| \le L_4(L_2 + L_1 L_3) \cdot U(N_i(t), \delta).$$

The last step is on establishing a Lipschitz condition of $\frac{|S|}{2p_S - 1}$ on p_S so we finish the proof. First since $p_i > 1/2, \forall i$, denote the slack by ϵ such that $p_i \ge 1/2 + \epsilon$. We can then easily prove for any S, $p_S \ge 1/2 + \epsilon$. Then according to Lemma 7 in (Liu and Liu 2015) we know for any subset S

$$\begin{aligned}&|\frac{|S|}{2\tilde{p}_S(\{\tilde{q}_i(t)\}_{i \in S}) - 1} - \frac{|S|}{2p_S - 1}| \\ \le &\frac{|S|}{\epsilon^2}|\tilde{p}_S(\{\tilde{q}_i(t)\}_{i \in S}) - p_S| \\ \le &\frac{N}{\epsilon^2}\sum_{i \in S}|\tilde{p}_i(t) - p_i(t)| \;,\end{aligned}$$

where the first inequality is due to mean value theorem, and the boundedness in p_S, and the second inequality is due to Lemma 7 in (Liu and Liu 2015).

□

Argument for Δ/C gap bandit selection

At time t when $i \neq 1$ is selected we have w.h.p.,(as similarly argued in (Jamieson and Nowak 2014))

$$\begin{aligned}&-\mathbb{E}[X_i] + (2 + \beta)CU(N_i(t), \delta/N_S) \ge \tilde{X}_i(t) + (1 + \beta)CU(N_i(t), \delta/N_S) \\ &\ge \tilde{X}_1(t) + (1 + \beta)CU(N_1(t), \delta/N_S) \ge \mathbb{E}[X_1] + \beta CU(N_1(t), \delta/N_S),\end{aligned}$$

from which we see when $(2 + \beta)CU(N_i(t), \delta/N_S) \ge \Delta_i$ no regret will be incurred.