No-Regret and Incentive-Compatible Prediction with Expert Advice

Rupert Freeman
Microsoft Research
New York, NY 10011
rupert.freeman@microsoft.com

David M. Pennock
Microsoft Research
New York, NY 10011
dpennock@microsoft.com

Chara Podimata
Harvard University
Cambridge, MA 02138
podimata@g.harvard.edu

Jennifer Wortman Vaughan
Microsoft Research
New York, NY 10011
jenn@microsoft.com

Abstract

We study the problem of prediction with expert advice in a setting in which experts act strategically to maximize their influence on the learning algorithm's prediction by potentially misreporting their beliefs about the binary event to be realized. Our goal is twofold. On the one hand, we want a learning algorithm to be no-regret when comparing against the best fixed expert in hindsight. On the other, we want to guarantee that each expert's best strategy, irrespective of the strategies of other experts, is to report their true belief about the realization of the event. Towards achieving this goal, we first show that any expert learning algorithm's update rule has an equivalent interpretation as a wagering mechanism, a type of multi-agent scoring rule. When experts are *myopic* (i.e., wishing to maximize their influence while looking only one step into the future), we show that using a variant of a known wagering mechanism, one can achieve both incentive-compatibility and asymptotically optimal regret guarantees. When experts are not myopic, we identify an incentive-compatible algorithm with low regret in practice.

1 Introduction

Prediction with expert advice is a fundamental problem in online learning [2, 3, 7, 12, 13]. In a nutshell, it is based on the following protocol: a learner wishes to predict an unknown sequence of outcomes, $(r_t)_{t=1}^T, r_t \in \{0,1\}$, and at each $t \in [T]$ she has access to a pool of experts $i \in [K]$, each of which issues a prediction $p_{i,t} \in [0,1]$ for r_t. The learner chooses her report $\bar{p}_t$, and subsequently, the true outcome r_t is revealed. Finally, both the learner and the experts get scored according to loss function $\ell : [0,1] \times \{0,1\} \to [0,1]$. The learner's goal is to report a sequence of predictions $(\bar{p}_t)_{t=1}^T$ that is not much worse than the sequence of predictions of the best-fixed expert in hindsight.

But what if the experts that the learner consults are strategic actors who may make predictions that do not represent their true beliefs? As was pointed out by Roughgarden and Schrijvers [10], such perverse incentives could occur when the learner is not only making predictions, but is also (implicitly or explicitly) ranking the experts. For example, the website `FiveThirtyEight`[1] not only predicts election results by aggregating different pollsters, but also publicly scores the pollsters, in a way that correlates with the amount of influence that the pollsters have over the `FiveThirtyEight` aggregate. Pollsters may, therefore, reasonably seek to maximize their score as a way of enhancing

[1] https://fivethirtyeight.com/

NeurIPS 2019 Workshop on Machine Learning with Guarantees, Vancouver, Canada.

their reputation and credibility. In a similar manner, competitors in IARPA-funded geopolitical forecasting contests typically elicit predictions from individual people, and aggregate them into a single prediction. More accurate forecasters have greater influence over the aggregate prediction, and in many cases receive monetary payments and valuable perks.[2]

When the learning algorithm is not designed in such a way that reporting true beliefs is aligned with the experts' motivation, we say that it fails to be *incentive compatible*. Perhaps not surprisingly, all standard prediction from expert advice algorithms fail incentive compatibility, in the sense that experts can sometimes achieve a greater influence on the algorithm's prediction by lying about their belief. Incentive compatibility is desirable for several reasons. First, when experts do not report truthfully, then the learner's prediction may be harmed, since the reports do not faithfully represent beliefs. Second, learning algorithms that fail incentive compatibility place an additional layer of cognitive burden on the experts, who must now reason about the details of the algorithm and other experts' reports and beliefs in order to decide how to act optimally. Our goal in this work is to design incentive-compatible algorithms for prediction from expert advice, without compromising on the quality of the algorithm's predictions, as measured by the notion of *strategic* regret (i.e., regret with respect to the *beliefs*, rather than the *reports* of the experts).

Our work is most closely related to the work of Roughgarden and Schrijvers [10], but differs from theirs in several important ways. Most crucially, in the model of Roughgarden and Schrijvers [10], the experts' incentives are only affected by the *unnormalized* weights that the learner's algorithm assigns to them, while in our work, the experts' incentives are tied to the *normalized* weights (i.e., the probabilities). We argue that normalized weights better reflect experts' incentives in reality, since reputation tends to be relative more than absolute; put another way, doubling the unnormalized weight of every expert should not increase the experts' utility, since their ultimate influence over the learner's prediction remains the same. Under Roughgarden and Schrijver's model, the design problem is fairly simple when the loss function is a proper loss [9]—that is, one that can be elicited by a proper scoring rule [1, 4, 8, 11], such as the quadratic loss function—and can be solved with a multiplicative weights algorithm. Because of this, they focus primarily on the absolute loss function, which is not a proper loss. In contrast, in our model, the design problem is nontrivial even for these "easier" proper loss functions.

Our contributions. Our goal is to study how the performance guarantee of a learner degrades—if at all—once she wants to guarantee that the experts that she consults are giving her honest predictions. Somewhat surprisingly, we show that there exists an algorithm that simultaneously satisfies incentive compatibility, guaranteeing experts' truthfulness, and guarantees asymptotically optimal competition with the best fixed expert in hindsight. Our results are enabled by a novel connection between the literature on *wagering mechanisms* [5, 6] and online learning. Additionally, our proof for the no-regret guarantee of our algorithm is significantly simpler than previously known proofs for no-regret. Following Roughgarden and Schrijvers [10], we focus primarily on the case in which experts strategize only one round into the future (i.e., the myopic case). However, we also obtain a partial extension to the non-myopic case, by identifying an algorithm that is incentive compatible and consistently achieves low regret in simulations. Proving that it is no regret is a direction of active inquiry.

2 Model and Preliminaries

Fix a bounded proper loss function ℓ. The protocol of the repeated interaction between the learner, the experts and the environment is as follows.

For all timesteps $t \in [T]$:

1. (a) Each expert $i \in [K]$ has a private belief $b_{i,t} \in [0, 1]$ about the outcome.
   (b) Each expert $i \in [K]$ reports prediction $p_{i,t} \in [0, 1]$ to the learner.
2. The learner chooses a prediction $\bar{p}_t \in [0, 1]$.
3. The learner and the experts observe the event realization $r_t \in \{0, 1\}$.

[2]One such competitor, the Good Judgment Project, now employs a network of so-called "Superforecasters".

4. The learner and the experts incur losses $\ell(\bar{p}_t, r_t)$ and $\ell(p_{i,t}, r_t), \forall i \in [K]$.

As is common in the literature, we restrict our attention to online prediction algorithms in which the learner maintains a timestep-specific probability distribution over the experts $\boldsymbol{\pi}_t = (\pi_{1,t}, \ldots, \pi_{K,t})$, and chooses her prediction $\bar{p}_t$ according to this distribution. Our results hold for both the case in which the learner randomly chooses to follow a single expert according to this distribution (so $\bar{p}_t = p_{i,t}$ with probability $\pi_{i,t}$) and the case in which the learner predicts a linear combination the experts' predictions using $\boldsymbol{\pi}_t$ as weights (so $\bar{p}_t = \sum_{t\in[T]} \pi_{i,t} p_{i,t}$).

We assume that at each timestep t, the experts act strategically in an effort to maximize the probability assigned to them at timestep $t+1$. Informally, an algorithm is incentive compatible, if all experts $i \in [K]$ maximize their (expected) probability at timestep $t+1$ when reporting $p_{i,t} = b_{i,t}$, irrespective of the reports of the others and the history of reports and realizations.

Definition 2.1 (Incentive Compatibility). *An online learning algorithm is said to be* incentive compatible *if and only if, for every expert i with belief $b_{i,t}$, every timestep $t \in [T]$, every report $p_{i,t}$, every possible vector of reports of the other experts $\mathbf{p}_{-i,t}$, and all possible histories of reports $(\mathbf{p}_{t'})_{t'<t}$ and outcomes $(r_{t'})_{t'<t}$, it holds that*

$$\begin{aligned}\mathop{\mathbb{E}}_{r_t\sim\mathtt{Bern}(b_{i,t})}&[\pi_{i,t+1}|\left(b_{i,t}, \mathbf{p}_{-i,t}\right), r_t, (r_{t'})_{t'<t}, (\mathbf{p}_{t'})_{t'<t}] \\ &\geq \mathop{\mathbb{E}}_{r_t\sim\mathtt{Bern}(b_{i,t})}[\pi_{i,t+1}|\left(p_{i,t}, \mathbf{p}_{-i,t}\right), r_t, (r_{t'})_{t'<t}, (\mathbf{p}_{t'})_{t'<t}]\end{aligned}$$

Incentive compatibility is not our only desideratum; our online prediction algorithms are also evaluated using the notion of strategic regret, formally defined below.

Definition 2.2 (Strategic Regret).

$$R\left(T\right) = \sum_{t\in[T]} \ell(\bar{p}_t, r_t) - \arg\min_{j\in[K]} \sum_{t\in[T]} \ell(b_{j,t}, r_t)$$

We remark here that since our mechanisms are incentive compatible, the definition of strategic regret coincides with the standard definition of external regret in our case.

3 Wagering Mechanisms and Online Learning

In this section, we outline our online prediction algorithm that is both incentive compatible and no regret. To do so, we first draw a connection between online prediction algorithms and *wagering mechanisms* [5, 6], a class of one-shot probabilistic elicitation mechanisms, that allow experts to bet on the quality of their prediction against others.

More formally, in the one-shot wagering setting that we consider, experts hold a belief $b_i, \forall i \in [K]$ regarding the expected value of a realization $r \in \{0, 1\}$, and they report to the mechanism a probability p_i (potentially with $p_i \neq b_i$) and a wager $w_i \geq 0$. A wagering mechanism, Γ maps reports $\mathbf{p} = (p_1, \ldots, p_K)$, wagers $\mathbf{w} = (w_1, \ldots, w_K)$, and the realization r to payments $\Gamma_i(\mathbf{p}, \mathbf{w}, r)$ for each expert i. The purpose of the wager is to allow each expert to set a maximum allowable loss, which we capture by imposing the constraint that $\Gamma_i(\mathbf{p}, \mathbf{w}, r) \geq 0, \forall i \in [K]$. We restrict our attention to *budget balanced* wagering mechanisms for which $\sum_{i\in[K]} \Gamma_i(\mathbf{p}, \mathbf{w}, r) = \sum_{i\in[K]} w_i$.

Definition 3.1 (Incentive Compatibility for Wagering Mechanisms). *A wagering mechanism Γ is said to be* incentive compatible *if and only if, for every expert $i \in [K]$, with belief $b_i \in [0, 1]$, every report $p_i \in [0, 1]$ and every possible vector of reports of the other experts $\mathbf{p}_{-i}$, it holds that*

$$\mathop{\mathbb{E}}_{r\sim\mathtt{Bern}(b_i)}\left[\Gamma_i\left(\left(b_i, \mathbf{p}_{-i}\right), \mathbf{w}, r\right)\right] \geq \mathop{\mathbb{E}}_{r\sim\mathtt{Bern}(b_i)}\left[\Gamma_i\left(\left(p_i, \mathbf{p}_{-i}\right), \mathbf{w}, r\right)\right].$$

Our key observation is that we can define a black-box reduction that transforms any wagering mechanism to an online prediction algorithm. Consider some budget balanced wagering mechanism Γ. Then we can define an online prediction algorithm that sets $\boldsymbol{\pi}_{t+1} = \Gamma(\mathbf{p}, \boldsymbol{\pi}_t, r_t)$. In words, we are thinking of an expert's weight according to distribution $\boldsymbol{\pi}$ as their currency, with each expert wagering $\boldsymbol{\pi}_t$ and receiving payoff $\boldsymbol{\pi}_{t+1}$, where the payoff depends on the reports of the experts $\mathbf{p}$

and the realization r_t.[3] It is easy to see that any online prediction algorithm that is derived from an incentive-compatible wagering mechanism will in turn be incentive compatible, because any misreport that increases weight $\boldsymbol{\pi}_{t+1}$ would also be a successful misreport in the wagering setting.

We now turn our attention to our second objective: guaranteeing vanishing regret compared to the best expert in hindsight. To do so, we will draw upon a class of wagering mechanisms known as *Weighted Score Wagering Mechanisms (WSWMs)* [5, 6]. A WSWM pays each expert

$$\texttt{WSWM}_i(\mathbf{w}, \mathbf{p}, r) = w_i \left(1 - \ell(p_i, r) + \sum_{j \in [K]} w_j \ell(p_j, r) \right) \tag{1}$$

The intuition behind a weighted score wagering mechanism is that an expert makes a profit (i.e., receives payment greater than her wager), whenever her loss is smaller than the wager-weighted average loss. Therefore, experts with accurate reports will tend to increase their wealth (which, in the context of online prediction algorithms, is equivalent to their probability of being followed by the learner), while those with less accurate reports will tend to lose wealth. Lambert et al. [5] proved that the `WSWM` defined in Equation (1) is incentive compatible.

As a first attempt, and due to the aforementioned reduction, one might hope that using Equation (1) would directly yield a no-regret online prediction algorithm. But this is not the case, due to the fact that an expert who makes an inaccurate prediction in the first timestep can lose almost all of her wealth/probability. This occurs if $\ell_{i,1}$ is close to 1, but $\ell_{j,1}$ is close to 0 for all other experts $j \neq i$. Then, the multiplying term inside the parentheses of Equation (1) is very small, meaning that i loses most of her wealth. Even if she is very accurate in the subsequent rounds, it can take a long time for her wealth to recover, in turn resulting in the learner placing insufficient probability on her reports. To handle this, we define an adjusted wagering mechanism where experts keep a $1 - \eta$ fraction of their wealth, and wager only an η fraction, for an adjustable value of $\eta \in (0, 0.5]$. This guarantees that any expert with probability $\pi_{i,t}$ close to zero must have made a long series of inaccurate predictions, not just one. Formally, the update rule of our online prediction algorithm is defined by:

$$\pi_{i,t+1} = \eta \texttt{WSWM}_i(\boldsymbol{\pi}, \mathbf{p}_t, r_t) + (1 - \eta)\pi_{i,t} \tag{2}$$

Theorem 3.1. *For step size $\eta = \sqrt{\frac{\log K}{T}}$, the online prediction algorithm defined by Equation* (2) *is* incentive compatible *in the experts' reports, and yields regret $R(T) = O\left(\sqrt{T \log K}\right)$.*

The proof is omitted due to space constraints. At a high level, we first show that Equation (2) can be rewritten as $\pi_{i,t+1} = \pi_{i,t}(1 - \eta(\ell(p_{i,t}, r_t) - \sum_{j \in [K]} \pi_{j,t} \ell(p_{j,t}, r_t)))$, which is reminiscent of the multiplicative weights update rule, modulo a non-negative shift. We then define and analyze the standard potential function $\Phi_t = \sum_{i \in [K]} \pi_{i,t}$. An interesting aspect of our proof is that, contrary to the usual proof that multiplicative weights update is no regret, $\Phi_t = 1$ for all time steps $t \in [T]$. As a result the standard analysis of the upper bound on Φ_T becomes trivial, so we only need to analyze the lower bound of Φ_T in order to prove the vanishing regret. To our knowledge, our proof is the simplest proof of no regret for the prediction with expert advice problem.

4 Ongoing and Future Work

The results presented above are only valid for the case that the experts are *myopic*, i.e., they only care about the probabilities that the prediction algorithm assigns to them at the immediate, next timestep. In many real-life applications, however, the experts can plan their predictions for a known time horizon, and they ultimately care about their relative ranking to the rest of the experts at the end of this horizon. Observe that this is a step closer to the real-life incentives of pollsters in online poll aggregators like `FiveThirtyEight`, who provide pollster rankings on a continuous basis[4]. Can one achieve both no-regret and non-myopic incentive compatibility for such settings? We have identified the `ELF` mechanism of Witkowski et al. [14] as a candidate mechanism for this setting. It is known to satisfy non-myopic incentive compatibility, but we have so far been unable to theoretically analyze its regret guarantee. However, preliminary experimental results suggest that this mechanism may be no-regret.

[3] It is worth pointing out that not all online prediction algorithms can be derived in this manner, as our construction only uses $\boldsymbol{\pi}_t$ rather than the full history of reports and realizations.

[4] https://projects.fivethirtyeight.com/pollster-ratings/

References

[1] Glenn W Brier. Verification of forecasts expressed in terms of probability. *Monthly weather review*, 78(1):1–3, 1950.

[2] Nicolo 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 (JACM)*, 44(3):427–485, 1997.

[3] Yoav Freund and Robert E Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, 55(1):119–139, 1997.

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

[5] Nicolas S Lambert, John Langford, Jennifer Wortman, Yiling Chen, Daniel Reeves, Yoav Shoham, et al. Self-financed wagering mechanisms for forecasting. In *Proceedings of the 9th ACM conference on Electronic commerce*, pages 170–179. ACM, 2008.

[6] Nicolas S Lambert, John Langford, Jennifer Wortman Vaughan, Yiling Chen, Daniel M Reeves, Yoav Shoham, and David M Pennock. An axiomatic characterization of wagering mechanisms. *Journal of Economic Theory*, 156:389–416, 2015.

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

[8] John McCarthy. Measures of the value of information. *Proceedings of the National Academy of Sciences of the United States of America*, 42(9):654, 1956.

[9] Mark D Reid and Robert C Williamson. Surrogate regret bounds for proper losses. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 897–904. ACM, 2009.

[10] Tim Roughgarden and Okke Schrijvers. Online prediction with selfish experts. In *Advances in Neural Information Processing Systems*, pages 1300–1310, 2017.

[11] Leonard J Savage. Elicitation of personal probabilities and expectations. *Journal of the American Statistical Association*, 66(336):783–801, 1971.

[12] Vladimir Vovk. A game of prediction with expert advice. *Journal of Computer and System Sciences*, 56(2):153–173, 1998.

[13] Volodimir G Vovk. Aggregating strategies. *Proc. of Computational Learning Theory, 1990*, 1990.

[14] Jens Witkowski, Rupert Freeman, Jennifer Wortman Vaughan, David M Pennock, and Andreas Krause. Incentive-compatible forecasting competitions. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.