Peer Prediction for Learning Agents

Shi Feng[1] Fang-Yi Yu[2] Yiling Chen[3]

[1]IIIS, Tsinghua University

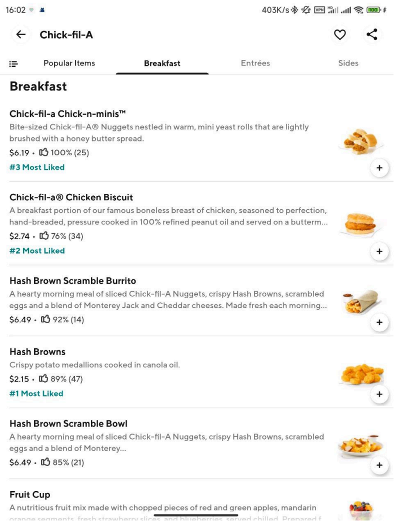
[2]CS Department, George Mason University

[3]SEAS, Harvard University

Yao Class Seminar, Sept. 2022

Overview

How to Incentivize Customers to Rate Truthfully?



DoorDash Food Ratings

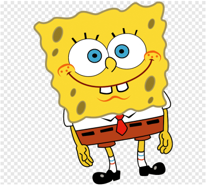
Dianping Ratings

Overview

Problem Settings

What we observe: $\hat{x}_t \in \{0,1\}$ and $\hat{y}_t \in \{0,1\}$

Alice's Private Signal: $X_t \in \{0,1\}$



Bob's Private Signal: $Y_t \in \{0,1\}$

- In each round t, Alice and Bob receive signal X_t and Y_t from a distribution $P_{X,Y}$ over $\{0,1\}^2$.
- Alice has a random variable $\hat{X}_t \in \{0,1\}$ for her report and send its realization $\hat{x}_t$ to us. So does Bob.
- **Assumption** (positive correlated signals): The distribution $P_{X,Y}$ is positively correlated, i.e.,
  $\min\{P_{X,Y}(1,1), P_{X,Y}(0,0)\} > \max\{P_{X,Y}(1,0), P_{X,Y}(0,1)\}$.

Consistent Strategy Assumption

X_t \ $\hat{X}_t$	0	1
0	$P(\hat{X}_t = 0 \| X_t = 0)$	$P(\hat{X}_t = 1 \| X_t = 0)$
1	$P(\hat{X}_t = 0 \| X_t = 1)$	$P(\hat{X}_t = 1 \| X_t = 1)$

- In all previous theory works, **Consistent Strategy Assumption** is pivotal.
- We call a strategy profile σ for Alice or Bob to be a 2×2 probability matrix giving probability of reporting $\hat{X}$ given private signal X.
- The above table is a strategy profile of Alice in the t^{th} round.
- **Consistent Strategy Assumption**: Alice and Bob use consistent (unchanged) strategy profiles σ_X and σ_Y respectively over all rounds.
- Unrealistic!!!

Goal of Peer Prediction

- In short, the final goal of peer prediction is incentivizing agents to report truthfully.
- We want to design a mechanism $\mathcal{M} = \{M_t : t \geq 1\}$ where $M_t : \{0, 1\}^{2 \times k} \to [-1, 1]$.
- After Alice and Bob reporting $\hat{\mathbf{x}}_{\leq t}$ and $\hat{\mathbf{y}}_{\leq t}$ in round t, $\mathcal{M}$ computes $(r_t, s_t) := M_t(\hat{\mathbf{x}}_{[t-k+1,t]}, \hat{\mathbf{y}}_{[t-k+1,t]})$. We call such $\mathcal{M}$ rank k mechanism.
- In total,if the number of rounds is T, we pay Alice $r_1 + r_2 + \cdots + r_T$ and pay Bob $s_1 + s_2 + \cdots + s_T$.
- We hope truthfully reporting is a BNE in this game. Or more strictly, we hope that $\mathcal{M}$ is strongly truthful.

Strongly Truthful

Definition (Strongly truthful)

In a peer prediction game, if agents are using consistent strategies, a mechanism is *strongly truthful* if and only if truthtelling is a BNE and also guarantees larger agent welfare than any non-permutation equilibrium. Here, welfare is defined by each agent's expected payoff so that is to say, the expected payoff of each agent using truthtelling strategy profile is strictly higher than the expected payoff using non-permutation equilibrium.

X_t \ $\hat{X}_t$	0	1
0	1	0
1	0	1

X_t \ $\hat{X}_t$	0	1
0	1	0
1	0	1

Permutation Strategy Profiles

Correlated Agreement (CA) Mechanism

- One strongly truthful mechanism is CA mechanism M^{CA} (rank 2).
- In the t^{th} round,

$$M_t^{CA}(\hat{\mathbf{x}}_{\leq t}, \hat{\mathbf{y}}_{\leq t}) = (\underbrace{\mathbb{I}[\hat{x}_t = \hat{y}_t] - \mathbb{I}[\hat{x}_t = \hat{y}_{t-1}]}_{r_t}, \underbrace{\mathbb{I}[\hat{y}_t = \hat{x}_t] - \mathbb{I}[\hat{y}_t = \hat{x}_{t-1}]}_{s_t}).$$

- M^{CA} is proved to be strongly truthful!
- According to consistent strategy assumption, we have $\mathbb{E}[r_2] = \mathbb{E}[r_3] = \cdots$ and $\mathbb{E}[s_2] = \mathbb{E}[s_3] = \cdots$. Thus total expected rewards of Alice and Bob are $(T-1)\mathbb{E}[r_t]$ and $(T-1)\mathbb{E}[s_t]$ for an arbitrary t, correspondingly.

An Example of Strongly Truthfulness

X_t \ $\hat{X}_t$	0	1
0	p_0	$1-p_0$
1	$1-p_1$	p_1

Consistent Strategy σ_X of Alice

Y_t \ $\hat{Y}_t$	0	1
0	q_0	$1-q_0$
1	$1-q_1$	q_1

Consistent Strategy σ_Y of Bob

- $M_t^{CA}(\hat{\mathbf{x}}_{\leq t}, \hat{\mathbf{y}}_{\leq t}) = (\underbrace{\mathbb{I}[\hat{x}_t = \hat{y}_t] - \mathbb{I}[\hat{x}_t = \hat{y}_{t-1}]}_{r_t}, \underbrace{\mathbb{I}[\hat{y}_t = \hat{x}_t] - \mathbb{I}[\hat{y}_t = \hat{x}_{t-1}]}_{s_t})$.

$$P_{X,Y}(0,0) = 0.4, P_{X,Y}(0,1) = 0.1,$$
$$P_{X,Y}(1,0) = 0.1, P_{X,Y}(1,1) = 0.4.$$

- $\mathbb{E}[r_t] = \mathbb{E}[s_t] = 0.3(p_1 + p_0 - 1)(q_1 + q_0 - 1) \leq 0.3$ and the equality holds if and only if $p_0 = p_1 = q_0 = q_1 = 1$ or $p_0 = p_1 = q_0 = q_1 = 0$ (permutation strategy profiles).

Replacement of Consistent Strategy Assumption

- Instead of sticking with one mixed strategy, real-world agents are learning to earn money!
- Can CA mechanism still guarantee that agents converge to truthfulness?
- No-regret behavior assumption on agents? ✗
- Agents using reward-based online learning algorithms? ✓

Overview

No Regret is Not Enough!

Theorem (Negative Result)

For any sequential information elicitation mechanism $\mathcal{M}$ of rank $k \in \mathbb{N}$, there exist no-regret algorithms for Alice and Bob so that $\mathcal{M}$ cannot achieve truthful convergence.

Intuition:

- No regret assumption does not prevent correlation between Alice's and Bob's strategy profiles.
- Alice and Bob can generate X'_t, Y'_t from $P_{X,Y}$ by themselves and report $\hat{X}_t = X'_t, \hat{Y}_t = Y'_t$ truthfully.
- In expectation, Alice and Bob should get same rewards as truthfully report $\hat{X}_t = X_t, \hat{Y}_t = Y_t$.

Overview

Accumulated Rewards

Agents have four pure strategies:

- truthtelling: opt_1,
- flip the private signal (report $\hat{X} = 1 - X$): opt_2,
- always report 1: opt_3,
- always report 0: opt_4.

We define $r_{i,t}$ as the reward of Alice if she uses opt_i in the t^{th} round while $\hat{\mathbf{x}}_{\leq t-1}$ and $\hat{\mathbf{y}}_{\leq t}$ are fixed. The accumulated rewards of four options for Alice is $R_{i,t} = r_{i,1} + \cdots + r_{i,t}$.
Similarly, we are able to compute $s_{i,t}$ and $S_{i,t}$.

Agents use $R_{i,t-1}, S_{i,t-1}$ to decide what to do in the t^{th} round!

Reward-Based Online Learning Algorithms

Take Alice as an example:

- Alice uses an update function $f\colon \mathbb{R}^4 \to \triangle^3$, and chooses opt_i with probability $f_i(R_{1,t-1}, R_{2,t-1}, R_{3,t-1}, R_{4,t-1})$ for $i \in [4]$ in the t^{th} round.
- **Exchangeability of** f: for any $R_1, R_2, R_3, R_4 \in \mathbb{R}$ and an arbitrary permutation of them $R_{i_1}, R_{i_2}, R_{i_3}, R_{i_4}$, $f_{i_j}(R_1, R_2, R_3, R_4) = f_j(R_{i_1}, R_{i_2}, R_{i_3}, R_{i_4})$ for all $j \in [4]$.
- **Order preservation of** f: for any $R_1, R_2, R_3, R_4 \in \mathbb{R}$ and suppose that $R_{i_1}, R_{i_2}, R_{i_3}, R_{i_4}$ is a non-increasing order of them, for f we have $f_{i_1}(R_1, R_2, R_3, R_4) \geq f_{i_2}(R_1, R_2, R_3, R_4) \geq f_{i_3}(R_1, R_2, R_3, R_4) \geq f_{i_4}(R_1, R_2, R_3, R_4)$.
- **Full exploitation of** f: $\lim_{R_1 - \max\{R_2, R_3, R_4\} \to +\infty} f_1(R_1, R_2, R_3, R_4) = 1$.
- One can directly verify that this algorithm family contains replicator dynamics, hedge algorithms and follow the perturbed leader algorithm.

Overview

Main Result

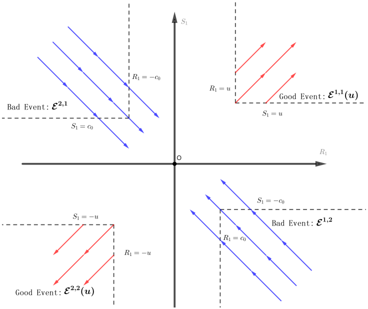
Theorem

Under the assumptions we introduced, the binary-signal, sequential CA mechanism M^{CA} achieves truthful convergence when agents use reward-based algorithms A_f and A_g, where the update functions f and g satisfy the properties in the last slide.

A few observations:

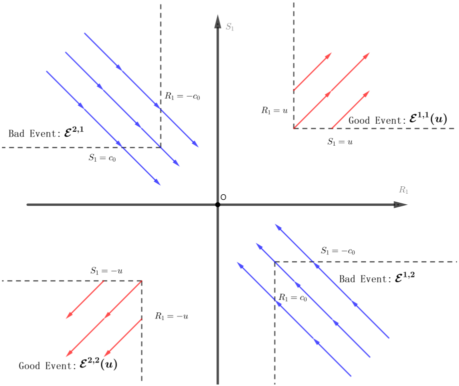
- $R_{1,t} + R_{2,t} = R_{3,t} + R_{4,t} = 0$ and $S_{1,t} + S_{2,t} = S_{3,t} + S_{4,t} = 0$.
- $R_{3,t}, R_{4,t} \in [-1, 1]$ and $S_{3,t}, S_{4,t} \in [-1, 1]$.
- Hence, a state of Alice's strategy can be approximately represented by $R_{1,t-1}$. When $R_{1,t-1} >> 1$, Alice chooses opt_1 whp; when $R_{1,t-1} << -1$, Alice chooses opt_2 whp.

Proof Idea



- **Step 1**: Prove that S_1 and R_1 cannot be on both sides of 0 and both far from 0. ($\rightarrow$)
- **Step 2**: Prove that when one of S_1, R_1 is not far from 0, they will eventually get into a state such that R_1, S_1 on both sides of 0 and both far from 0.
- **Step 3**: Prove that when R_1, S_1 are in state $\mathcal{E}^{1,1}(u)$ or $\mathcal{E}^{2,2}(u)$, R_1, S_1 will both become further from 0 in the following rounds. ($\rightarrow$)

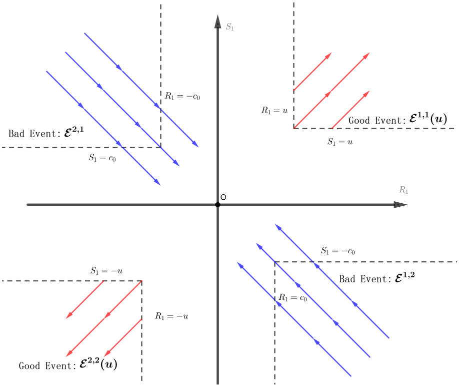
Step 1



Lemma

Given the game we defined, $\Pr\left\{\limsup_{t\to\infty} \overline{\mathcal{E}_t^{1,2} \vee \mathcal{E}_t^{2,1}} = 1\right\} = 1.$

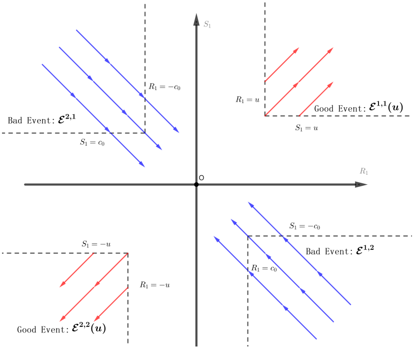
Step 2



Lemma

Given the game we defined, for all u there exists λ_u so that for any T with history $\mathcal{H}_T \in \overline{\mathcal{E}_T^{1,2} \vee \mathcal{E}_T^{2,1}}$, we have
$$\Pr\left\{\left(\vee_{i=T}^{T+4(u+c_0)+100}\mathcal{E}_t^{1,1}(u)\right) \vee \left(\vee_{i=T}^{T+4(u+c_0)+100}\mathcal{E}_t^{2,2}(u)\right) = 1 \middle| \mathcal{H}_T\right\} \geq \lambda_u.$$

Step 3

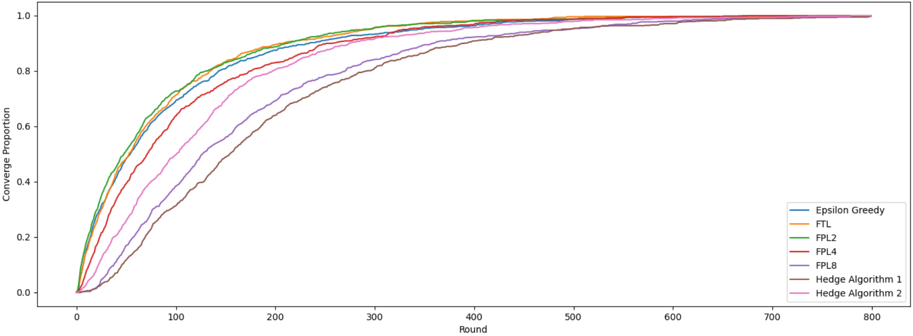


Lemma

Given the game we defined, for all $\epsilon > 0$ *there exists* $u \in \mathbb{N}^+$ *such that given a history* $\mathcal{H}_T \in \mathcal{E}_T^{1,1}(u) \vee \mathcal{E}_T^{2,2}(u)$, *we have*

$$\Pr\left\{\forall i \in \mathbb{N}, \mathcal{E}^{1,1}_{T+\left(\lceil\frac{1000}{\gamma_1-\gamma_2}\rceil+1\right)i}\left(\lfloor\frac{u}{2}\rfloor + i\right) \vee \mathcal{E}^{2,2}_{T+\left(\lceil\frac{1000}{\gamma_1-\gamma_2}\rceil+1\right)i}\left(\lfloor\frac{u}{2}\rfloor + i\right) = 1 \middle| \mathcal{H}_T\right\} \geq 1 - \varepsilon.$$

Simulations



All converge! Moreover, ϵ-greedy also converges, we expect an almost identical proof works.

Overview

Future Work

- Non-binary private signals X_t, Y_t for Alice and Bob?
- Do other mechanisms for peer prediction guarantee truthful convergence?
- More general family of learning algorithms?
- ...

Overview

Q&A

Questions?

Thank you!