Strategic Usage in a Multi-Learner Setting

Eliot Seo Shekhtman
Sarah Dean

INTRODUCTION

INTRODUCTION • SETTING • RESULTS

ONLINE MARKETPLACE SETTING

- Retailers (users) want listings (legitimate or scam) to be successful
- Platforms (services) don’t want to host scams/spam (audience trust)

- Platforms want to learn to filter out scam listings
- Retailers want to adapt strategically

[FDA APPROVED] [EPA APPROVED] [GMO FREE] sweatpants [RELIABLE]

M. Hardt, N. Megiddo, C. Papadimitriou, and M. Wootters. Strategic classification. In Proceedings of the 2016 ACM conference on innovations in theoretical computer science, pages 111–122, 2016.

MOTIVATION

Strategic Classification: single-service setting

- Studied what if retailers adapt through feature manipulation?
- Retailers make listings more believable to trick the platform

M. Hardt, N. Megiddo, C. Papadimitriou, and M. Wootters. Strategic classification. In Proceedings of the 2016 ACM conference on innovations in theoretical computer science, pages 111–122, 2016.
M. Hardt, M. Jagadeesan, and C. Mendler-Dünner. Performative Power. Advances in Neural Information Processing Systems, 36, 2022.

MOTIVATION

Strategic Classification: single-service setting

Performative Power: service's ability to impact the market

- Feature manipulation can be costly!
- In a multi-service setting, retailers might change platforms instead

M. Hardt, N. Megiddo, C. Papadimitriou, and M. Wootters. Strategic classification. In Proceedings of the 2016 ACM conference on innovations in theoretical computer science, pages 111–122, 2016.
M. Hardt, M. Jagadeesan, and C. Mendler-Dünner. Performative Power. Advances in Neural Information Processing Systems, 36, 2022.

MOTIVATION

Strategic Classification: single-service setting

Performative Power: service's ability to impact the market

Our work: multi-service setting

- Retailers only post on a platform if advantageous
- Platforms learn to filter based on their listings

MAIN RESULTS (Informal)

When services retrain naïvely:

- Retailers might avoid suppression by switching platforms endlessly

When services remember past timesteps:

- Services will learn to make accurate assessments
- Scam retailers will leave the market

SETTING

FORMALIZED SETTING

n users with d features $x_i \in \mathscr{X}$ and a label $y_i \in \{-1, 1\}$

m services with classifiers $h_j : \mathscr{X} \rightarrow \{+1, -1\}$, $h \in \boldsymbol{H}$

- Example features: listing descriptions, reviews, number of listings
- Label: “scam” or “legitimate” retailer

FORMALIZED SETTING

n users with d features $x_i \in \mathscr{X}$ and a label $y_i \in \{-1, 1\}$

m services with classifiers $h_j : \mathscr{X} \rightarrow \{+1, -1\}$, $h \in \boldsymbol{H}$

We assume realizability!

USER OBJECTIVE

USER OBJECTIVE

Users receive utility from positive classification $u : \mathscr{X} \times \boldsymbol{H} \to \mathbb{R}$

- *Assume sign of u is shared with h*
- Example: projected number of clicks on a listing, how strict their filter is

USER OBJECTIVE

Users receive utility from positive classification $u : \mathcal{X} \times \boldsymbol{H} \rightarrow \mathbb{R}$

- *Assume sign of u is shared with h*
- Users assign usage A to services that give them utility:

  This incurs cost! $1/q\ (\sum_j^m A_{ij})^q$

  Example: effort to join a platform

USER OBJECTIVE

Users receive utility from positive classification $u : \mathscr{X} \times \boldsymbol{H} \to \mathbb{R}$

- *Assume sign of u is shared with h*
- Users assign usage A to services that give them utility:

  This incurs cost! $1/q \left(\sum_j^m A_{ij}\right)^q$

Users allocate usage to maximize:

$$\sum_{j=1}^{m} A_{ij} u(x_i, h_j) - \frac{1}{q}\Big(\sum_{j=1}^{m} A_{ij}\Big)^q$$

SERVICE OBJECTIVE

SERVICE OBJECTIVE

Services observe user usages to learn about the user distribution

$$M^t = \frac{A^t}{1+p} + \frac{pM^{t-1}}{1+p}$$

SERVICE OBJECTIVE

Services observe user usages to learn about the user distribution

$$M^t = \frac{A^t}{1+p} + \frac{pM^{t-1}}{1+p}$$

Services optimize over non-negative loss $\ell : \boldsymbol{H} \times \mathscr{X} \times \mathscr{Y} \longrightarrow \mathbb{R}$

- *Utility has a strict monotonic relationship with* $-y\ell(h, x, y)$
- *There exists a* $v>0$ *such that* $u(x, h) = 0$ *when* $\ell(h, x, y) = v$

SERVICE OBJECTIVE

Services observe user usages to learn about the user distribution

$$M^t = \frac{A^t}{1+p} + \frac{pM^{t-1}}{1+p}$$

Services optimize over non-negative loss $\ell : \boldsymbol{H} \times \mathscr{X} \times \mathscr{Y} \rightarrow \mathbb{R}$

Services update to minimize the following formula:

$$\sum_{i=1}^{n} \frac{M_{ij}^t}{\sum_{k=1}^{n} M_{kj}^t} \ell(h_j, x_i, y_i)$$

FULL INTERACTION DYNAMICS

At timestep t:

$$A^t \in \operatorname*{argmax}_{A \in \mathbb{R}_+^{n \times m}} \sum_{i=1}^{n} \left[\sum_{j=1}^{m} A_{ij} u(x_i, h_j^t) - \frac{1}{q} \Big[\sum_{j=1}^{m} A_{ij} \Big]^q \right]$$

$$M^t = \frac{A^t}{1+p} + \frac{pM^{t-1}}{1+p}$$

$$H^{t+1} \in \operatorname*{argmin}_{H \in \mathcal{H}^m} \sum_{j=1}^{m} \sum_{i=1}^{n} \frac{M_{ij}^t}{\sum_{k=1}^{n} M_{kj}^t} \ell(h_j, x_i, y_i)$$ *

FULL INTERACTION DYNAMICS

At timestep t:

$$A^t \in \operatorname*{argmax}_{A \in \mathbb{R}_+^{n \times m}} \sum_{i=1}^{n} \left[\sum_{j=1}^{m} A_{ij} u(x_i, h_j^t) - \frac{1}{q} \Big[\sum_{j=1}^{m} A_{ij} \Big]^q \right]$$

$$M^t = \frac{A^t}{1+p} + \frac{pM^{t-1}}{1+p}$$

$$H^{t+1} \in \operatorname*{argmin}_{H \in \mathcal{H}^m} \sum_{j=1}^{m} \sum_{i=1}^{n} \frac{M_{ij}^t}{\sum_{k=1}^{n} M_{kj}^t} \ell(h_j, x_i, y_i)^{*}$$

* (tiebreaking must be sticky!)

RESULTS

ZERO-LOSS STATE

Definition 2. A state (H,A) is zero-loss if all services j satisfy:

1. $A_{ij}\ell(h_j, x_i, y_i) = 0$ for all $i \in \{1, \ldots, n\}$
2. $u(x_i, h_j) \leq 0$ for all i with $y_i = -1$.

ZERO-LOSS STATE

Definition 2. A state (H,A) is zero-loss if all services j satisfy:

1. $A_{ij}\ell(h_j,x_i,y_i) = 0$ for all $i \in \{1, \ldots, n\}$
2. $u(x_i,h_j) \leq 0$ for all i with $y_i = -1$.

All services make accurate classifications on the populations they observe

ZERO-LOSS STATE

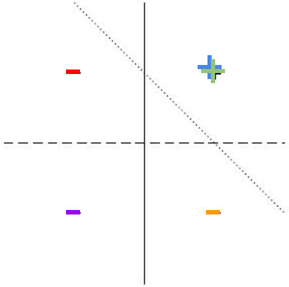
Definition 2. A state (H,A) is zero-loss if all services j satisfy:

1. $A_{ij}\ell(h_j, x_i, y_i) = 0$ for all $i \in \{1, \ldots, n\}$
2. $u(x_i, h_j) \leq 0$ for all i with $y_i = -1$.

All services make accurate classifications on the populations they observe

All negative users receive zero utility and will not use any service

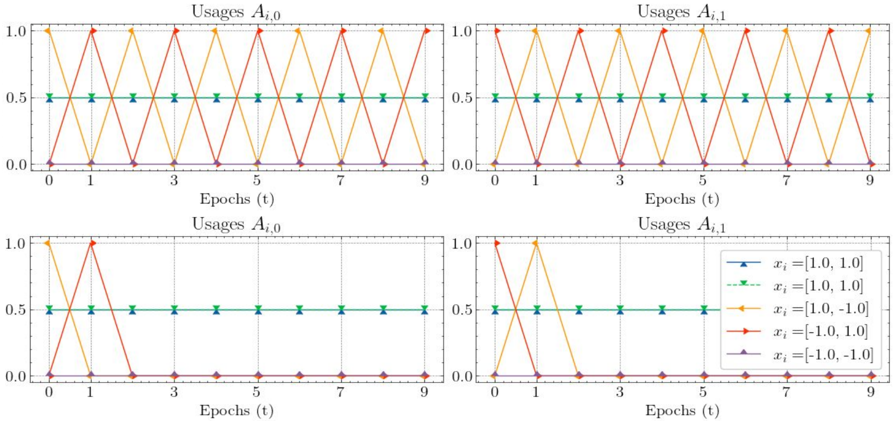
IMPOSSIBILITY RESULT



Filled-in service | Dashed service

Memoryless (p=0):

With memory (p=0.5):



Without memory, negative users (orange and red) switch between services endlessly!

CONVERGENCE RESULT

Definition 2. A state (H,A) is zero-loss if all services j satisfy:

1. $A_{ij}\ell(h_j,x_i,y_i) = 0$ for all $i \in \{1, \ldots, n\}$
2. $u(x_i,h_j) \leq 0$ for all i with $y_i = -1$.

Theorem 5. Given nonzero memory $p > 0$, there is a finite time $t \in \mathbb{N}$ after which for all $\tau > t$, (H^τ, A^τ) is zero-loss.

PROOF STRUCTURE

PROOF STRUCTURE

- **Proposition 3:** once reached, future timesteps will be zero-loss states
    - Proof concept: no mistakes means no service update; users are in an equivalence class and are would’ve incurred loss before if possible

PROOF STRUCTURE

- **Proposition 3:** once reached, future timesteps will be zero-loss states
  - Proof concept: no mistakes means no service update; users are in an equivalence class and are would’ve incurred loss before if possible
- **Lemma 4:** no services observing new users implies a zero-loss state
  - Proof concept: services already do well on users they saw

PROOF STRUCTURE

- **Proposition 3:** once reached, future timesteps will be zero-loss states
  - Proof concept: no mistakes means no service update; users are in an equivalence class and are would’ve incurred loss before if possible
- **Lemma 4:** no services observing new users implies a zero-loss state
  - Proof concept: services already do well on users they saw
- **Theorem 5:** zero-loss state occurs in finite time
  - Proof concept: there are only nm new users that can be introduced to services

BANKNOTE FORGERY EXPERIMENT

Depositors (users) want to deposit banknotes

- Some depositors are forgers!

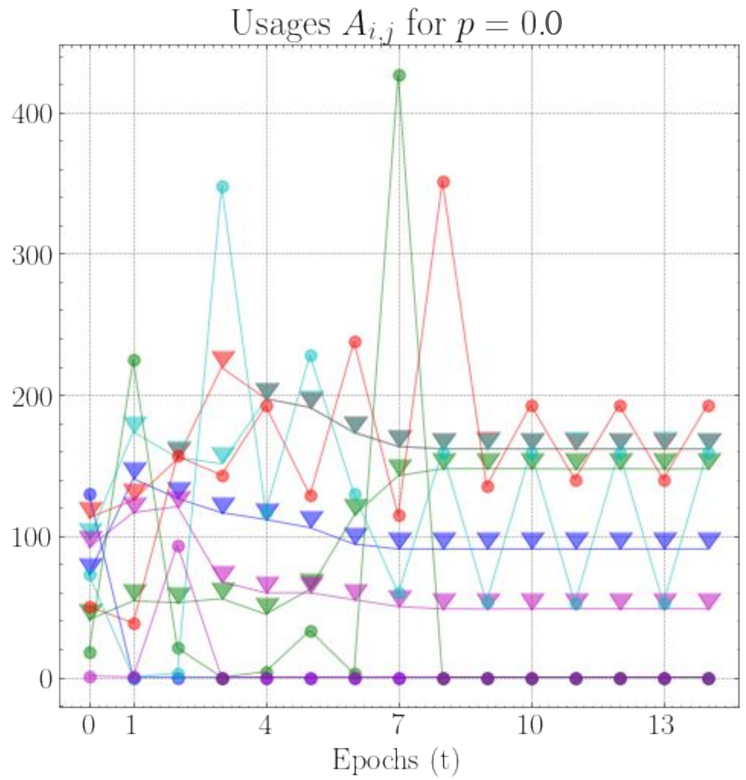
Banks (services) don’t want to accept forgeries

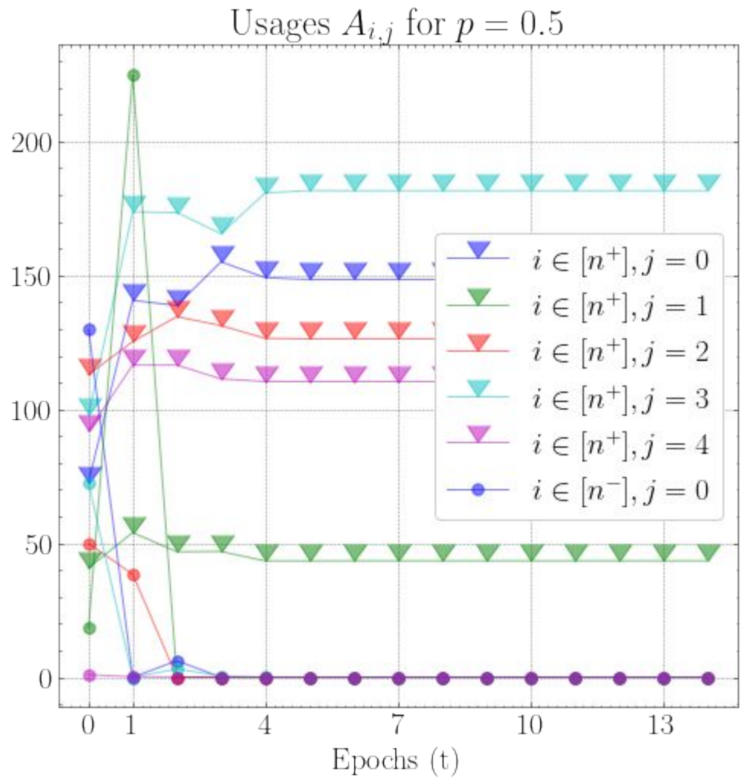
- Want to learn classifiers to vet banknotes

Positives: legal banknotes

Negatives: forgeries

BANKNOTE FORGERY EXPERIMENT





FUTURE DIRECTIONS

Generalizations:

- Sampled population
- Non-realizable user distribution
- Users react to noisy classifiers

New settings:

- Explicit competition between services
- Long-term strategic planning of users