Strategic Usage in a Multi-Learner Setting

Eliot Seo Shekhtman

in Partial Fulfillment of the Requirements for the Degree of Master of Science
Major Chair: Sarah Dean*
Minor Chair: John Whitman

* coauthor of our work under the same name

INTRODUCTION

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

This 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 $h_j : \mathscr{X} \to \{+1, -1\}$, $h \in \boldsymbol{H}$

- 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 $u : \mathscr{X} \times \boldsymbol{H} \to \mathbb{R}$

- *Assume sign of* u *is shared with* h
- Example: projected number of clicks on a listing, how strict their filter is
  - 0-1 utility: $u(x, \theta) = \mathbf{1}\{\theta^{\mathrm{T}}\varphi(x) > 0\}$
  - Linear utility: $u(x, \theta) = \theta^{\mathrm{T}}\varphi(x)$

USER OBJECTIVE

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

- *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 $u : \mathscr{X} \times \boldsymbol{H} \rightarrow \mathbb{R}$

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

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

Users allocate usage to maximize:

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

USER OBJECTIVE

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

Users allocate usage to maximize:

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

- **Lemma 8:** users choose services with maximum utility
  - Proof concept: the only stable equilibrium point when taking gradient descent on the user objective is when all weight is in services granting maximum utility
- **Corollary 9:** services that give no utility to a user will receive no usage

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 $\ell : \boldsymbol{H} \times \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$

- *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 $\ell : \boldsymbol{H} \times \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$

- *Utility has a strict monotonic relationship with* $-y\ell(h, x, y)$
- *There exists a* $v>0$ *such that* $u(x, h) = 0$ *when* $\ell(h, x, y) = v$
  - 0-1 loss: $\ell(\theta, x, y) = \mathbf{1}\{\theta^T\varphi(x) \cdot y > 0\}$
  - Hinge loss: $\ell(\theta, x, y) = \max\{1 - \theta^T\varphi(x) \cdot y, 0\}$

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:

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

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 update to minimize the following formula:

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

- **Lemma 2:** services have zero loss for users they've seen
  - Proof concept: if we saw the point before, realizability gives that the objective wouldn't be minimized if it wasn't classified correctly

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:

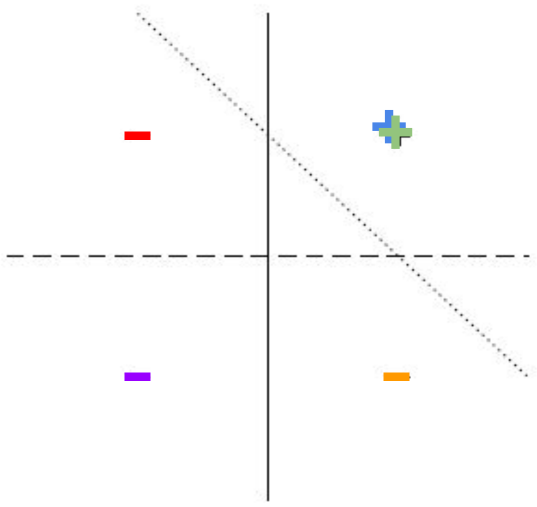
1. $A_{ij}\ell(h_j, x_i, y_i) = 0$ for all $i \in \{1, \ldots, n\}$
2. $u(x_i, h_j) \leq 0$ for all i with $y_i = -1$.

All services make accurate classifications on the populations they observe

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

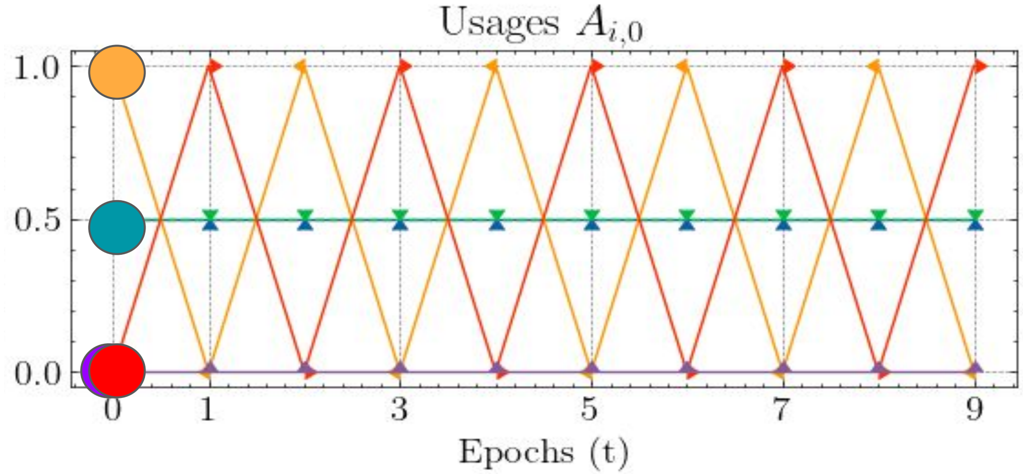
IMPOSSIBILITY RESULT

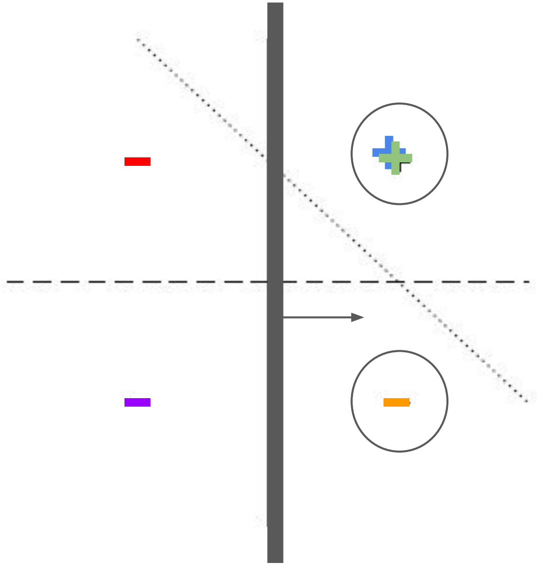
Memoryless (p=0):



IMPOSSIBILITY RESULT

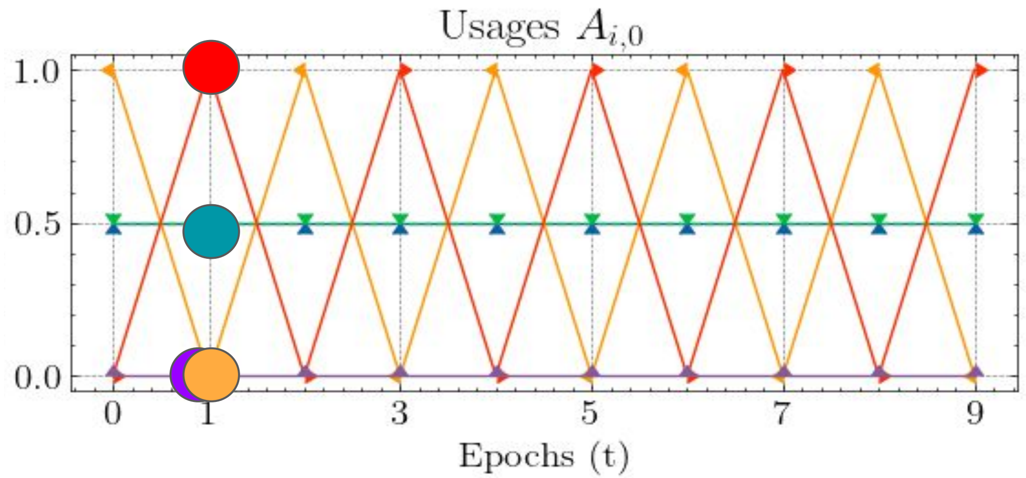
Memoryless (p=0):

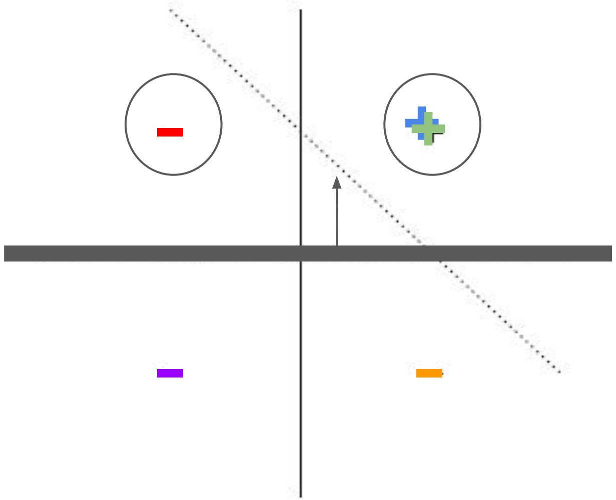




IMPOSSIBILITY RESULT

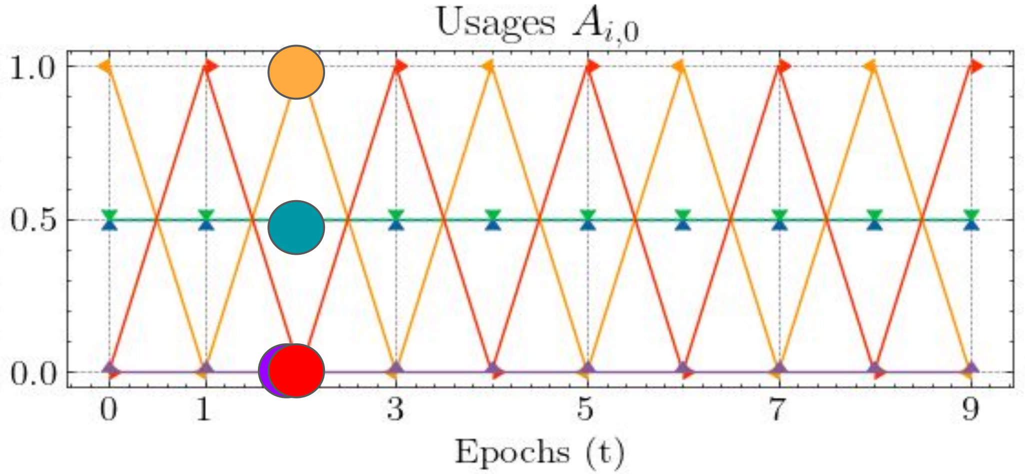
Memoryless (p=0):

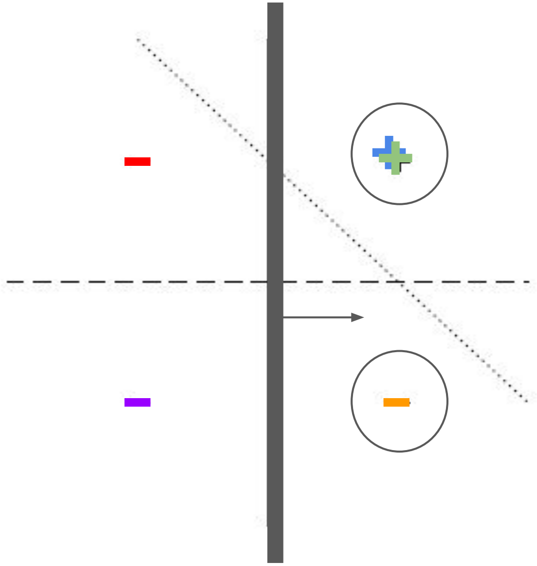




IMPOSSIBILITY RESULT

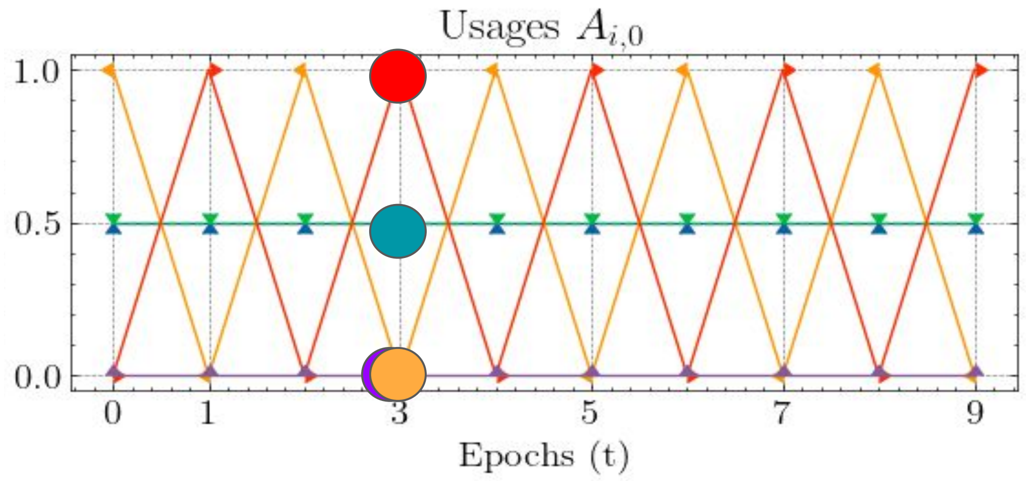
Memoryless (p=0):

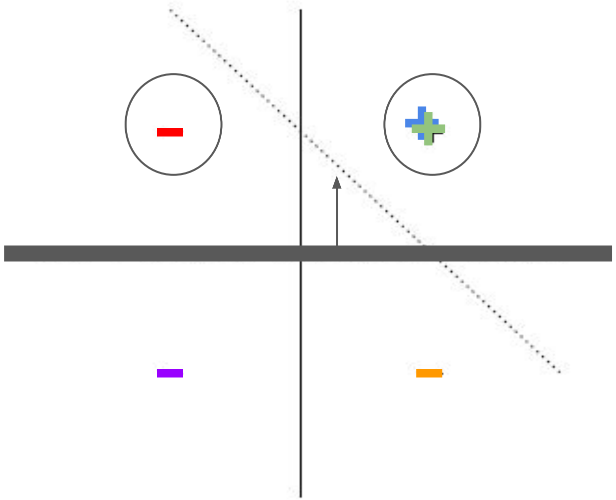




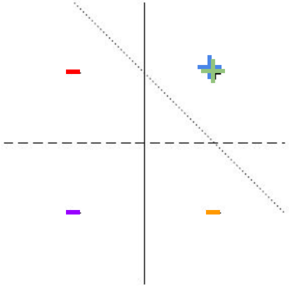
IMPOSSIBILITY RESULT

Memoryless (p=0):





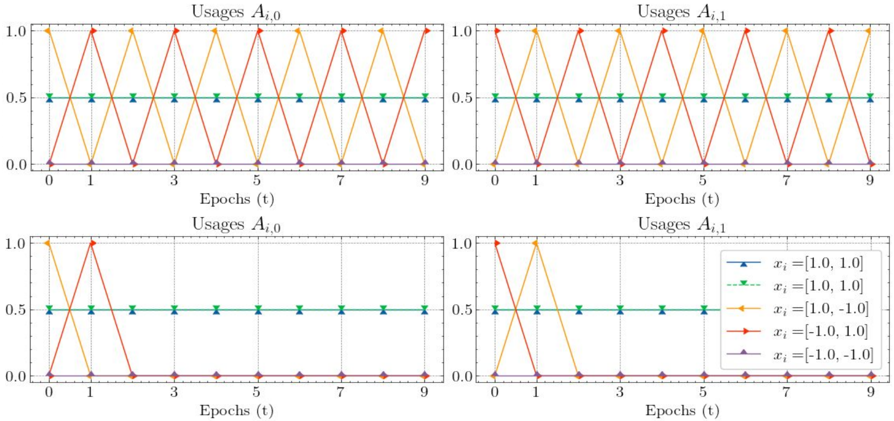
IMPOSSIBILITY RESULT



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 6. 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 5:** 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 5:** no services observing new users implies a zero-loss state
  - Proof concept: services already do well on users they saw
- **Theorem 6:** 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!

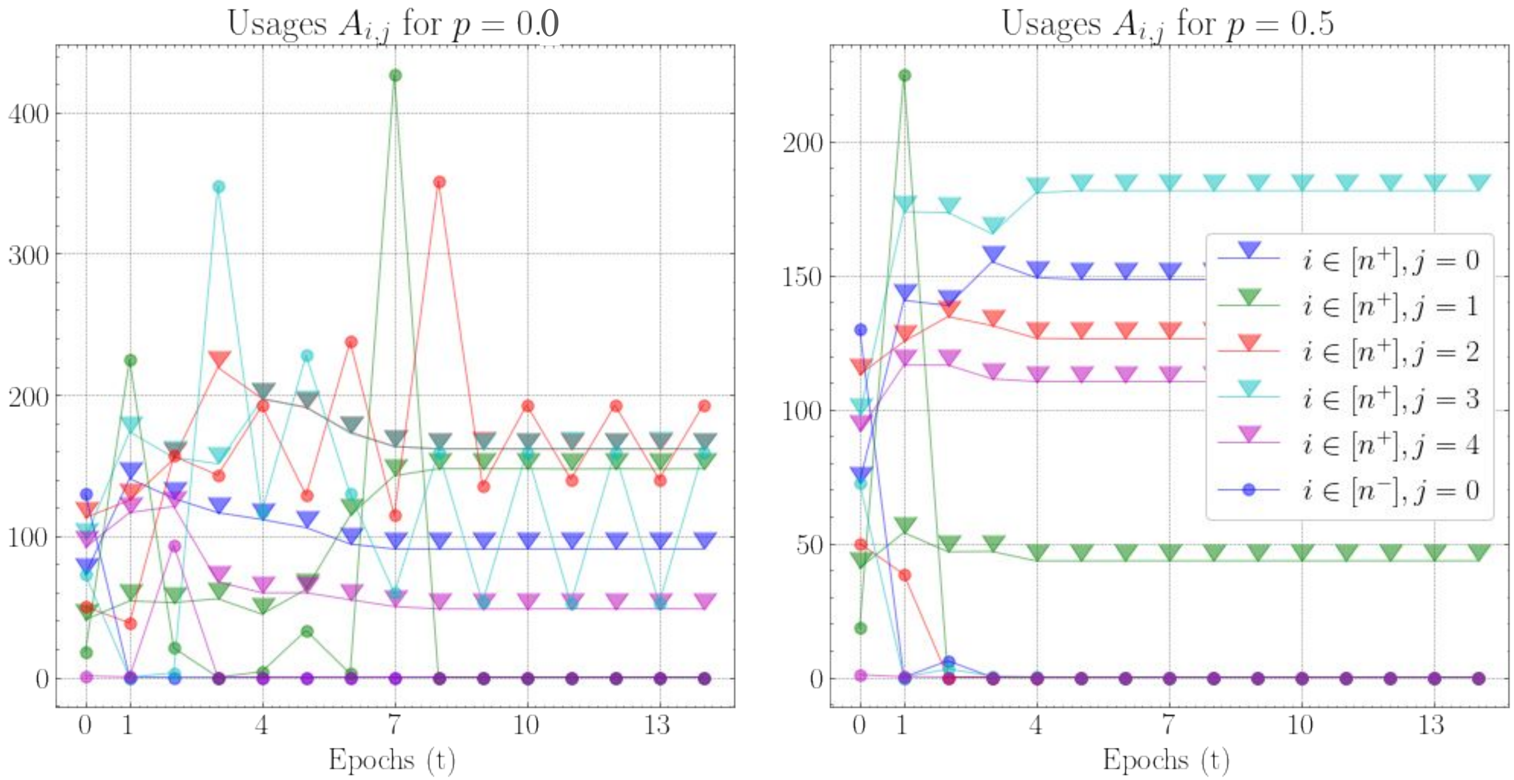
Banks (services) don’t want to accept forgeries

- Want to learn classifiers to vet banknotes

Positives: legal banknotes

Negatives: forgeries

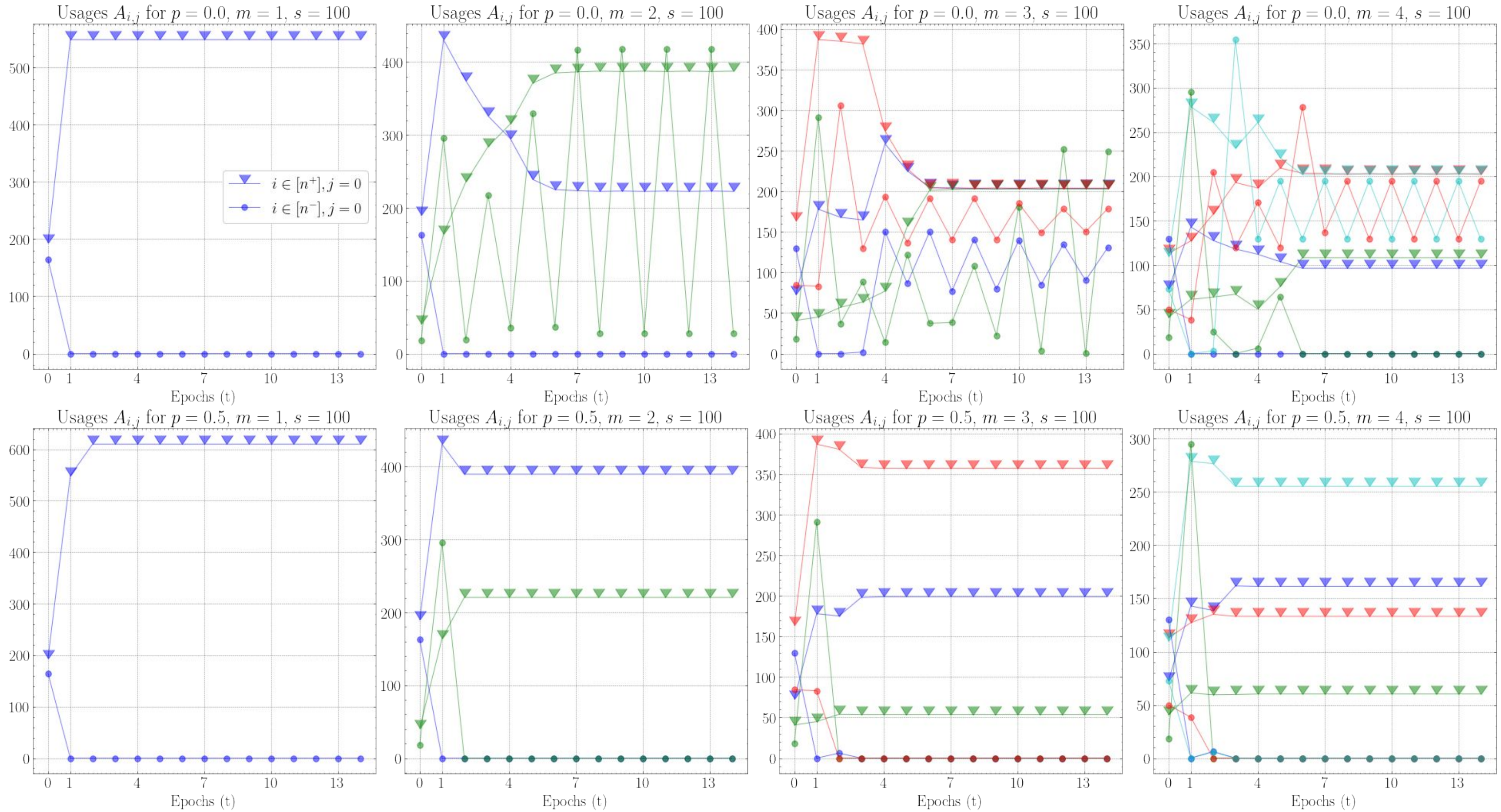
BANKNOTE FORGERY EXPERIMENT



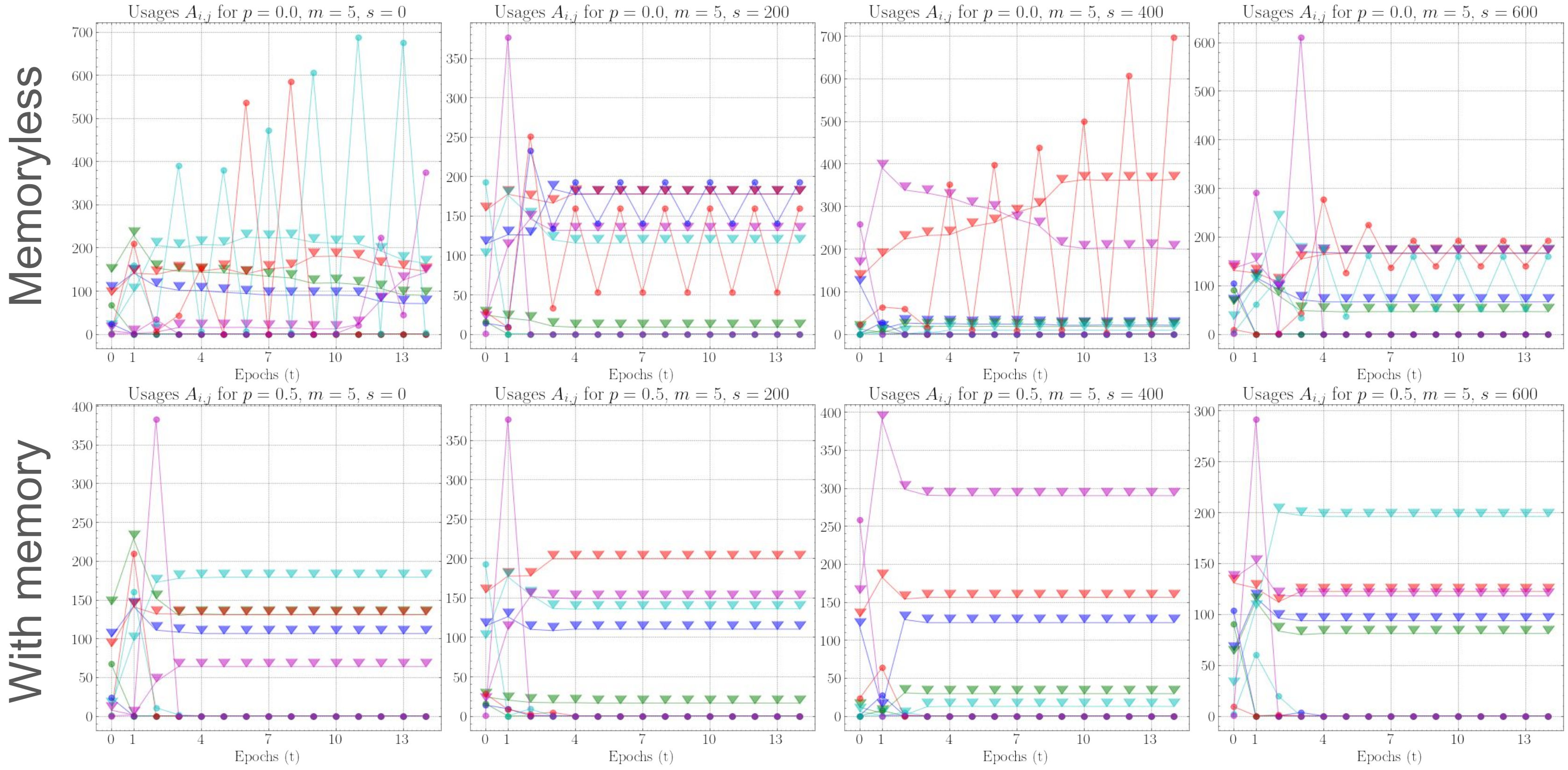
BANKNOTE FORGERY ABLATION (m)

Memoryless

With memory



BANKNOTE FORGERY ABLATION (s)



BANK ACCOUNT FRAUD EXPERIMENT

Clients (users) want to open bank accounts

- Some of these clients and openings are fraudulent

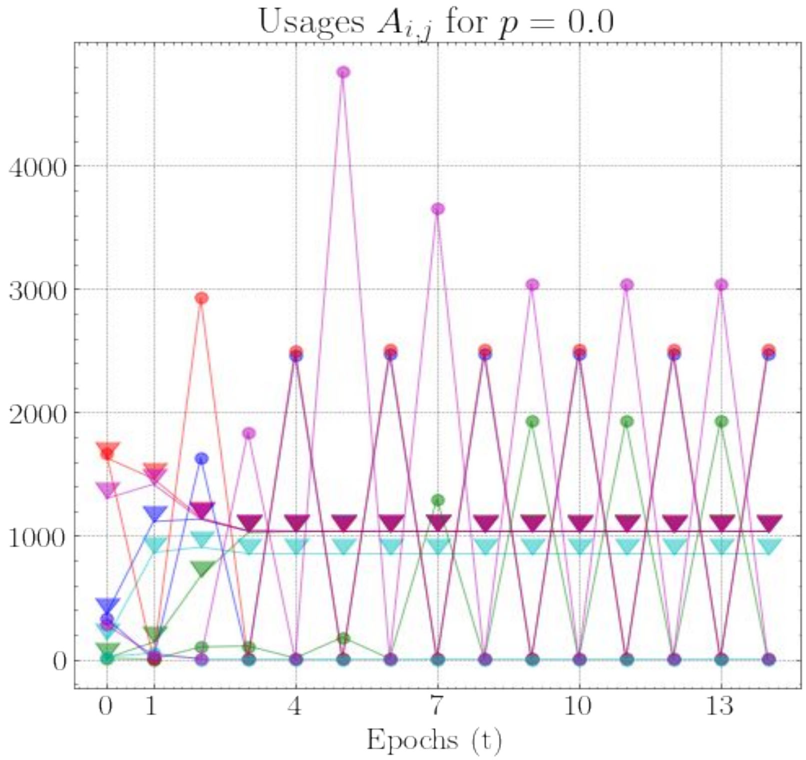
Banks (services) don't want to accept these clients

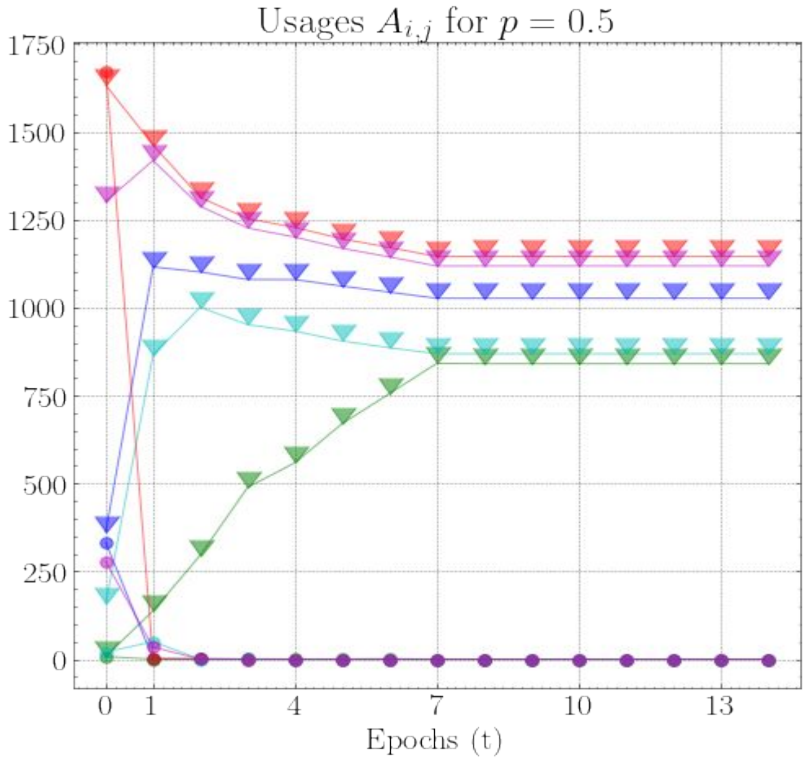
- Want to learn classifiers to vet bank account openings

Positives: legal banknotes

Negatives: fraudulent openings

BANK ACCOUNT FRAUD EXPERIMENT





RELAXATIONS

ROUND ROBIN UPDATES

Usage reallocation and service retraining isn’t so synchronous in real life!

- What if users and services update fully at random?

Need a new time concept: **rounds**

ROUND ROBIN UPDATES

Usage reallocation and service retraining isn't so synchronous in real life!

- What if users and services update fully at random?

Need a new time concept: **rounds**

- Generalize timesteps to contain a full set of user updates and service updates
- **Proposition 7.** Given nonzero memory $p > 0$, there are a finite number of rounds $r \in \mathbb{N}$ after which for all $\rho > r$, (H^ρ, A^ρ) is zero-loss.

UNREALIZABLE DISTRIBUTION

Many interesting extensions involve removing the realizability guarantee

- Users not revealing labels
- Stochastic user labels
- Noisy observations

Inherently violates the existence of a zero-loss equilibrium!

UNREALIZABLE DISTRIBUTION

Many interesting extensions involve removing the realizability guarantee

Definition 3. A state (H^τ, A^τ) is a **stable state** if for all $H > \tau$, $H^t = H^\tau$.

Even so, we cannot make any guarantees for this new convergence definition!

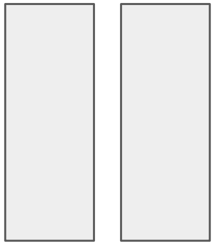
UNREALIZABLE DISTRIBUTION

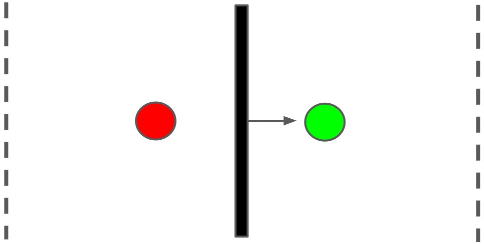
Many interesting extensions involve removing the realizability guarantee

Definition 3. A state (H^τ, A^τ) is a **stable state** if for all $H > \tau$, $H^t = H^\tau$.

Even so, we cannot make any guarantees for this new convergence definition!

MEMORY





UNREALIZABLE DISTRIBUTION

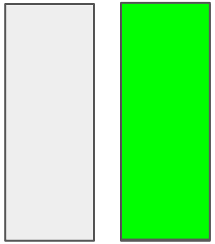
Many interesting extensions involve removing the realizability guarantee

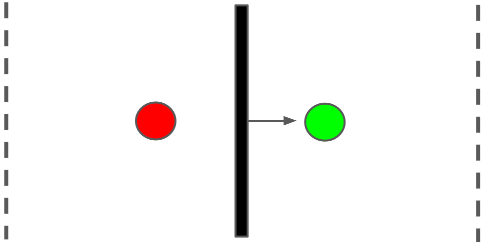
Definition 3. A state (H^τ, A^τ) is a **stable state** if for all $H > \tau$, $H^t = H^\tau$.

Even so, we cannot make any guarantees for this new convergence definition!

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

MEMORY





UNREALIZABLE DISTRIBUTION

Many interesting extensions involve removing the realizability guarantee

Definition 3. A state (H^τ, A^τ) is a **stable state** if for all $H > \tau$, $H^t = H^\tau$.

Even so, we cannot make any guarantees for this new convergence definition!

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

UNREALIZABLE DISTRIBUTION

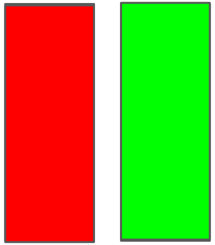
Many interesting extensions involve removing the realizability guarantee

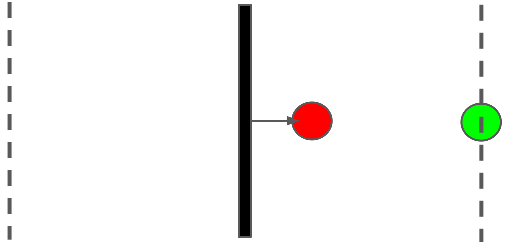
Definition 3. A state (H^τ, A^τ) is a **stable state** if for all $H > \tau$, $H^t = H^\tau$.

Even so, we cannot make any guarantees for this new convergence definition!

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

MEMORY





UNREALIZABLE DISTRIBUTION

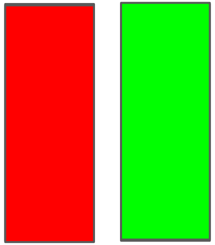
Many interesting extensions involve removing the realizability guarantee

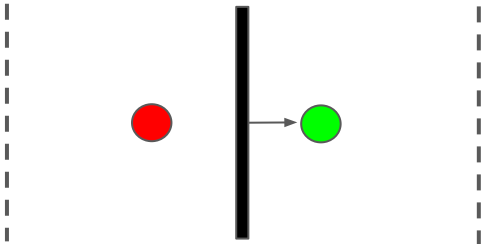
Definition 3. A state (H^τ, A^τ) is a **stable state** if for all $H > \tau$, $H^t = H^\tau$.

Even so, we cannot make any guarantees for this new convergence definition!

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

MEMORY





UNREALIZABLE DISTRIBUTION

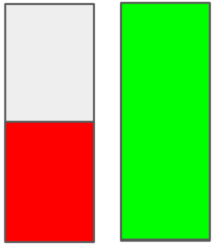
Many interesting extensions involve removing the realizability guarantee

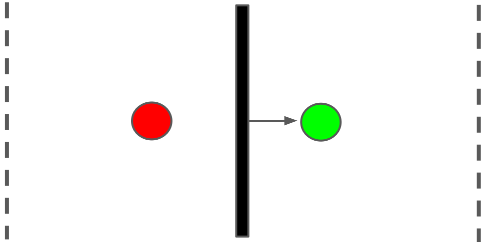
Definition 3. A state (H^τ, A^τ) is a **stable state** if for all $H > \tau$, $H^t = H^\tau$.

Even so, we cannot make any guarantees for this new convergence definition!

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

MEMORY





UNREALIZABLE DISTRIBUTION

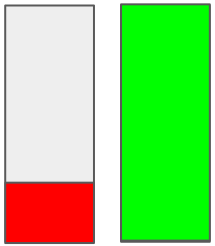
Many interesting extensions involve removing the realizability guarantee

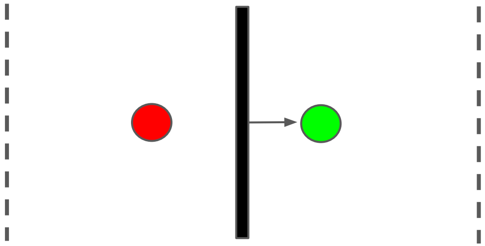
Definition 3. A state (H^τ, A^τ) is a **stable state** if for all $H > \tau$, $H^t = H^\tau$.

Even so, we cannot make any guarantees for this new convergence definition!

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

MEMORY





UNREALIZABLE DISTRIBUTION

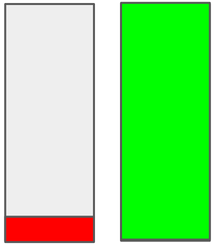
Many interesting extensions involve removing the realizability guarantee

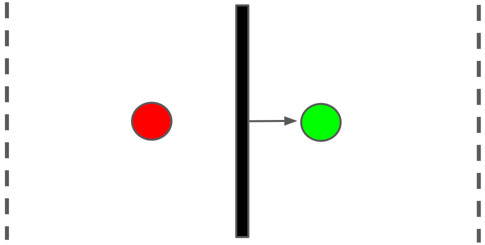
Definition 3. A state (H^τ, A^τ) is a **stable state** if for all $H > \tau$, $H^t = H^\tau$.

Even so, we cannot make any guarantees for this new convergence definition!

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

MEMORY





UNREALIZABLE DISTRIBUTION

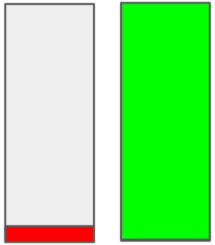
Many interesting extensions involve removing the realizability guarantee

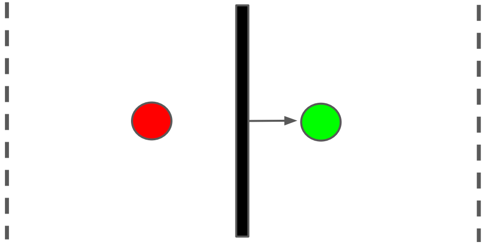
Definition 3. A state (H^τ, A^τ) is a **stable state** if for all $H > \tau$, $H^t = H^\tau$.

Even so, we cannot make any guarantees for this new convergence definition!

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

MEMORY





UNREALIZABLE DISTRIBUTION

Many interesting extensions involve removing the realizability guarantee

Definition 3. A state (H^τ, A^τ) is a **stable state** if for all $H > \tau$, $H^t = H^\tau$.

Even so, we cannot make any guarantees for this new convergence definition!

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

UNREALIZABLE DISTRIBUTION

Many interesting extensions involve removing the realizability guarantee

Definition 3. A state (H^τ, A^τ) is a **stable state** if for all $H > \tau$, $H^t = H^\tau$.

Even so, we cannot make any guarantees for this new convergence definition!

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

MEMORY

UNREALIZABLE DISTRIBUTION

Many interesting extensions involve removing the realizability guarantee

Definition 3. A state (H^τ, A^τ) is a **stable state** if for all $H > \tau$, $H^t = H^\tau$.

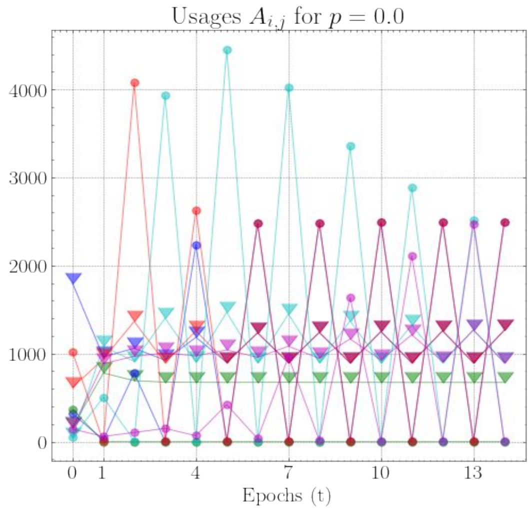
Even so, we cannot make any guarantees for this new convergence definition!

0-1 Memory:

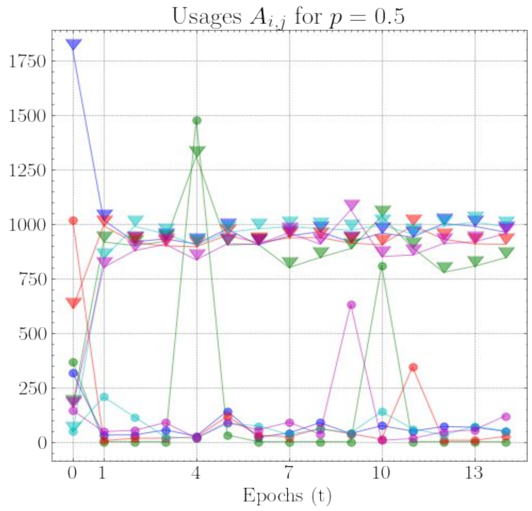
$$M^t = \begin{cases} A^t & \text{if } p = 0 \\ \mathbf{1}\left\{\max(A^t, M^{t-1})\right\} & \text{if } p > 0 \end{cases}$$

UNREALIZABLE DISTRIBUTION

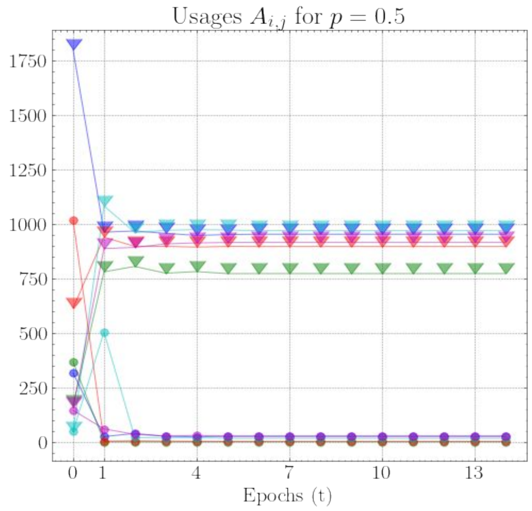
Memoryless



Weighted Memory



0-1 Memory



SAMPLED USERS

Users might enter or leave a system over time, leading to a non-static dataset

This means there’s no longer a finite number of service-user introductions!

A mistake bound analysis is required!

SAMPLED USERS

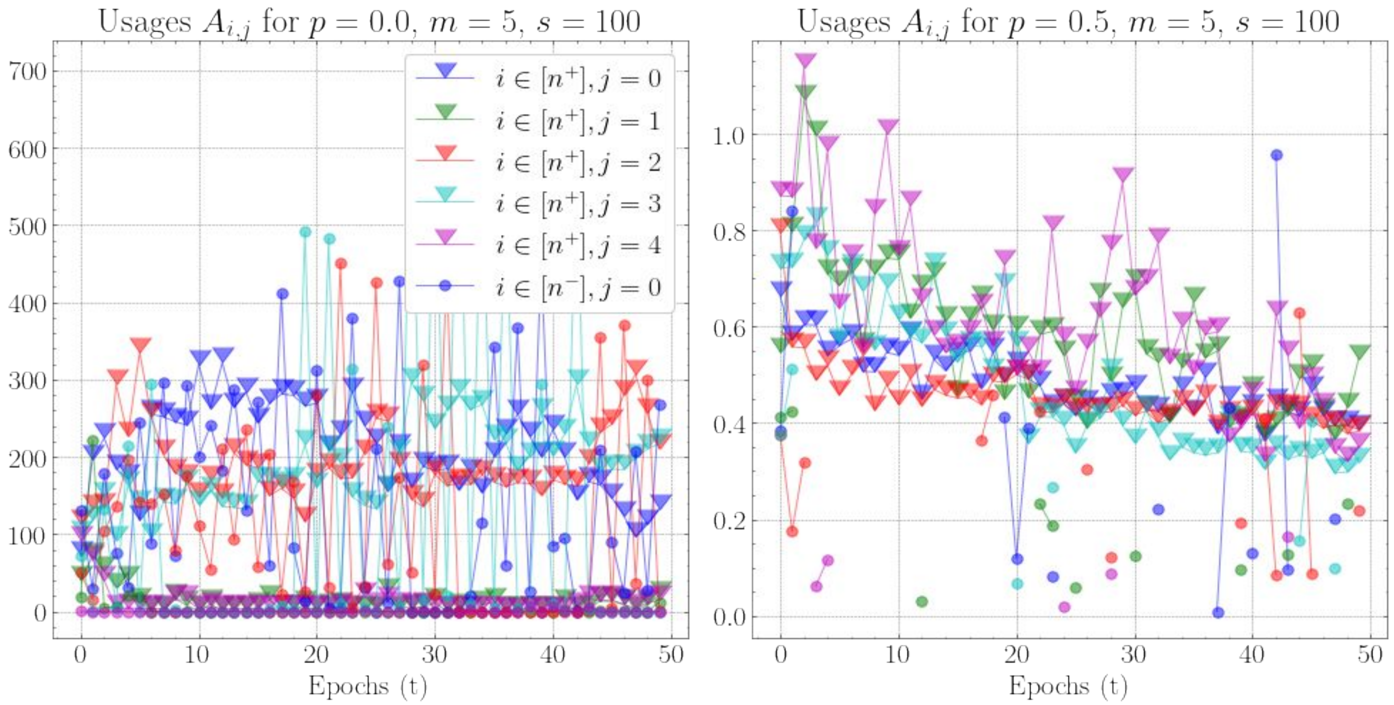
Users might enter or leave a system over time, leading to a non-static dataset

This means there's no longer a finite number of service-user introductions!

A mistake bound analysis is required!

- Even so, could make an infinite number of mistakes with no further constraints on the model class

SAMPLED USERS



FUTURE DIRECTIONS

- Further theoretical exploration of the relaxations!

Extensions:

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

THANK YOU!

ADVISORS • FRIENDS • COWORKERS • FAMILY