Grinding the Space: Learning to Classify Against Strategic Agents

Chara Podimata, Harvard

Joint work with
Yiling Chen (Harvard), and Yang Liu (UC Santa Cruz)
ML Against Strategic Agents

**Strength of agent according to what they report**

- Stochastic (report ~ Distribution)
- Strategic (report ~ utility maximizer)
- Adversarial (any report)

Learner’s Goal: Either induce truthfulness (learn as if ‘clean’ dataset) or Learn by exploiting the structure of the agents’ behavior

We address the problem of online learning in repeated classification settings against strategic agents.
Model

Protocol
1. At timestep $t \in [T]$, learner draws action $\alpha_t \in \mathcal{A} \subseteq [-1,1]^{d+1}$.
2. An agent observes $\alpha_t$, draws $(x_t, y_t)$, with $x_t \sim \mathcal{X} \subseteq ([0,1]^d, 1)$, $y_t = \{-1, 1\}$.
3. Agent reports $z_t(\alpha_t; x_t)$ (potentially different from $x_t$).
4. Learner observes true label of $z_t(\alpha_t; x_t)$ ($\hat{y}_t$); incurs classification loss:
   \[ \ell(\alpha_t, z_t(\alpha_t; x_t)) = 1\{\hat{y}_t \cdot \langle \alpha_t, z_t(\alpha_t; x_t) \rangle = -1 \}. \]

Agents’ Behavior
- General form of utility functions: $u_t(\alpha, z_t(\alpha_t; x_t)) = v_t(\alpha, z_t(\alpha_t; x_t)) - c_t(\alpha, z_t(\alpha_t; x_t))$
- Value $\in [0,1]$, e.g., $\delta \cdot 1\{(\alpha, z_t(\alpha; x_t)) \geq 0\} \cdot 1\{\hat{y}_t \neq y_t\}$
- Cost $\in [0,1]$, e.g., $(x_t - z_t(\alpha; x_t))^2$
- Myopic & rational, i.e., report:
  \[ r_t(\alpha_t; x_t) = \operatorname{arg\ max}_{z_t \in \mathcal{X}; x_t} u_t(\alpha, z_t(\alpha_t; x_t)) \]
Goal

Minimize Stackelberg Regret:

\[ R(T) = \sum_{t=1}^{T} \ell(\alpha_t, r_t(\alpha_t)) - \min_{\alpha \in \mathcal{A}} \sum_{t=1}^{T} \ell(\alpha, r_t(\alpha)) \]


[Dong, Roth, Schutzman, Waggoner, Wu ’18] found appropriate conditions for the best-response of the agents, such that the problem becomes online learning with convex losses.

More broadly...

Strategic Classification: [Meir, Almagor, Michaely, Rosenschein ’11], [Meir, Procaccia, Rosenschein ’12], [Hardt, Meggido, Papadimitriou, Wooters ’16]...

Strategic Linear Regression: [Perote, Perote-Pena ‘04], [Dekel, Fischer, Procaccia ‘10], [Cummings, Ioannidis, Ligett ‘15], [Cai, Daskalakis, Papadimitriou ‘15], [Chen, P., Procaccia, Shah ‘18]...
Minimize Stackelberg Regret:

\[ R(T) = \sum_{t=1}^{T} \ell(\alpha_t, r_t(\alpha_t)) - \min_{\alpha \in \mathcal{A}} \sum_{t=1}^{T} \ell(\alpha, r_t(\alpha)) \]

1. External Regret

\[ R(T) = \sum_{t=1}^{T} \ell(\alpha_t, r_t(\alpha_t)) - \min_{\alpha \in \mathcal{A}} \sum_{t=1}^{T} \ell(\alpha, r_t(\alpha)) \]

is worst-case incompatible with Stackelberg Regret

2. \( \ell(\alpha, r_t(\alpha)) \) is not even Lipschitz (wrt to \( \alpha \)) in general
What can we infer from the agents’ behavior?
What can we infer from the agents’ behavior?

\[ \alpha \]

\[ x_t \]

\[ r_t(\alpha) \]

\[ \delta \]

\[ 2\delta \]

\[ \alpha \]

\[ x_t \]

\[ r_t(\alpha) \]

\[ \delta \]

\[ 2\delta \]
What can we infer from the agents’ behavior?

Actions outside of $2\delta$-ball are safe to update! \( \rightarrow \) conservative estimate, no bias added
What can we infer from the agents’ behavior?

- Actions outside of $2\delta$-ball are safe to update! \(\rightarrow\) conservative estimate, no bias added
- Actions $\alpha$ inside the region: label for $r_t(\alpha) = +1$
- Actions $\alpha$ inside the region: label for $r_t(\alpha) = -1$
Translation into the learner’s action space

\[ P_t^u \]

\[ P_t^l \]

\[ x_t \]

\[ r_t(\alpha) \]

\[ \alpha \]

\[ \gamma \]

\[ 2\delta \]

\[ \delta \]
Translation into the learner’s action space

\[ r_t(\alpha) \]

\[ x_t \]

\[ \delta \]

\[ 2\delta \]
Translation into the learner's action space

\[ \alpha \]

\[ x_t \]

\[ r_t(\alpha) \]

\[ 2\delta \]

\[ \delta \]

\[ 1 \]

\[ -1 \]

\[ 1 \]

\[ -1 \]
Translation into the learner’s action space

\[
\alpha = x_t + \delta, \quad r_t(\alpha) = x_t + 2\delta
\]
The Grinding Algorithm

For all $t \in [T]$:

1. Choose polytope $p_t \sim \pi_t(p) = (1 - \gamma)q_t(p) + \frac{\gamma \lambda(p)}{\lambda(A)}$

2. Commit to $\alpha_t \sim Unif(p_t)$

3. Observe $(r_t(\alpha_t), y_t)$ & grind $A$ according to $p^u_t, p^l_t, p^m_t$.

4. Update all polytopes with their IPS. (note: upper, and lower $\rightarrow$ full info!)

5. $\forall p$: $q_{t+1}(p) \propto \lambda(p) \exp(-\eta \sum_{i=1}^{t} \ell(p, r_t(p)))$. 
Grinding Algorithm adaptively discretizes the continuous action space & achieves Stackelberg regret:

\[
O\left(\sqrt{\max_{t \in [T]} \left\{ 8\log \left( \frac{4\lambda(\mathcal{A})}{\lambda(p')} \right) + \lambda(\mathcal{P}^m_{t,G_T}) \right\} \log \left( \frac{4\lambda(\mathcal{A})}{\lambda(p')} \right) T} \right)
\]

\(\lambda(S)\): Lebesgue measure of \(S\)

\(p'\): polytope with the smallest Lebesgue measure
Simulations – 60% probability of being spammer

Discrete, Well-Defined Action Set

Continuous Action Set

\[ p = 0.6, \delta = 1.00 \]

In both cases, Grinding significantly outperforms EXP3, despite not having access to an omnipotent oracle.
Future Directions + Open Questions

A lot of interesting technical questions from the paper. Some examples:

- Can we learn non-linear classifiers?
- What if we have to concurrently find the appropriate $\delta$?
- Formal proof that approximation oracles will not destroy our regret bound.
- To what extend can the Smoothing techniques [Krishnamurthy, Langford, Slivkins, Zhang ‘19] help the Grinding Algo?

More broadly...

- Learning against agents with different behavioral models (e.g., not fully rational, approximately maximizing utility functions etc.)
Thank You!