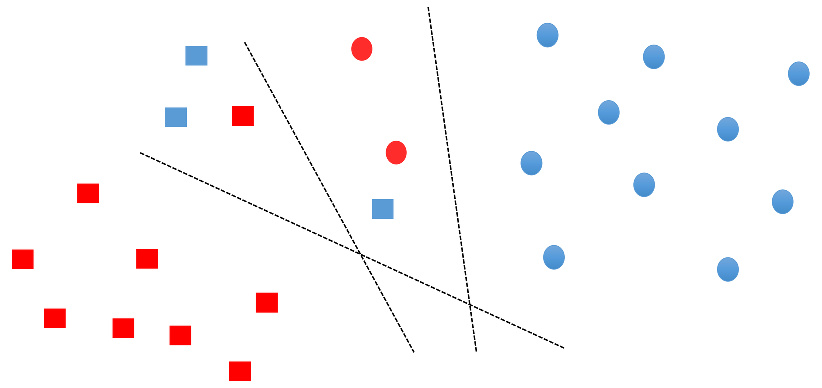
A Bandit Framework for Strategic Regression

Yang Liu and Yiling Chen
School of Engineering and Applied Science & CRCS, Harvard

yangl@seas.harvard.edu
www.umich.edu/~youngliu



Introduction

Training data for a series of regression problems are collected from strategic sources, such as via crowdsourcing or survey.

Such data suffers from quality issues:

- Intrinsic noise: different worker expertise
- Strategic noise: due to lack of monitoring, incentives etc.

How to control the quality for learning?

- Lack of prior knowledge of workers
- Lack of ground truth for quality verification
- Lack of monitoring: e.g., effort is not observable
- Lack of incentives

Objectives

Elicit high quality data for training regression model with performance guarantee.

- How to incentivize effort from workers?
- Any other incentives besides one-step payment?
- A robust mechanism or algorithm.
- Easy to implement.

Acknowledge

We acknowledge the support of NSF grant CCF-1301976.

References

[1] Peter Auer, Nicolo Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine learning*, 47(2-3):235–256, 2002.
[2] Yang Cai, Constantinos Daskalakis, and Christos H Papadimitriou. Optimum statistical esti-mation with strategic data sources. *COLT 2015.*
[3] Rachel Cummings, Stratis Ioannidis, and Katrina Ligett. Truthful linear regression. In *Proceedings of The 28th Conference on Learning Theory, COLT 2015, Paris, France, July 3-6, 2015*, pages 448–483, 2015.
[4] Nolan Miller, Paul Resnick, and Richard Zeckhauser. Eliciting informative feedback: The peer-prediction method. *Management Science*, 51(9):1359–1373.
[5] Stratis Ioannidis and Patrick Loiseau. Linear regression as a non-cooperative game. In *Web and Internet Economics*, pages 277–290. Springer, 2013.

NIPS'16

Strategic regression and our goal

The learner has a regression problem in mind.

- Assign data to each worker to label
- Targeting a good regression with training data collected from workers

$$f : \mathbb{R}^d \rightarrow \mathbb{R} \text{ that } \mathbf{y(x)} = f(x) + z$$

Workers are effort sensitive, and can be incentivized via monetary payment

- Higher effort leads to smaller variance in data

$$\tilde{y}_i(x,e) = f(x) + z + z_i(e), \ \text{Var}(z_i(e)) = \sigma_\mathbf{i}(\mathbf{e})$$

- Utility function: Payment – Effort over tasks.

$$U_i(e) = \sum_k (\mathbb{E}[\text{Payment}(e_k)] - e_k)$$

$$\mathbf{e}^* \in \text{argmin}_{\{e(x)\}_{x\in X}} \texttt{ERROR}(\tilde{f}(\{x, \tilde{y}(x, e(x))\}_{x\in X})) + \lambda \cdot \texttt{PAYMENT}(\{e(x)\}_{x\in X})$$

One-step payment function can be designed to pay each contributed data point to elicit effort [2,3]

Our approach: long-term incentive… instead of immediate payment

Future job opportunity

- In a stable market, workers care about *future job opportunities*.
- Or in other words, they care about their *reputation*.

Future selection ⇔ Bandit (Multi-Armed Bandit)

- Each worker will be taken as an "arm".
- Maintain online "score", which the future selection will be based on.
- But we do not directly observe workers' performance => scoring rule

SR-UCB: a scoring rule aided UCB

Step 1. For each worker, train a reference estimator using data from other workers.

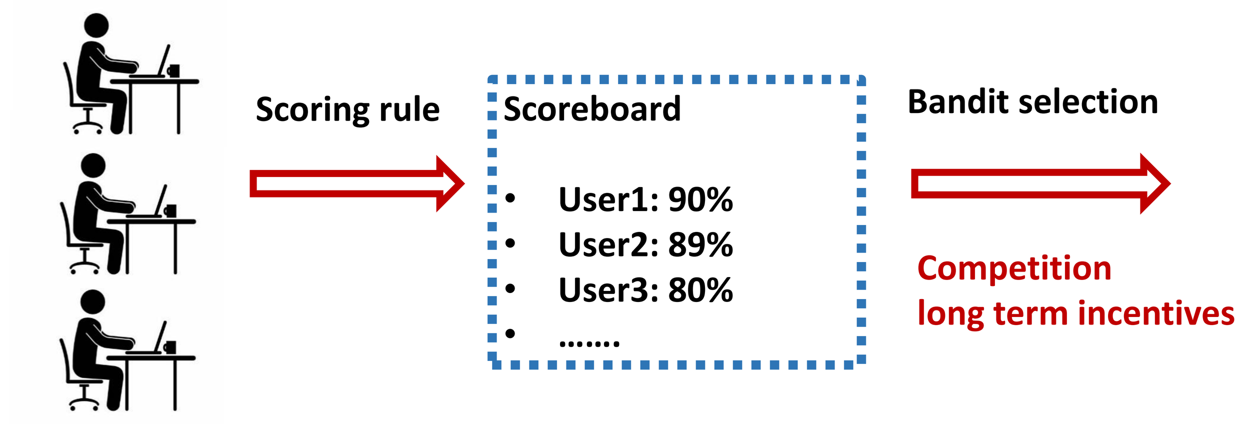
$$\tilde{f}_{-i,t}(\{x_j(n) : 1 \leq n \leq t-1, \ j \in d(n), j \neq i\})$$

Step 2. Then compute the following index for worker

$$\frac{1}{n_i(t)} \sum_{n=1}^{t} 1(i \in d(n)) \left[\mathbf{a} - \mathbf{b}\left(\tilde{\mathbf{f}}_{-\mathbf{i},\mathbf{t}}(\mathbf{x_i(n)}) - \tilde{\mathbf{y}}_\mathbf{i}(\mathbf{n}, \mathbf{e_i(n)})\right)^2\right] + \mathbf{c}\sqrt{\frac{\log \mathbf{t}}{\mathbf{n_i(t)}}}$$

Step 3. Select the following workers to assign data

$$d(t) := \{j : I_j(t) \geq \max_i I_i(t) - \tau(t)\}$$



For workers with homogeneous expertise:

- Pay each worker

$$p_i = e^* + \gamma, \ \gamma = \Omega(\sqrt{\log T/T})$$

- Form a competition: **Exerting efforts guarantees linear number of selections.**

Lemma 1 *If every worker exerts effort level* $e_i(t) = e^*, \forall t$, *there exists a constant* $\delta_U > 0$ *such that for any* i, j *that* $i \neq j$ *we have probability at least* $1 - O(\frac{1}{t^2})$, $n_i(t) \leq (1+\delta_U) n_j(t)$.

- Slack of: selection bounded sub-linearly
- **Theorem:** Workers exerting e^* is an approximate BNE.

With heterogeneous workers:

- Targeting the *best two: sufficient and necessary*
- Form competition & data is coming from the most competitive workers

Extension, computation & privacy

Ridge regression

$$\tilde{\theta} = \text{argmin}_{\hat{\theta}\in\mathbb{R}^d} \sum_{x\in X} (y(x) - \hat{\theta}^T x)^2 + \rho||\hat{\theta}||_2^2$$

- Biased reference model
- To account for bias: a larger confidence term

Non linear regression

- Inspired by consistency of M-estimator.
- Works for parametric families such that

$$|\tilde{f}_{-i,t}(x) - f(x)| \leq L_N ||\tilde{\theta}_i(t) - \theta||_2$$

Computational issues

- Online model update

$$\tilde{\theta}_{-i}^{\text{online}}(t+1) := \tilde{\theta}_{-i}^{\text{online}}(t) - \eta_t \cdot \nabla_{\tilde{\theta}_{-i}^{\text{online}}(t)} \left[(\theta^T x_{-i}(t) - \tilde{y}_{-i}(t))^2 + \rho||\theta||_2^2\right]$$

- Online score update

$$S_i^{\text{online}}(t) := \frac{1}{n_i(t)} \sum_{n=1}^{t-1} 1(i \in d(n)) \left[a - b\left((\tilde{\theta}_{-\mathbf{i}}^{\text{online}}(\mathbf{n}))^T x_i(n) - \tilde{y}_i(n, e_i(n))\right)^2\right]$$

Privacy preserving

- Preserving privacy in a sequential index system
- Partial sum idea: separate indexes into partial sums; only add noise to each partial sum.

$$S_i^{\text{online}}(t) = \frac{1}{n_i(t)}\left(\sum_{n=q(t)+1}^{t} dS(n) + \sum_{n=q(q(t))+1}^{q(t)} dS(n) + ... + \sum_{n=0}^{0} dS(n)\right)$$

What we achieve

- Show a bandit framework can help provide long-term incentive for such regression problems.
- A long-term, quantitative reputation system.
- Robust to different regression models & can be maintained efficiently.
- Preserve privacy in workers' data.