Bandit in Crowdsourcing

Yang Liu

Harvard University

ACK: This tutorial received a lot of information from CJ

Disclaimer

- This is *not* intended to be
  - either a technical lecture
  - or a systematic review of results

- What this tutorial is trying to provide?
  - several pointers to interesting challenges

Crowdsourcing

“Crowdsourcing, a modern business term coined in 2005,[1] is defined by Merriam-Webster as the process of obtaining needed services, ideas, or content by soliciting contributions from a large group of people, especially an online community, rather than from employees or suppliers”

[Wiki]

Bandit(Multi-armed Bandit, MAB)

- MAB is a decision making & learning framework,
  - Make a sequence of decision on selections, when facing multiple options with unknown statistics.
  - **Q**: which one to select next
  - **Goal**: Maximize total payoff or minimize regret

Formulation

- N options, with unknown reward
  - Observe one sample if pulled once

$$X_1, ..., X_N \text{ with mean } \mu_1 > ... > \mu_N$$

  - IID sequence (existing results also cover MC samples).
  - Select: $a(1), ..., a(t), ...$
- Weak regret:

$$R(T) = T \cdot \mu_1 - \sum_{t=1}^{T} \mathbb{E}[X_{a(t)}]$$

- Goal: $R(T) = o(T)$

Upper**C**onfidence**B**ound1 [Auer et al 2002]

- Initialization: for $t \leq N$, play arm/choice $t, t = t + 1$
- When $t > N$:
  - for each choice k, calculate its sample mean:

$$\bar{X}_k(t) = \frac{X_k(1) + ... + X_k(n_k(t))}{n_k(t)}$$

  - its index

$$I_k(t) = \bar{X}_k(t) + \sqrt{\frac{L \log t}{n_k(t)}}, \forall k.$$

  - play the arm with the highest index; $t = t + 1$.

Why it works

- Regret bound: $R(T) \leq O(\log T)$

$$\text{w.h.p.}, \bar{X}_1(t) + \sqrt{\frac{L \log t}{n_1(t)}} \geq \mu_1$$

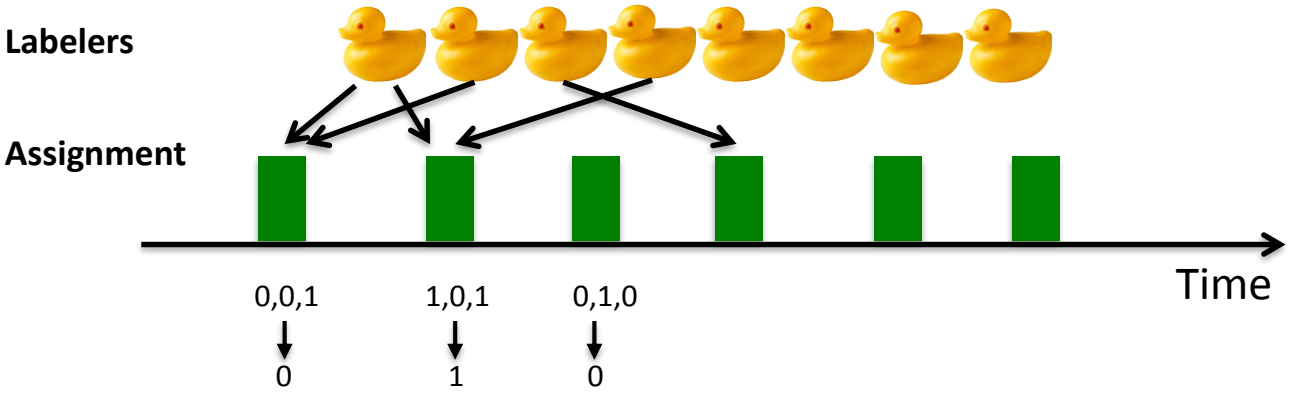
$$\text{w.h.p.}, \bar{X}_k(t) + \sqrt{\frac{L \log t}{n_k(t)}} \leq \mu_k + 2\sqrt{\frac{L \log t}{n_k(t)}}$$

$$\text{When } 2\sqrt{\frac{L \log t}{n_k(t)}} < \mu_1 - \mu_k \Rightarrow \text{no regret.}$$

- **Applications in crowdsourcing**

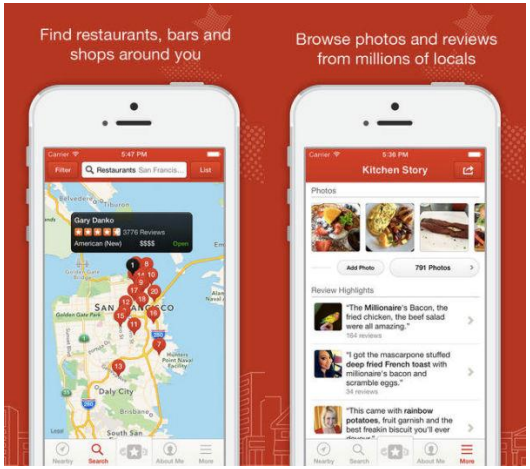
Application I: decision makings in crowdsourcing

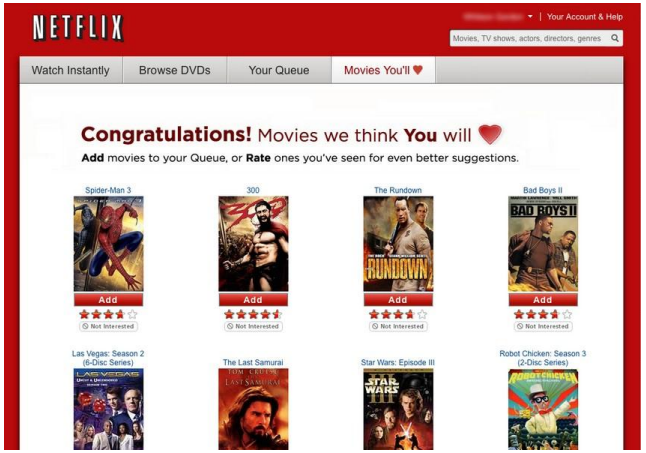
- Who to send our request to? [Ho and Vaughan 12, Tran-Thanh et al.14, Abraham et al. 13, Liu and Liu 15]
  - Unknown performance



- How much should we offer (pricing, contracts)? [Singla and Krause 13, Chawla et al. 15, Ho et al. 16]
  - Unknown incentives

- Which option is better? [Li et al. 10, Massoulié et al. 15, Bresler et al.15]
  - Unknown preference





Application II: Long term incentives

- Inducing high quality contribution from crowdsourcing
  - One-shot payment (scoring rule, e.g.)

- User-generated content,crowdsourced labels,…etc
  - Unknown effort

- What about future job opportunities? [Ghosh and Hummel 13, Liu and Chen 16]
  - You will be selected in future,if you do well.
  - Reputation => Index, future job => bandit selection

In applying bandit … challenges (outline)

- Large space (*metric bandit*)
- Budget constraints (*Knaps. Bandit*)
- Incentivize exploration (*Strategic exploration*)
- Partial information (dueling bandit)
- Bandit w/o ground-truth
  - *Decision making*
- Long term incentive (*endogenous bandit*)
  - *Incentives & Reputations*

- **Decision making**
  - *Full information*
  - *Strategic information*
  - *Partial information*

Full information

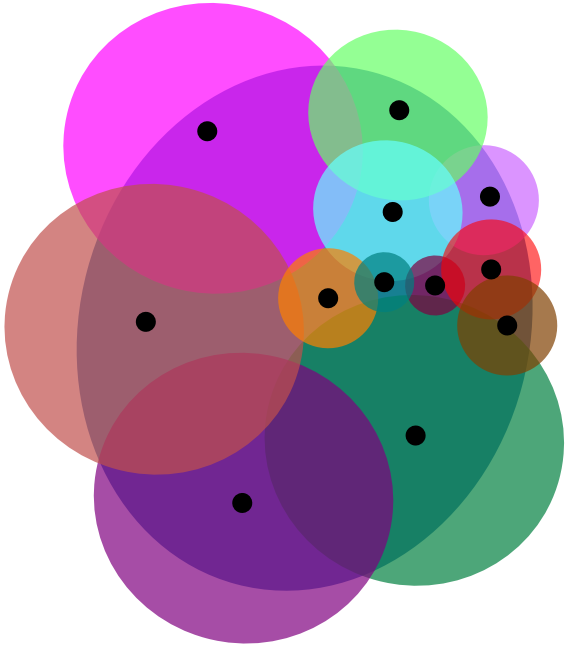
- Setting is similar to classical bandit setting
  - E.g., for each offered price, we observe workers' action in accepting or not.

- Additional challenge in crowdsourcing
  - Larger exploration space (price, contextual information)
  - Budget constraints (budget)

Large exploration space: metric bandit

- Intuitions:
  - The payoffs of "nearby" arms could be similar
  - Each pull learns the payoffs of nearby arms
  - Need only focus on more promising regions of arms

- Payoff structures:
  - Lipschitz condition [Kleinberg et al. 08, Bubeck et al. 08]
  - Tree structures [Slivkins 11, Munos 11]
  - Uncertain but learnable structure [Ho et al. 16]



Budget constraints: bandit with knapsacks

- Example: Knapsack bandit [Tran-Thanh et al. 12, Badanidiyuru et al. 13]
  - A stochastic version of knapsack problem.
  - Each arm pull consumes resources.
  - Exploration-exploitation under budget constraints

$$\sum_{t=1}^{T} c_i(t) \leq B_i, \forall i.$$

- Intuition:
  - Calculate the UCB value of the arm payoff
  - Estimate the "unit cost" of the arms via the dual problem
  - Select the arm with the maximum "bang-per-buck" index

Incentivizing explorations: BIC-MAB

- Decide to sample option k [Mansour et al. 15, Mansour et al. 16]
- E.g., Google wants to know ratings of restaurant k
  - Want to ask a user to "sample"
  - User can choose a different option.

- Randomize exploration with exploitation, to take advantages of workers' limited belief.
  - As a user: not sure about whether I'm being explored, or this is indeed the best option.

Partial information

- More likely for a crowdsourcing setting.
- E.g.1, you don't observe the sample realization $X_i(\omega)$
  - but you observe

$$1(X_i(\omega_i) > X_j(\omega_j))$$

  - Common in
    - recommendation elicitation (which movie to recommend)
    - ranking elicitation (which one to vote)

- E.g.2, learner wants to explore a diverse crowd of workers
  - Assign tasks, and get back with the labels,
  - but how well do they perform? (or how to update index)

Pairwise comparison: Dueling bandit

- Choose two options each step
- Goal: target the best option via comparisons [Yue et al. 12, Zoghi et al 15, Zoghi et al. 15]
  - Condorcet winner

$$p_{i,j} := \Pr(X_i > X_j) > 1/2, \forall j \neq i$$

  - Copeland winner

$$\text{argmax}_i \sum_{j \neq i} 1(p_{i,j} > 1/2)$$

- E.g., Copeland Confidence Bound [Zoghi et al. 15]
  - Confidence bounds over preference matrices
  - Choose from a likely winner set, and an adversary from a likely "discreditor" set

Missing ground-truth

- Infer the ground-truth [Abraham et al. 13, Liu and Liu 15]
- Repeated test over labels & aggregate (sequential hypothesis testing, “crowd within”), e.g.,

$$1(\sum_{t=1}^{T} X(t)/T \geq 0.5)$$

  - Serve as a noisy ground-truth.
  - Surrogate index.

So far

- Indeed, bandit can be applied to various decision making problems in crowdsourcing
- Unique challenges
  - Full information
  - Partial information
  - Strategic information

- **Long term incentive/reputation system**

Information elicitation for ML

- Information elicitation when its quality depends on endogenous variables
  - E.g, quality of works depends on *effort*, which is not directly observable.
  - w/ or w/o ground-truth: one step payment often suffices. (*scoring rule, peer prediction, etc*)
  - What about future job opportunity?

Basic idea: endogenous arms

- Form a bandit on quality of works [Ghosh and Hummel 13, Liu and Chen 16]
  - Each worker is now an arm.

- Full observation
  - Index policy (*reputation score*)
  - $Q_i(t|e) + \mathrm{Rad}_i(t)$ (*Empirical quality + confidence)*
  - Selection ⇔ Future job opportunity
  - *Form a competition*

- Partial observation (no ground-truth)
  - Peer prediction aided index rule

$$S_i(t|e_i, e_{-i}) + \mathrm{Rad}_i(t)$$

  - Each arm's reward distribution depends also on others' action
  - More convoluted argument

Looking forward…

- Online learning with limited feedbacks
  - Fundamental limit of crowd wisdom in a bandit setting?
- What is the best worker behavior model (arm)?
- Incentive compatible bandit?
- Gossiping..
- Other novel applications of bandit.
- …..

Thank you.

References

- Abraham, I., Alonso, O., Kandylas, V., & Slivkins, A. (2013, February). Adaptive Crowdsourcing Algorithms for the Bandit Survey Problem. In *COLT*(pp. 882-910).
- Auer, P., Cesa-Bianchi, N., & Fischer, P. (2002). Finite-time analysis of the multiarmed bandit problem. *Machine learning*, *47*(2-3), 235-256.
- Badanidiyuru, A., Kleinberg, R., & Slivkins, A. (2013, October). Bandits with knapsacks. In Foundations of Computer Science (FOCS), 2013 IEEE 54th Annual Symposium on (pp. 207-216). IEEE.
- Bresler, G., Shah, D., & Voloch, L. F. (2015). Collaborative Filtering with Low Regret. *arXiv preprint arXiv:1507.05371*.
- Bubeck, S., Stoltz, G., Szepesvári, C., & Munos, R. (2009). Online optimization in X-armed bandits. In Advances in Neural Information Processing Systems (pp. 201-208).
- Chawla, S., Hartline, J. D., & Sivan, B. (2015). Optimal crowdsourcing contests. *Games and Economic Behavior*.
- Ghosh, A., & Hummel, P. (2013, January). Learning and incentives in user-generated content: Multi-armed bandits with endogenous arms. In *Proceedings of the 4th conference on Innovations in Theoretical Computer Science* (pp. 233-246).
- Ho, C. J., Slivkins, A., & Vaughan, J. W. (2016). Adaptive contract design for crowdsourcing markets: Bandit algorithms for repeated principal-agent problems. *Journal of Artificial Intelligence Research*, *55*, 317-359.
- Ho, Chien-Ju, and Jennifer Wortman Vaughan. "Online Task Assignment in Crowdsourcing Markets." *AAAI*. Vol. 12. 2012.

References (cont.)

- Kleinberg, R., Slivkins, A., & Upfal, E. (2008, May). Multi-armed bandits in metric spaces. In Proceedings of the fortieth annual ACM symposium on Theory of computing (pp. 681-690). ACM.
- Li, L., Chu, W., Langford, J., & Schapire, R. E. (2010, April). A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th international conference on World wide web* (pp. 661-670).
- Liu, Y., & Chen, Y. (2016). A Bandit Framework for Strategic Regression. *NIPS 2016, Barcelona, Spain*.
- Liu, Yang, and Mingyan Liu. "An online learning approach to improving the quality of crowd-sourcing." *ACM SIGMETRICS Performance Evaluation Review*. Vol. 43. No. 1. ACM, 2015.
- Massoulié, L., Ohannessian, M. I., & Proutière, A. (2015, June). Greedy-Bayes for targeted news dissemination. In *ACM SIGMETRICS Performance Evaluation Review* (Vol. 43, No. 1, pp. 285-296). ACM.
- Munos, R. (2011). Optimistic optimization of deterministic functions without the knowledge of its smoothness. In Advances in neural information processing systems.
- Mansour, Y., Slivkins, A., & Syrgkanis, V. (2015, June). Bayesian incentive-compatible bandit exploration. In Proceedings of the Sixteenth ACM Conference on Economics and Computation (pp. 565-582). ACM.
- Mansour, Y., Slivkins, A., Syrgkanis, V., & Wu, Z. S. (2016). Bayesian Exploration: Incentivizing Exploration in Bayesian Games. *arXiv preprint arXiv:1602.07570*.

References (cont.)

- Singla, A., & Krause, A. (2013, May). Truthful incentives in crowdsourcing tasks using regret minimization mechanisms. In Proceedings of the 22nd international conference on World Wide Web (pp. 1167-1178). ACM.
- Slivkins, A. (2011). Multi-armed bandits on implicit metric spaces. In Advances in Neural Information Processing Systems (pp. 1602-1610).
- Tran-Thanh, L., Chapman, A., Rogers, A., & Jennings, N. R. (2012). Knapsack based optimal policies for budget-limited multi-armed bandits.arXiv preprint arXiv:1204.1909.
- Tran-Thanh, L., Stein, S., Rogers, A., & Jennings, N. R. (2014). Efficient crowdsourcing of unknown experts using bounded multi-armed bandits. Artificial Intelligence, 214, 89-111.2.
- Yue, Y., Broder, J., Kleinberg, R., & Joachims, T. (2012). The k-armed dueling bandits problem. Journal of Computer and System Sciences, 78(5), 1538-1556.
- Zoghi, M., Karnin, Z. S., Whiteson, S., & De Rijke, M. (2015). Copeland dueling bandits. In Advances in Neural Information Processing Systems (pp. 307-315).
- Zoghi, M., Whiteson, S., Munos, R., & Rijke, M. D. (2014). Relative upper confidence bound for the k-armed dueling bandit problem. In JMLR Workshop and Conference Proceedings (No. 32, pp. 10-18). JMLR.