Leveraging Good Representations in Linear Contextual Bandits

Matteo Papini[§¶], **Andrea Tirinzoni**[§], **Marcello Restelli**[§],
Alessandro Lazaric[†], **Matteo Pirotta**[†]

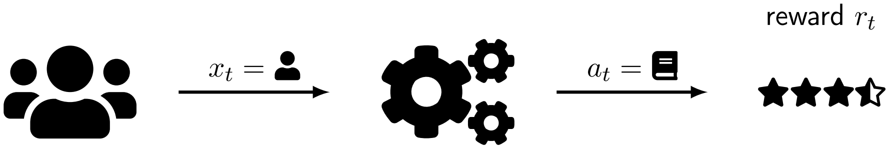
[†] Facebook AI Research, [§] Politecnico di Milano
[¶] work done while at Facebook

International Conference on Machine Learning (ICML), 2021

FACEBOOK AI

Contextual Linear Bandits $r(x, a) = \langle \phi(x, a), \theta^\star \rangle$

- $x_t \sim \rho$
- $a_t \in \{1, \ldots, A\}$
- $r_t = r(x_t, a) + \text{noise}$



minimize $\quad R(n) = \sum_{t=1}^{n} r(x_t, a^\star_{x_t}) - r(x_t, a_t)$

Several algorithms have been designed for this problem, e.g., LinUCB [Chu et al., 2011], OFUL [Abbasi-Yadkori et al., 2011], Thompson Sampling [Abeille and Lazaric, 2017]

* $a^\star_x = \underset{a}{\operatorname{argmax}}\{r(x, a)\}$

Contextual Linear Bandit: LinUCB

[Chu et al., 2011, Abbasi-Yadkori et al., 2011]

Setting:

$$r(x_t, a_t) = \langle \phi(x_t, a_t), \theta^\star \rangle$$

- known realizable d_ϕ-dimensional representation ϕ
- unknown parameter $\theta^\star \in \mathbb{R}^{d_\phi}$

* minimum gap assumption: $r(x, a_x^\star) - r(x, a) > \Delta$ for all $x, a \neq a_x^\star$

Contextual Linear Bandit: LinUCB

[Chu et al., 2011, Abbasi-Yadkori et al., 2011]

Setting:

$$r(x_t, a_t) = \langle \phi(x_t, a_t), \theta^\star \rangle$$

- known realizable d_ϕ-dimensional representation ϕ
- unknown parameter $\theta^\star \in \mathbb{R}^{d_\phi}$

LinUCB:

- Estimate $\theta^\star$ to build *Upper Confidence Bound* $U_t \geq r$ with high probability
- $a_t = \max_a U_t(x_t, a)$

then, LinUCB suffers a *problem-dependent* regret:

$$R_n \lesssim \frac{d_\phi^2}{\Delta} \ln^2(n)$$

* minimum gap assumption: $r(x, a_x^\star) - r(x, a) > \Delta$ for all $x, a \neq a_x^\star$

Contextual Linear Bandit: LinUCB

[Chu et al., 2011, Abbasi-Yadkori et al., 2011]

Setting:

$$r(x_t, a_t) = \langle \phi(x_t, a_t), \theta^\star \rangle$$

- known realizable d_ϕ-dimensional representation ϕ
- unknown parameter $\theta^\star \in \mathbb{R}^{d_\phi}$

LinUCB:

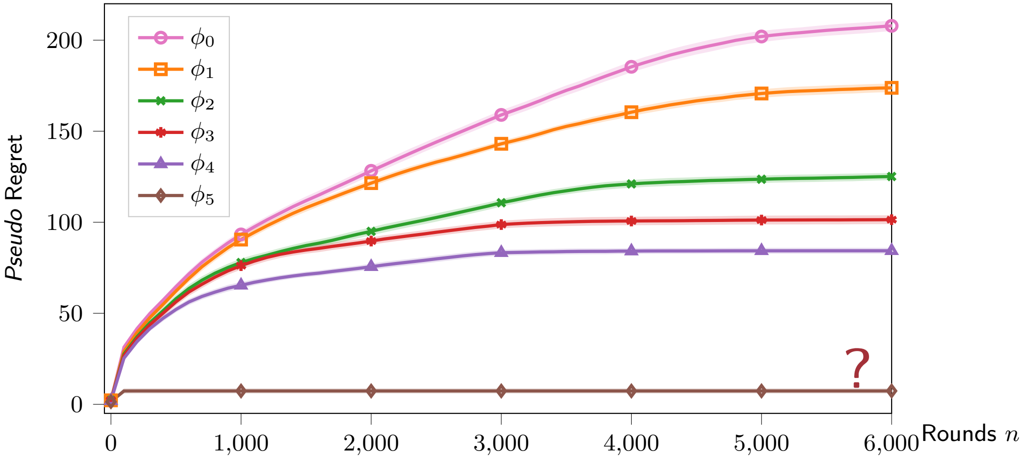
- Estimate $\theta^\star$ to build *Upper Confidence Bound* $U_t \geq r$ with high probability
- $a_t = \max_a U_t(x_t, a)$

then, LinUCB suffers a *problem-dependent* regret:

$$R_n \lesssim \frac{d_\phi^2}{\Delta} \ln^2(n)$$

* minimum gap assumption: $r(x, a_x^\star) - r(x, a) > \Delta$ for all $x, a \neq a_x^\star$

What is a Good Representation?



All realizable! Same dimension! Same algorithm! (LinUCB)

Good Representations for LinUCB

We say ϕ is good if:*

$$\mathrm{span}\{\phi(x, a_x^\star) \mid x \in \mathrm{supp}(\rho)\} = \mathbb{R}^{d_\phi}$$

Our result

LinUCB achieves `CONSTANT` regret if and only if ϕ is good:

$$R_n \lesssim \frac{d_\phi^2}{\Delta} \ln^2(\tau_\phi)$$

where

$$\tau_\phi \lesssim \left(\frac{d_\phi^2}{\lambda_\phi \Delta}\right)^2 \qquad \text{(constant)}$$

$$\lambda_\phi = \lambda_{\min}\Big(\mathbb{E}_\rho[\phi(x, a_x^\star)\phi(x, a_x^\star)^\top]\Big) \qquad (\lambda_\phi > 0 \text{ iff } \phi \text{ good})$$

*good feature ϕ introduced in [Hao et al., 2020]. ** $a_x^\star = \underset{a}{\mathrm{argmax}}\{r(x,a)\}$

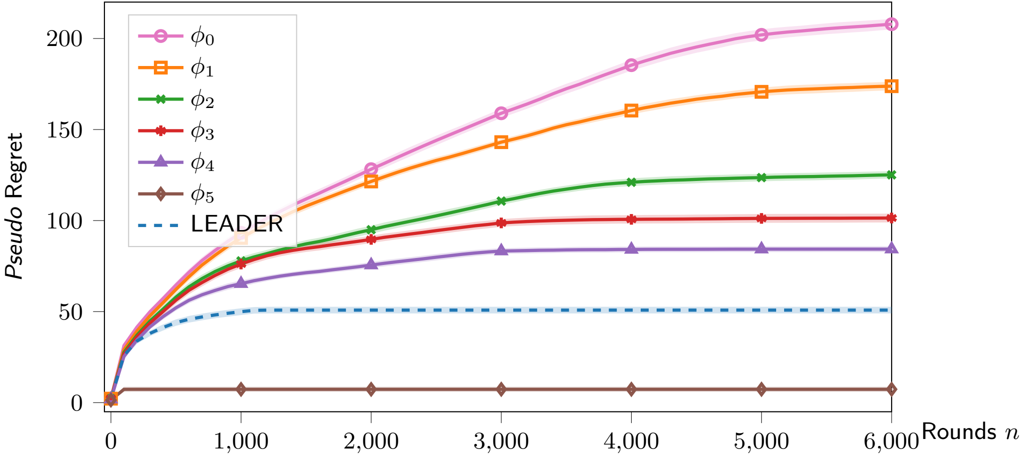
Strong Assumption? Not really...

We fit neural networks to real datasets from [Bietti et al., 2018]

Network size	Good representations	λ_ϕ (median)	$\mathbf{R^2}$
$(16, 16)$	88.8%	0.001833	0.67
$(32, 32)$	86.4%	0.001502	0.76
$(64, 64)$	79.9%	0.001004	0.83

⚠ Good representations seem to be quite common in practice!

Can we Leverage Good Representations?



Given M realizable representations $\phi_1, \ldots, \phi_M$ (good or not)

The LEADER Algorithm

Given M realizable representations $\phi_1, \ldots, \phi_M$

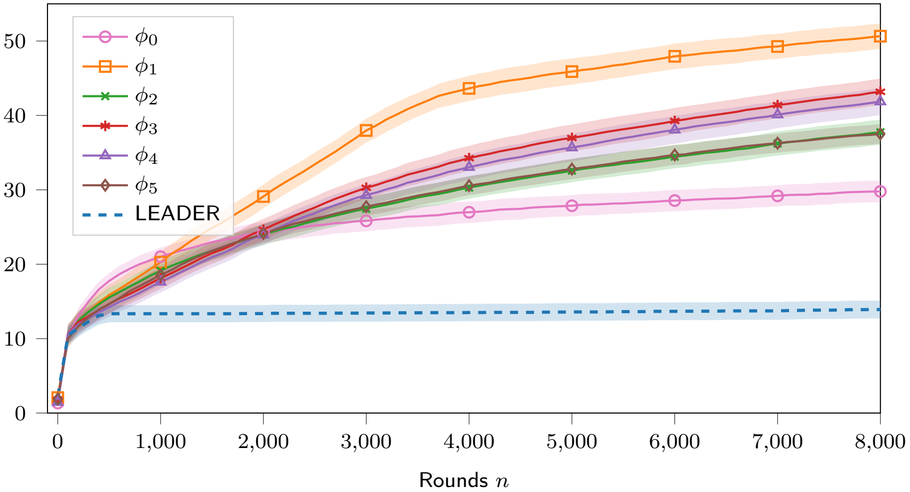
1. Estimate θ_i for each ϕ_i using all data
2. Build UCB $U_{ti} \geq r$ with high probability
3. $a_t = \max_a \min_i U_{ti}(x_t, a)$

Regret of LEADER

$$R_n(\text{LEADER}) \leq \ln(M) \min_{i \in [M]} R_n(\phi_i)$$

and is constant if at least one of the M representations is good

Mixing Representations



$\phi_1, \ldots, \phi_M$ are all NOT good

Still, LEADER achieves constant regret

Find Out How...

- LEADER can mix representations to form a good representation and achieve constant regret
- Redundant representations can also be good
- The minimum gap assumption can be removed
- Misspecified representations can be eliminated fast
- Constant regret is observed in real problems with neural networks
- LEADER is competitive with model-selection algorithms

And stay tuned for representation selection in Linear MDPs

Details are in the paper:

Leveraging Good Representations in Linear Contextual Bandits

ICML 2021

Matteo Papini, Andrea Tirinzoni, Marcello Restelli, Alessandro Lazaric, Matteo Pirotta

Thank you

Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. Improved algorithms for linear stochastic bandits. In *NIPS*, pages 2312–2320, 2011.

Marc Abeille and Alessandro Lazaric. Linear thompson sampling revisited. In *AISTATS*, volume 54 of *Proceedings of Machine Learning Research*, pages 176–184. PMLR, 2017.

Alberto Bietti, Alekh Agarwal, and John Langford. A contextual bandit bake-off. *arXiv preprint arXiv:1802.04064*, 2018.

Wei Chu, Lihong Li, Lev Reyzin, and Robert E. Schapire. Contextual bandits with linear payoff functions. In *AISTATS*, volume 15 of *JMLR Proceedings*, pages 208–214. JMLR.org, 2011.

Botao Hao, Tor Lattimore, and Csaba Szepesvári. Adaptive exploration in linear contextual bandit. In *AISTATS*, volume 108 of *Proceedings of Machine Learning Research*, pages 3536–3545. PMLR, 2020.