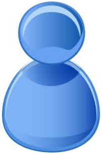
Multi-task Linear Bandits

Marta Soare — *INRIA Lille - Nord Europe*
Alessandro Lazaric — *INRIA Lille - Nord Europe*
Ouais Alsharif — *McGill University & Google*
Joelle Pineau — *McGill University*

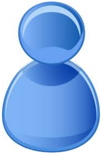
NIPS2014 Workshop on Transfer and Multi-task Learning : Theory meets Practice — Montreal, December 2014

Motivating example: Movie recommendation system



- Features: age, profession, ...
- Ratings
- **Learn his preference as fast as possible!**

Motivating example: Movie recommendation system



- Features: age, profession, ...
- Ratings
- **Learn his preference as fast as possible!**

- Features: age, profession, ...
- **Learn his preference faster (with fewer ratings)!**

Sequential multi-task linear bandit

Linear bandit task:

- Set of arms $\mathcal{X} \subseteq \mathbb{R}^d$, $|\mathcal{X}| = K$ and $||x||_2 \leq L$, $\forall x \in \mathcal{X}$.
- Reward model for task j

$$r_j(x) = x^\top \theta_j + \eta$$

where $\theta_j \in \mathbb{R}^d$ is **unknown** and η is an R sub-Gaussian noise.

- Cumulative regret w.r.t. the optimal arm $x_j^* = \arg\max_{x \in \mathcal{X}} x^\top \theta_j$

$$R_{n_j} = \sum_{s=1}^{n_j} (x_j^* - x_s)^\top \theta_j$$

Sequential multi-task linear bandit

Linear bandit task:

- Set of arms $\mathcal{X} \subseteq \mathbb{R}^d$, $|\mathcal{X}| = K$ and $||x||_2 \leq L$, $\forall x \in \mathcal{X}$.
- Reward model for task j

$$r_j(x) = x^\top \theta_j + \eta$$

  where $\theta_j \in \mathbb{R}^d$ is **unknown** and η is an R sub-Gaussian noise.

- Cumulative regret w.r.t. the optimal arm $x_j^* = \arg\max_{x \in \mathcal{X}} x^\top \theta_j$

$$R_{n_j} = \sum_{s=1}^{n_j} (x_j^* - x_s)^\top \theta_j$$

Task Similarity: there exists $\varepsilon > 0$ such that for any pair of tasks (θ, θ') we have $||\theta - \theta'||_2 \leq \varepsilon$.

Sequential multi-task linear bandit

Linear bandit task:

- Set of arms $\mathcal{X} \subseteq \mathbb{R}^d$, $|\mathcal{X}| = K$ and $||x||_2 \leq L$, $\forall x \in \mathcal{X}$.
- Reward model for task j

$$r_j(x) = x^\top \theta_j + \eta$$

where $\theta_j \in \mathbb{R}^d$ is **unknown** and η is an R sub-Gaussian noise.

- Cumulative regret w.r.t. the optimal arm $x_j^* = \arg\max_{x \in \mathcal{X}} x^\top \theta_j$

$$R_{n_j} = \sum_{s=1}^{n_j} (x_j^* - x_s)^\top \theta_j$$

Task Similarity: there exists $\varepsilon > 0$ such that for any pair of tasks (θ, θ') we have $||\theta - \theta'||_2 \leq \varepsilon$.

Improve performance on a particular task by leveraging information from previous tasks!

Single task estimates:

OLS estimate: For any task j, the estimate of θ_j after n rewards is

$$\hat{\theta}_{j,n} = A_{j,n}^{-1} b_{j,n}$$

$$A_{j,n} = \sum_{t=1}^{n} x_t x_t^\top \qquad b_{j,n} = \sum_{t=1}^{n} x_t r_t$$

Prediction error: Thm.2 in [Abbasi Yadkori et al., NIPS 2012] (w.p. $1-\delta$)

$$\left|x^\top \theta_j - x^\top \hat{\theta}_{j,n}^{\lambda}\right| \leq \underbrace{||x||_{(A_{j,n}^{\lambda})^{-1}} \left(R\sqrt{d \log\left(\frac{1+nL^2/\lambda}{\delta}\right)} + \lambda^{1/2}||\theta_j|| \right)}_{B_{j,n}(x)}$$

Multi-task estimates :

$$\widetilde{\theta}_{m+1,t} = \widetilde{A}_{m+1,t}^{-1}\widetilde{b}_{m+1,t} \qquad \mathbb{E}\big[\widetilde{\theta}_{m+1,t}\big] = \widetilde{\theta}^{*}_{m+1,t}$$

$$\widetilde{A}_{m+1,t} = \sum_{j=1}^{m} A_j + A_{m+1,t} \qquad \widetilde{b}_{m+1,t} = \sum_{j=1}^{m} b_j + b_{m+1,t}$$

Theorem

Let $\widetilde{\theta}^{\lambda}_{m+1,t}$ be the multi-task regularized least-squares estimate. Then, for any $\delta \geq 0$, for any $t \geq 1$, with probability greater than $1-\delta$ it holds that:

$$|x^{\top}(\widetilde{\theta}^{\lambda}_{m+1,t} - \theta_{m+1})|$$

$$\leq \underbrace{||x||_{(\widetilde{A}^{\lambda}_{m+1,t})^{-1}} \left(R\sqrt{d \log\Big(\frac{\det(\widetilde{A}^{\lambda}_{m+1,t})^{1/2}\det(\lambda I)^{-1/2}}{\delta}\Big)} + \lambda^{1/2} S \right) + \varepsilon||x||}_{\widetilde{B}_{m+1,t}(x)}.$$

Multi-task estimates :

$$\widetilde{\theta}_{m+1,t} = \widetilde{A}^{-1}_{m+1,t}\widetilde{b}_{m+1,t} \qquad \mathbb{E}\big[\widetilde{\theta}_{m+1,t}\big] = \widetilde{\theta}^*_{m+1,t}$$

$$\widetilde{A}_{m+1,t} = \sum_{j=1}^{m} A_j + A_{m+1,t} \qquad \widetilde{b}_{m+1,t} = \sum_{j=1}^{m} b_j + b_{m+1,t}$$

Theorem

Let $\widetilde{\theta}^{\lambda}_{m+1,t}$ be the multi-task regularized least-squares estimate. Then, for any $\delta \geq 0$, for any $t \geq 1$, with probability greater than $1-\delta$ it holds that:

$$|x^\top(\widetilde{\theta}^{\lambda}_{m+1,t} - \theta_{m+1})|$$

$$\leq \underbrace{||x||_{(\widetilde{A}^{\lambda}_{m+1,t})^{-1}} \left(R\sqrt{d\log\left(\frac{\det(\widetilde{A}^{\lambda}_{m+1,t})^{1/2}\det(\lambda I)^{-1/2}}{\delta}\right)} + \lambda^{1/2} S \right) + \varepsilon||x||}_{\widetilde{B}_{m+1,t}(x)}.$$

Multi-task estimates :

$$\widetilde{\theta}_{m+1,t} = \widetilde{A}_{m+1,t}^{-1}\widetilde{b}_{m+1,t} \qquad \mathbb{E}\left[\widetilde{\theta}_{m+1,t}\right] = \widetilde{\theta}^{*}_{m+1,t}$$

$$\widetilde{A}_{m+1,t} = \sum_{j=1}^{m} A_j + A_{m+1,t} \qquad \widetilde{b}_{m+1,t} = \sum_{j=1}^{m} b_j + b_{m+1,t}$$

Theorem

Let $\widetilde{\theta}^{\lambda}_{m+1,t}$ be the multi-task regularized least-squares estimate. Then, for any $\delta \geq 0$, for any $t \geq 1$, with probability greater than $1-\delta$ it holds that:

$$|x^{\top}(\widetilde{\theta}^{\lambda}_{m+1,t} - \theta_{m+1})|$$

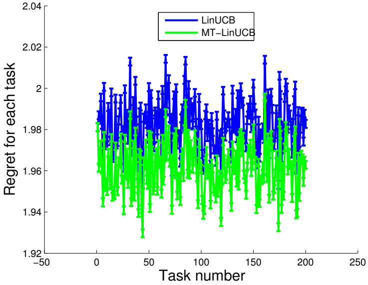
$$\leq \underbrace{||x||_{(\widetilde{A}^{\lambda}_{m+1,t})^{-1}} \left(R\sqrt{d \log\left(\frac{\det(\widetilde{A}^{\lambda}_{m+1,t})^{1/2}\det(\lambda I)^{-1/2}}{\delta}\right)} + \lambda^{1/2} S \right) + \varepsilon ||x||.}_{\widetilde{B}_{m+1,t}(x)}$$

MT-LinUCB

Input: budgets $\{n_j\}_j$, arms $\mathcal{X} \subset \mathbb{R}^d$, regularizer λ
$j = 1$
$A = \lambda I_d, \widetilde{b} = b = 0_d, \widetilde{A}_j = \lambda I_d, \hat{\theta}_j = A^{-1} b$
for $j = 1, \ldots, m+1$ **do**
Compute the multi-task OLS solution:
$\widetilde{A}_j = \widetilde{A}_j + A - \lambda I_d,\ \widetilde{b} = \widetilde{b} + b,\ \widetilde{\theta}_j = \widetilde{A}_j^{-1} \widetilde{b}$
Compute the task-specific OLS solution:
$A = \lambda I_d, b = 0_d, \hat{\theta}_j = A^{-1} b$
for $t = 1, \ldots, n_j$ **do**
Select arms according to the tightest bound:
$x_t = \arg\max_{x \in \mathcal{X}} \min \left[x^\top \hat{\theta}_j + B_{j,t}(x);\ x^\top \widetilde{\theta}_j + \widetilde{B}_{j,t}(x) \right]$
Observe reward: $r_t = x_t^\top \theta_j + \eta_t$
Update: $A, b, \hat{\theta}_j, B_{j,t}, \widetilde{A}_j, \widetilde{b}, \widetilde{\theta}_j, \widetilde{B}_{j,t}$
end for
end for

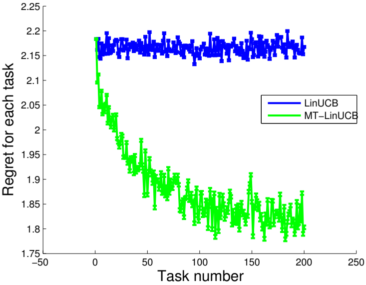
Experiments

- 200 tasks with $\theta_1, \ldots, \theta_{200} \in \mathbb{R}^2$ randomly generated
- $\max_\theta ||\theta||_2 = 1.1 \cdot \sqrt{2}$
- 100 samples for each task
- $\varepsilon = 0.8 \cdot \sqrt{2}$
- MT-LinUCB is never worse than LinUCB.



Experiments

- As the difference between tasks reduces, the advantage of MT-LinUCB becomes more evident.
- For $\varepsilon = 0.2\sqrt{2}$, the regret of MT-LinUCB decreases with every additional task, while the regret for LinUCB remains constant over time.



Follow up work

- Regret bound for MT-LinUCB.
- Additional experiments on datasets (see Boston Housing example on the poster).
- Weighting scheme according to task relevance.

Thank you!