6998-2 Advanced Machine Learning for Personalization Home Work #2

va2361: Vishal Anand

March 25, 2018

Contextual bandit: LinUCB

Max Percentage: 94.762%

α **search**			
α	Cumulative-Take-Rate	Reward	Match
$1/\sqrt{t}$	0.91763	947	1032
$0.1/\sqrt{t}$	0.94589	979	1035
$\mathbf{0.09}/\sqrt{t}$	**0.94762**	**977**	**1031**
$8.08/\sqrt{t}$	0.94547	971	1027
$0.085/\sqrt{t}$	0.94731	971	1025
$0.095/\sqrt{t}$	0.94762	977	1031
$10.088/\sqrt{t}$	0.60176	615	1022
$0.5088/t$	0.91185	931	1021
$5.5088/t$	0.93641	972	1038
$3.6088/t$	0.93695	966	1031
$50.6088/t$	0.91023	943	1036
$50.6088/2t$	0.91779	949	1034
$25.6088/(\sqrt{t}+2t)$	0.93030	961	1033
$\mathbf{10.6088}/\ (\sqrt{t})+\mathbf{2t}$	**0.93743**	**974**	**1039**
$18.6088/(\sqrt{t}+t^3)$	0.92210	947	1027
$0.085/t^{0.25}$	0.92843	960	1034
$\mathbf{0.0085}/t^{0.25}$	**0.94680**	**979**	**1034**
$0.0085/t^{0.125}$	0.94731	971	1025
$0.0085/t^{0.2975}$	0.94498	979	1036
$0.0085/t^{0.2999}$	0.94083	970	1031

Table 1: Alphas and Cumulative-Take-Rates

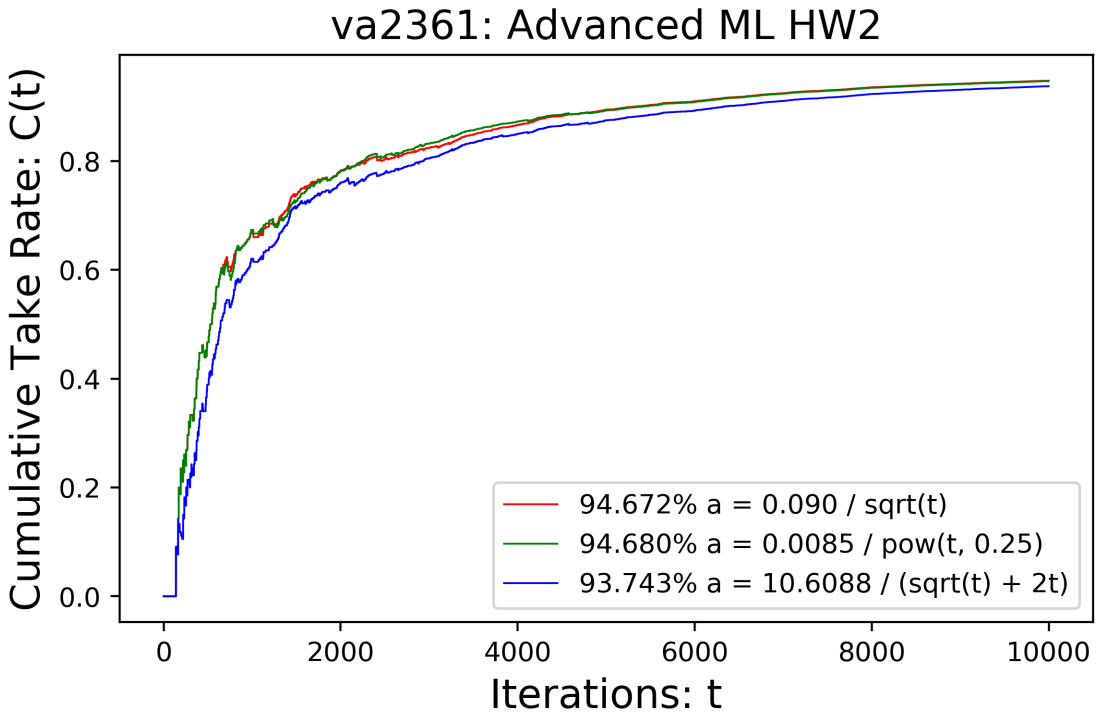
Algorithm 1 LinUCB with disjoint linear models **modified for current dataset**

Inputs: $\alpha \in \mathbb{R}_+$
for $t = 1, 2, 3, ..., T$ **do**
  Observe features of all arms $a \in \mathbb{A}_t : x_{t,a} \in \mathbb{R}^d$
  for all $a \in \mathbb{A}_t$ **do**
    if a is new **then**
      $A_a \leftarrow I_d$ (d-dimensional identity matrix)
      $b_a \leftarrow 0_{d \times 1}$ (d-dimensional zero matrix)
    end if
    $\hat{\theta} \leftarrow \mathbb{A}_a^{-1} b_a$
    $p_{t,a} \leftarrow \hat{\theta}^T x_t + \alpha \sqrt{x_t^T \mathbb{A}_a^{-1} x_t}$
  end for
  Choose arm $a_t = \arg max_a \in \mathbb{A}_t p_{t,a}$ with ties broken arbitrarily, and observe a real-valued payoff r_t
  $\mathbb{A}_{a_t} \leftarrow \mathbb{A}_{a_t} + x_t x_t^T$
  $b_{a_t} \leftarrow b_{a_t} + r_t x_t$
end for

Algorithm 2 Policy evaluator (with finite data stream)

bandit algorithm A; stream of events S of length L
$h_0 \leftarrow \emptyset$ An initially empty history
$\hat{\mathbb{G}}_A \leftarrow 0$ An initially zero total payoff
$T \leftarrow 0$ An initially zero counter of valid events
for $t = 1, 2, 3, ..., T$ **do**
  Get the t-th event (x, a, r_a) from S
  if A$(h_{t-1}, x) =$ a **then**
    $h_t \leftarrow CONCATENATE(h_{t-1}, (a, a, r_a))$
    $\hat{\mathbb{G}}_A \leftarrow \hat{\mathbb{G}}_A + r_a$
    $T \leftarrow T + 1$
    $b_a \leftarrow 0_{d \times 1}$ (d-dimensional zero matrix)
  else
    $h_t \leftarrow h_{t-1}$
  end if
end for
Output: $\hat{\mathbb{G}}_A / T$

Graphs



Algorithms adapted from [1] and [2]

Code is attached with the zip file, and the same is also at `https://github.com/vishalanand/6998-HW2-linUCB`

References

[1] Lihong Li, Wei Chu, John Langford and Robert E. Schapire *A Contextual-Bandit Approach to Personalized News Article Recommendation*
`https://arxiv.org/abs/1003.0146`

[2] Lihong Li, Wei Chu and John Langford *An Unbiased, Data-Driven, Offline Evaluation Method of Contextual Bandit Algorithms*
`https://arxiv.org/abs/1003.5956`