Discounted UCB

Levente Kocsis and Csaba Szepesvári

MTA SZTAKI, Hungary
UCB1-tuned

Discounted UCB1-tuned

Experiments

Other algorithms

Conclusions
UCB1-tuned+

\[
s_{it} = \sum_{\tau=0}^{t} \mathbb{I}(l_\tau = i) x_\tau
\]

\[
n_{it} = \sum_{\tau=0}^{t} \mathbb{I}(l_\tau = i)
\]

\[
\mu_{it} = \frac{s_{it}}{n_{it}} \quad n_t = \sum_i n_{it}
\]

\[
l_{t+1} = \arg\max_i \left( \mu_{it} + \sqrt{\frac{\max(\mu_{it}(1 - \mu_{it}), 0.002) \ln n_t}{n_{it}}} \right)
\]

Levente Kocsis and Csaba Szepesvári  Discounted UCB
Discounted UCB1-tuned+

\[ S_{it} = \sum_{\tau=0}^{t} \mathbb{I}(l_{\tau} = i) \gamma^{t-\tau} x_{\tau} \]

\[ n_{it} = \sum_{\tau=0}^{t} \mathbb{I}(l_{\tau} = i) \gamma^{t-\tau} \]

\[ \mu_{it} = \frac{S_{it}}{n_{it}} \quad n_t = \sum_{i} n_{it} \]

\[ l_{t+1} = \arg\max_{i} \left( \mu_{it} + \sqrt{\frac{\max(\mu_{it}(1 - \mu_{it}), 0.002) \ln n_t}{n_{it}}} \right) \]
Experiments: Task 1 (averaged over 1000 seeds)
Experiments: Task 1 (averaged over test seeds)
Experiments: Task 2 (averaged over 1000 seeds)
Experiments: Task 2 (averaged over test seeds)
Experiments: Task 3 (averaged over 1000 seeds)
Experiments: Task 3 (averaged over test seeds)

![Chart showing regret vs iteration for UCB1-tuned, Exp3, and Exp3 with different gamma values.]
Other algorithms

- line fitting
- discounted UCB + exploiting periodicity
- adaptive discounted UCB
Conclusions

- Challenging challenge
- Task 4(?): mixing task 1 and 2
- Regret bounds depending on how fast the response rate vary?
- Universal algorithms (algorithms adapting to response rate)