Efficient-UCBV: An Almost Optimal Algorithm using Variance Estimates

Subhojyoti Mukherjee (IIT Madras)
Dr. K.P. Naveen (IIT Tirupati)
Dr. Nandan Sudarsanam (IIT Madras, RBC-DSAI)
Dr. Balaraman Ravindran (IIT Madras, RBC-DSAI)

AAAI 2018, New Orleans, Louisiana, USA

Feb 6, 2018

Overview

Stochastic Multi-Armed Bandit Problem (SMAB)

- A finite set of actions or arms belonging to set $\mathbb{A}$ such that $|\mathbb{A}| = K$.

Stochastic Multi-Armed Bandit Problem (SMAB)

- A finite set of actions or arms belonging to set $\mathbb{A}$ such that $|\mathbb{A}| = K$.
- The rewards for each of the arms, $X_{i,t} \sim^{i.i.d} D_i$.

Stochastic Multi-Armed Bandit Problem (SMAB)

- A finite set of actions or arms belonging to set $\mathbb{A}$ such that $|\mathbb{A}| = K$.
- The rewards for each of the arms, $X_{i,t} \sim^{i.i.d} D_i$.
- The learner does not know the mean r_i or the variance σ_i^2 of the distribution $D_i, \forall i \in \mathbb{A}$.

Stochastic Multi-Armed Bandit Problem (SMAB)

- A finite set of actions or arms belonging to set $\mathbb{A}$ such that $|\mathbb{A}| = K$.
- The rewards for each of the arms, $X_{i,t} \sim^{i.i.d} D_i$.
- The learner does not know the mean r_i or the variance σ_i^2 of the distribution $D_i, \forall i \in \mathbb{A}$.
- **Vital Assumption:** $D_i, \forall i \in \mathbb{A}$ are fixed throughout the time horizon denoted by T.

Problem Definition of SMAB

- **Primary aim:** Minimize the cumulative regret by identifying the arm whose expected mean is r^* such that $r^* > r_i, \forall i \in \mathbb{A}$.

Problem Definition of SMAB

- **Primary aim:** Minimize the cumulative regret by identifying the arm whose expected mean is r^* such that $r^* > r_i, \forall i \in \mathbb{A}$.
- **Condition:** This has to be achieved within a finite T timesteps.

Problem Definition of SMAB

- **Primary aim:** Minimize the cumulative regret by identifying the arm whose expected mean is r^* such that $r^* > r_i, \forall i \in \mathbb{A}$.
- **Condition:** This has to be achieved within a finite T timesteps.
- The expected regret of an algorithm after T timesteps is give by,

$$\mathbb{E}[R_T] = \sum_{i=1}^{K} \mathbb{E}[z_i(T)]\Delta_i,$$

where $\Delta_i = r^* - r_i$ is the gap.

Problem Definition of SMAB

- **Primary aim:** Minimize the cumulative regret by identifying the arm whose expected mean is r^* such that $r^* > r_i, \forall i \in \mathbb{A}$.
- **Condition:** This has to be achieved within a finite T timesteps.
- The expected regret of an algorithm after T timesteps is give by,

$$\mathbb{E}[R_T] = \sum_{i=1}^{K} \mathbb{E}[z_i(T)]\Delta_i,$$

  where $\Delta_i = r^* - r_i$ is the gap.
- $LB(R_T) \geq \Omega\left(\sqrt{KT}\right)$ and good learner should have atmost $UB(R_T) \leq O\left(\sqrt{KT \log T}\right)$.

Why Study this Problem?

- **General:** The simple SMAB forms the building block to the larger multi-state RL problems.

Why Study this Problem?

- **General:** The simple SMAB forms the building block to the larger multi-state RL problems.
- **Algorithmic:** No round-based arm-elimination variance-aware algorithm exists which performs well empirically.

Why Study this Problem?

- **General:** The simple SMAB forms the building block to the larger multi-state RL problems.
- **Algorithmic:** No round-based arm-elimination variance-aware algorithm exists which performs well empirically.
- **Theoretical:** No round-based arm-elimination variance-aware algorithm exists which reaches order-optimal regret bound of $O\left(\sqrt{KT}\right)$.

Contributions in SMAB

- We propose the Efficient-UCB-Variance (EUCBV) algorithm for the SMAB setting.

Contributions in SMAB

- We propose the Efficient-UCB-Variance (EUCBV) algorithm for the SMAB setting.
- EUCBV takes into account the empirical variances of the arms along with mean estimates to quickly find the optimal arm.

Contributions in SMAB

- We propose the Efficient-UCB-Variance (EUCBV) algorithm for the SMAB setting.
- EUCBV takes into account the empirical variances of the arms along with mean estimates to quickly find the optimal arm.
- It is the first variance-based arm elimination algorithm for the considered SMAB setting.

Contributions in SMAB

- We propose the Efficient-UCB-Variance (EUCBV) algorithm for the SMAB setting.
- EUCBV takes into account the empirical variances of the arms along with mean estimates to quickly find the optimal arm.
- It is the first variance-based arm elimination algorithm for the considered SMAB setting.
- It addresses an open problem discussed in Auer and Ortner (2010) of designing an algorithm that can eliminate arms based on variance estimates.

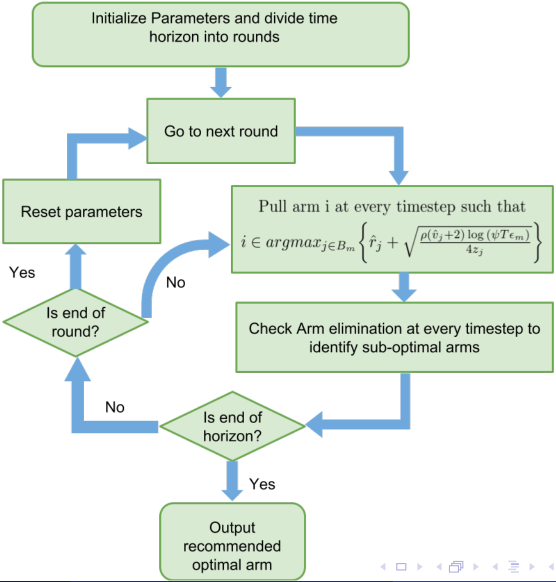
Contributions in SMAB

- We propose the Efficient-UCB-Variance (EUCBV) algorithm for the SMAB setting.
- EUCBV takes into account the empirical variances of the arms along with mean estimates to quickly find the optimal arm.
- It is the first variance-based arm elimination algorithm for the considered SMAB setting.
- It addresses an open problem discussed in Auer and Ortner (2010) of designing an algorithm that can eliminate arms based on variance estimates.
- Theoretically it achieves an order-optimal regret bound, the first for an arm elimination algorithm in SMAB setting.

Contributions in SMAB

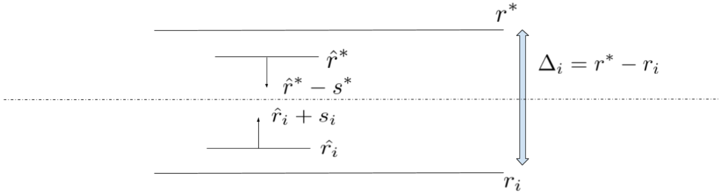
- We propose the Efficient-UCB-Variance (EUCBV) algorithm for the SMAB setting.
- EUCBV takes into account the empirical variances of the arms along with mean estimates to quickly find the optimal arm.
- It is the first variance-based arm elimination algorithm for the considered SMAB setting.
- It addresses an open problem discussed in Auer and Ortner (2010) of designing an algorithm that can eliminate arms based on variance estimates.
- Theoretically it achieves an order-optimal regret bound, the first for an arm elimination algorithm in SMAB setting.
- Empirically, it outperforms all the state-of-the-art algorithms for the considered environments.

EUCBV Algorithm for SMAB



EUCBV Arm Elimination

Arm Elimination: $\hat{r}_i + s_i < \max_{j \in B_m} \{\hat{r}_j - s_j\}$ **Definition:** $s_i = \dfrac{\rho(\hat{v}_i + 2)\log(T\epsilon_m)}{2n_i}$



Expected Regret of EUCBV

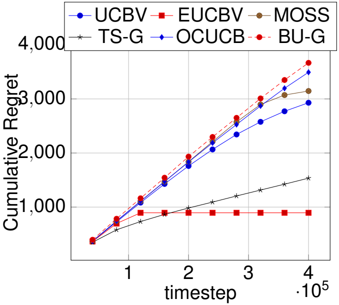
Corollary (***Gap-Independent Bound***)

The regret of EUCBV is upper bounded by the following gap-independent expression:

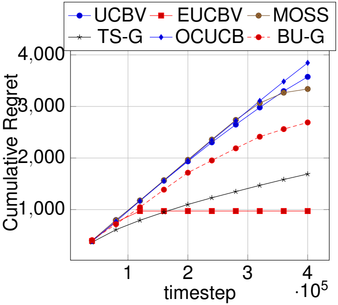
$$\mathbb{E}[R_T] \leq \frac{C_3 K^5}{T^{\frac{1}{4}}} + 80\sqrt{KT}.$$

Algorithm	GD Bound	GI Bound	Var
EUCBV	$O\left(\frac{K\sigma_{\max}^2 \log(\frac{T\Delta^2}{K})}{\Delta}\right)$	$O\left(\sqrt{KT}\right)$	Yes
UCBV	$O\left(\frac{K\sigma_{\max}^2 \log T}{\Delta}\right)$	$O\left(\sqrt{KT \log T}\right)$	Yes
MOSS	$O\left(\frac{K^2 \log(T\Delta^2/K)}{\Delta}\right)$	$O\left(\sqrt{KT}\right)$	No
OCUCB	$O\left(\frac{K \log(T/H_i)}{\Delta}\right)$	$O\left(\sqrt{KT}\right)$	No

Experiments in SMAB



(a) Expt-1: Failure of TS



(b) Expt-2: 3 Group Variance

Conclusions

- We proposed the EUCBV algorithm for the SMAB setting which uses variance and mean estimation along with arm elimination to find the optimal arm.

Conclusions

- We proposed the EUCBV algorithm for the SMAB setting which uses variance and mean estimation along with arm elimination to find the optimal arm.
- Theoretically, EUCBV achieves an order-optimal regret guarantees, but further studies are required to reduce the constants.

Conclusions

- We proposed the EUCBV algorithm for the SMAB setting which uses variance and mean estimation along with arm elimination to find the optimal arm.
- Theoretically, EUCBV achieves an order-optimal regret guarantees, but further studies are required to reduce the constants.
- A more detailed analysis of the non-uniform arm selection and parameter selection is also required for EUCBV.

Thank You