A Novel Confidence-Based Algorithm for Structured Bandits

Andrea Tirinzoni[1], Alessandro Lazaric[2], and Marcello Restelli[1]

[1] Politecnico di Milano
[2] Facebook AI Research

POLITECNICO
MILANO 1863

Motivation

Motivation


- Real problems often exhibit **structure** $\rightarrow$ Arms are *correlated*

Motivation


- Real problems often exhibit **structure** $\rightarrow$ Arms are *correlated*
  - User preference about one product might reveal preferences about other products

Motivation


- Real problems often exhibit **structure** $\rightarrow$ Arms are *correlated*
  - User preference about one product might reveal preferences about other products
  - Response to some medical treatment influences responses to other treatments

Motivation


- Real problems often exhibit **structure** $\rightarrow$ Arms are *correlated*
  - User preference about one product might reveal preferences about other products
  - Response to some medical treatment influences responses to other treatments
- Exploiting structure can significantly reduce **regret** of bandit algorithms

Bandit Structures

- Several *specific structures* studied in the literature

Bandit Structures

- Several *specific structures* studied in the literature
  - Linear [Dani et al., 2008, Chu et al., 2011, Abbasi-Yadkori et al., 2011]

Bandit Structures

- Several *specific structures* studied in the literature
  - Linear [Dani et al., 2008, Chu et al., 2011, Abbasi-Yadkori et al., 2011]
  - Combinatorial [Cesa-Bianchi and Lugosi, 2012]

Bandit Structures

- Several *specific structures* studied in the literature
  - Linear [Dani et al., 2008, Chu et al., 2011, Abbasi-Yadkori et al., 2011]
  - Combinatorial [Cesa-Bianchi and Lugosi, 2012]
  - Lipschitz [Magureanu et al., 2014]

Bandit Structures

- Several *specific structures* studied in the literature
  - Linear [Dani et al., 2008, Chu et al., 2011, Abbasi-Yadkori et al., 2011]
  - Combinatorial [Cesa-Bianchi and Lugosi, 2012]
  - Lipschitz [Magureanu et al., 2014]
  - ...

Bandit Structures

- Several *specific structures* studied in the literature
  - Linear [Dani et al., 2008, Chu et al., 2011, Abbasi-Yadkori et al., 2011]
  - Combinatorial [Cesa-Bianchi and Lugosi, 2012]
  - Lipschitz [Magureanu et al., 2014]
  - ...
- Algorithms to exploit any **general structures** are desirable in practice

Bandit Structures

- Several *specific structures* studied in the literature
  - Linear [Dani et al., 2008, Chu et al., 2011, Abbasi-Yadkori et al., 2011]
  - Combinatorial [Cesa-Bianchi and Lugosi, 2012]
  - Lipschitz [Magureanu et al., 2014]
  - ...
- Algorithms to exploit any **general structures** are desirable in practice
  - Learner knows a subset of realizable bandit problems

Bandit Structures

- Several *specific structures* studied in the literature
  - Linear [Dani et al., 2008, Chu et al., 2011, Abbasi-Yadkori et al., 2011]
  - Combinatorial [Cesa-Bianchi and Lugosi, 2012]
  - Lipschitz [Magureanu et al., 2014]
  - ...
- Algorithms to exploit any **general structures** are desirable in practice
  - Learner knows a subset of realizable bandit problems
  - Pulling an arm provides **information** about the bandit problem itself...

Bandit Structures

- Several *specific structures* studied in the literature
  - Linear [Dani et al., 2008, Chu et al., 2011, Abbasi-Yadkori et al., 2011]
  - Combinatorial [Cesa-Bianchi and Lugosi, 2012]
  - Lipschitz [Magureanu et al., 2014]
  - ...
- Algorithms to exploit any **general structures** are desirable in practice
  - Learner knows a subset of realizable bandit problems
  - Pulling an arm provides **information** about the bandit problem itself...
  - ...which in turn provides information about all other arms

Bandit Structures

- Several *specific structures* studied in the literature
  - Linear [Dani et al., 2008, Chu et al., 2011, Abbasi-Yadkori et al., 2011]
  - Combinatorial [Cesa-Bianchi and Lugosi, 2012]
  - Lipschitz [Magureanu et al., 2014]
  - ...
- Algorithms to exploit any **general structures** are desirable in practice
  - Learner knows a subset of realizable bandit problems
  - Pulling an arm provides **information** about the bandit problem itself...
  - ...which in turn provides information about all other arms
  - Constant regret (independent of horizon) is possible in certain structures [Lattimore and Munos, 2014]

Dealing with General Structures

- **Confidence-based** algorithms

Dealing with General Structures

- **Confidence-based** algorithms
  - Choose arms based on *confidence intervals* of the true bandit problem

Dealing with General Structures

- **Confidence-based** algorithms
  - Choose arms based on *confidence intervals* of the true bandit problem
  - Typically extend unstructured optimistic strategies (e.g., UCB)

Dealing with General Structures

- **Confidence-based** algorithms
  - Choose arms based on *confidence intervals* of the true bandit problem
  - Typically extend unstructured optimistic strategies (e.g., UCB)
  - Good **finite-time** performance but not asymptotically optimal in general [Lattimore and Szepesvari, 2017]

Dealing with General Structures

- **Confidence-based** algorithms
  - Choose arms based on *confidence intervals* of the true bandit problem
  - Typically extend unstructured optimistic strategies (e.g., UCB)
  - Good **finite-time** performance but not asymptotically optimal in general [Lattimore and Szepesvari, 2017]
- Algorithms from asymptotic **problem-dependent lower bounds**

Dealing with General Structures

- **Confidence-based** algorithms
  - Choose arms based on *confidence intervals* of the true bandit problem
  - Typically extend unstructured optimistic strategies (e.g., UCB)
  - Good **finite-time** performance but not asymptotically optimal in general [Lattimore and Szepesvari, 2017]
- Algorithms from asymptotic **problem-dependent lower bounds**
  - Asymptotically optimal for general structures...

Dealing with General Structures

- **Confidence-based** algorithms
  - Choose arms based on *confidence intervals* of the true bandit problem
  - Typically extend unstructured optimistic strategies (e.g., UCB)
  - Good **finite-time** performance but not asymptotically optimal in general [Lattimore and Szepesvari, 2017]
- Algorithms from asymptotic **problem-dependent lower bounds**
  - Asymptotically optimal for general structures...
  - ...but require **forced-exploration**

Dealing with General Structures

- **Confidence-based** algorithms
  - Choose arms based on *confidence intervals* of the true bandit problem
  - Typically extend unstructured optimistic strategies (e.g., UCB)
  - Good **finite-time** performance but not asymptotically optimal in general [Lattimore and Szepesvari, 2017]
- Algorithms from asymptotic **problem-dependent lower bounds**
  - Asymptotically optimal for general structures...
  - ...but require **forced-exploration**
  - Open question how well they perform in finite time

Dealing with General Structures

- **Confidence-based** algorithms
  - Choose arms based on *confidence intervals* of the true bandit problem
  - Typically extend unstructured optimistic strategies (e.g., UCB)
  - Good **finite-time** performance but not asymptotically optimal in general [Lattimore and Szepesvari, 2017]
- Algorithms from asymptotic **problem-dependent lower bounds**
  - Asymptotically optimal for general structures...
  - ...but require **forced-exploration**
  - Open question how well they perform in finite time
  - Not easy to handle structures where constant regret is achievable

Our Contributions

1. A novel **confidence-based** algorithm for general structures

Our Contributions

1. A novel **confidence-based** algorithm for general structures
   - Pull count of a sub-optimal arm can be reduced using the information of other arms

Our Contributions

1. A novel **confidence-based** algorithm for general structures
   - Pull count of a sub-optimal arm can be reduced using the information of other arms
2. Our algorithm suffers **constant regret** in certain structures

Our Contributions

1. A novel **confidence-based** algorithm for general structures
   - Pull count of a sub-optimal arm can be reduced using the information of other arms
2. Our algorithm suffers **constant regret** in certain structures
3. A matching **finite-time lower bound** for these structures

Formulation

- **Structure** Θ: set of possible bandit *models*

Formulation

- **Structure** Θ: set of possible bandit *models*
- Mappings $\mu_i : \Theta \to \mathbb{R}$ from models to mean rewards

Formulation

- **Structure** Θ: set of possible bandit *models*
- Mappings $\mu_i : \Theta \to \mathbb{R}$ from models to mean rewards
- Θ, μ_i **known** to the learner, true model θ^* *unknown*

Formulation

- **Structure** Θ: set of possible bandit *models*
- Mappings $\mu_i : \Theta \to \mathbb{R}$ from models to mean rewards
- Θ, μ_i **known** to the learner, true model θ^* *unknown*
- **Goal** minimize cumulative regret

$$R_n^\pi(\theta^*, \Theta) := n\mu^*(\theta^*) - \mathbb{E}_{\pi,\theta^*}\left[\sum_{t=1}^{n} \mu_{I_t}(\theta^*)\right]$$

Structured Arm Elimination (SAE)

Input: Set of models Θ, horizon n

Set of active arms: $\widetilde{\mathcal{A}}_0 \leftarrow \mathcal{A}^*(\Theta)$

Threshold: $\widetilde{\Gamma}_0 \leftarrow 1$

Foreach phase $h = 0, 1, \dots$ **do**

Play all active arms until $\mathcal{O}\left(\frac{\log n}{\widetilde{\Gamma}_h^2}\right)$ pulls are reached for all $i \in \widetilde{A}_h$

Update confidence set: $\widetilde{\Theta}_{h+1} \leftarrow \left\{\theta \in \Theta \;\middle|\; \forall i \in \mathcal{A} : |\hat{\mu}_{i,h} - \mu_i(\theta)| < \sqrt{\frac{\alpha \log n}{T_i(h)}}\right\}$

Update set of active arms: $\widetilde{A}_{h+1} = \mathcal{A}^*(\widetilde{\Theta}_{h+1}) \cap \widetilde{A}_h$

Decrease threshold: $\widetilde{\Gamma}_{h+1} \leftarrow \frac{\widetilde{\Gamma}_h}{2}$

End

Structured Arm Elimination (SAE)

Input: Set of models Θ, horizon n
Set of active arms: $\widetilde{\mathcal{A}}_0 \leftarrow \mathcal{A}^*(\Theta)$
Threshold: $\widetilde{\Gamma}_0 \leftarrow 1$
Foreach phase $h = 0, 1, \dots$ **do**
  Play all active arms until $\mathcal{O}\left(\frac{\log n}{\widetilde{\Gamma}_h^2}\right)$ pulls are reached for all $i \in \widetilde{A}_h$
  Update confidence set: $\widetilde{\Theta}_{h+1} \leftarrow \left\{\theta \in \Theta \;\middle|\; \forall i \in \mathcal{A} : |\hat{\mu}_{i,h} - \mu_i(\theta)| < \sqrt{\frac{\alpha \log n}{T_i(h)}}\right\}$
  Update set of active arms: $\widetilde{A}_{h+1} = \mathcal{A}^*(\widetilde{\Theta}_{h+1}) \cap \widetilde{A}_h$
  Decrease threshold: $\widetilde{\Gamma}_{h+1} \leftarrow \frac{\widetilde{\Gamma}_h}{2}$
End

Structured Arm Elimination (SAE)

Input: Set of models Θ, horizon n
Set of active arms: $\widetilde{\mathcal{A}}_0 \leftarrow \mathcal{A}^*(\Theta)$
Threshold: $\widetilde{\Gamma}_0 \leftarrow 1$
Foreach phase $h = 0, 1, \dots$ **do**
  Play all active arms until $\mathcal{O}\left(\frac{\log n}{\widetilde{\Gamma}_h^2}\right)$ pulls are reached for all $i \in \widetilde{A}_h$
  Update confidence set: $\widetilde{\Theta}_{h+1} \leftarrow \left\{\theta \in \Theta \;\middle|\; \forall i \in \mathcal{A} : |\hat{\mu}_{i,h} - \mu_i(\theta)| < \sqrt{\frac{\alpha \log n}{T_i(h)}}\right\}$
  Update set of active arms: $\widetilde{A}_{h+1} = \mathcal{A}^*(\widetilde{\Theta}_{h+1}) \cap \widetilde{A}_h$
  Decrease threshold: $\widetilde{\Gamma}_{h+1} \leftarrow \frac{\widetilde{\Gamma}_h}{2}$
End

Structured Arm Elimination (SAE)

Input: Set of models Θ, horizon n
Set of active arms: $\widetilde{\mathcal{A}}_0 \leftarrow \mathcal{A}^*(\Theta)$
Threshold: $\widetilde{\Gamma}_0 \leftarrow 1$
Foreach phase $h = 0, 1, \dots$ **do**
  Play all active arms until $\mathcal{O}\left(\frac{\log n}{\widetilde{\Gamma}_h^2}\right)$ pulls are reached for all $i \in \widetilde{A}_h$
  Update confidence set: $\widetilde{\Theta}_{h+1} \leftarrow \left\{\theta \in \Theta \;\middle|\; \forall i \in \mathcal{A} : |\hat{\mu}_{i,h} - \mu_i(\theta)| < \sqrt{\frac{\alpha \log n}{T_i(h)}}\right\}$
  Update set of active arms: $\widetilde{A}_{h+1} = \mathcal{A}^*(\widetilde{\Theta}_{h+1}) \cap \widetilde{A}_h$
  Decrease threshold: $\widetilde{\Gamma}_{h+1} \leftarrow \frac{\widetilde{\Gamma}_h}{2}$
End

Structured Arm Elimination (SAE)

Input: Set of models Θ, horizon n
Set of active arms: $\widetilde{\mathcal{A}}_0 \leftarrow \mathcal{A}^*(\Theta)$
Threshold: $\widetilde{\Gamma}_0 \leftarrow 1$
Foreach phase $h = 0, 1, \dots$ **do**

Play all active arms until $\mathcal{O}\left(\frac{\log n}{\widetilde{\Gamma}_h^2}\right)$ pulls are reached for all $i \in \widetilde{A}_h$

Update confidence set: $\widetilde{\Theta}_{h+1} \leftarrow \left\{\theta \in \Theta \;\middle|\; \forall i \in \mathcal{A} : |\hat{\mu}_{i,h} - \mu_i(\theta)| < \sqrt{\frac{\alpha \log n}{T_i(h)}}\right\}$

Update set of active arms: $\widetilde{A}_{h+1} = \mathcal{A}^*(\widetilde{\Theta}_{h+1}) \cap \widetilde{A}_h$

Decrease threshold: $\widetilde{\Gamma}_{h+1} \leftarrow \frac{\widetilde{\Gamma}_h}{2}$

End

Structured Arm Elimination (SAE)

Input: Set of models Θ, horizon n
Set of active arms: $\widetilde{\mathcal{A}}_0 \leftarrow \mathcal{A}^*(\Theta)$
Threshold: $\widetilde{\Gamma}_0 \leftarrow 1$
Foreach phase $h = 0, 1, \dots$ **do**

Play all active arms until $\mathcal{O}\left(\frac{\log n}{\widetilde{\Gamma}_h^2}\right)$ pulls are reached for all $i \in \widetilde{A}_h$

Update confidence set: $\widetilde{\Theta}_{h+1} \leftarrow \left\{\theta \in \Theta \;\middle|\; \forall i \in \mathcal{A} : |\hat{\mu}_{i,h} - \mu_i(\theta)| < \sqrt{\frac{\alpha \log n}{T_i(h)}}\right\}$

Update set of active arms: $\widetilde{A}_{h+1} = \mathcal{A}^*(\widetilde{\Theta}_{h+1}) \cap \widetilde{A}_h$

Decrease threshold: $\widetilde{\Gamma}_{h+1} \leftarrow \frac{\widetilde{\Gamma}_h}{2}$

End

Structured Arm Elimination (SAE)

Input: Set of models Θ, horizon n
Set of active arms: $\widetilde{\mathcal{A}}_0 \leftarrow \mathcal{A}^*(\Theta)$
Threshold: $\widetilde{\Gamma}_0 \leftarrow 1$
Foreach phase $h = 0, 1, \dots$ **do**

Play all active arms until $\mathcal{O}\left(\frac{\log n}{\widetilde{\Gamma}_h^2}\right)$ pulls are reached for all $i \in \widetilde{A}_h$

Update confidence set: $\widetilde{\Theta}_{h+1} \leftarrow \left\{\theta \in \Theta \;\middle|\; \forall i \in \mathcal{A} : |\hat{\mu}_{i,h} - \mu_i(\theta)| < \sqrt{\frac{\alpha \log n}{T_i(h)}}\right\}$

Update set of active arms: $\widetilde{A}_{h+1} = \mathcal{A}^*(\widetilde{\Theta}_{h+1}) \cap \widetilde{A}_h$

Decrease threshold: $\widetilde{\Gamma}_{h+1} \leftarrow \frac{\widetilde{\Gamma}_h}{2}$

End

Structured Arm Elimination (SAE)

Input: Set of models Θ, horizon n
Set of active arms: $\widetilde{\mathcal{A}}_0 \leftarrow \mathcal{A}^*(\Theta)$
Threshold: $\widetilde{\Gamma}_0 \leftarrow 1$
Foreach phase $h = 0, 1, \dots$ **do**

Play all active arms until $\mathcal{O}\left(\frac{\log n}{\widetilde{\Gamma}_h^2}\right)$ pulls are reached for all $i \in \widetilde{A}_h$

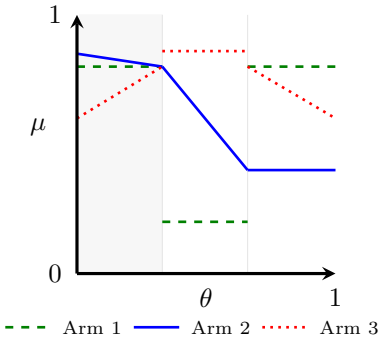
Update confidence set: $\widetilde{\Theta}_{h+1} \leftarrow \left\{\theta \in \Theta \;\middle|\; \forall i \in \mathcal{A} : |\hat{\mu}_{i,h} - \mu_i(\theta)| < \sqrt{\frac{\alpha \log n}{T_i(h)}}\right\}$

Update set of active arms: $\widetilde{A}_{h+1} = \mathcal{A}^*(\widetilde{\Theta}_{h+1}) \cap \widetilde{A}_h$

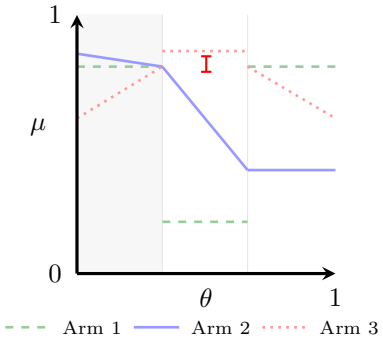
Decrease threshold: $\widetilde{\Gamma}_{h+1} \leftarrow \frac{\widetilde{\Gamma}_h}{2}$

End

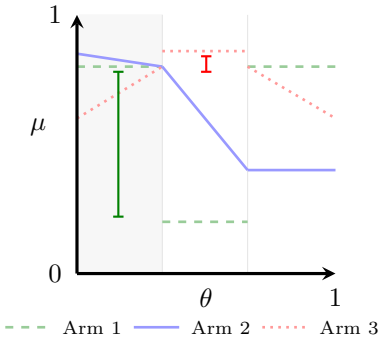
Example



Example



Example



Main Result

Theorem (Regret bound for SAE)

Let $\mathcal{A}_i^*$ be the set of arms that can be guaranteed to be active in the phase in which i is discarded, then

$$R_n^{\mathrm{SAE}}(\theta^*, \Theta) \leq \sum_{i \in \mathcal{A}^*(\Theta) \setminus \{i^*\}} \frac{c\Delta_i(\theta^*) \log n}{\inf_{\theta \in \Theta_i^*} \max_{j \in \mathcal{A}_i^*} |\mu_j(\theta^*) - \mu_j(\theta)|^2} + 2|\mathcal{A}^*(\Theta)|$$

Main Result

Theorem (Regret bound for SAE)

Let $\mathcal{A}_i^*$ be the set of arms that can be guaranteed to be active in the phase in which i is discarded, then

$$R_n^{\mathrm{SAE}}(\theta^*, \Theta) \leq \sum_{i \in \mathcal{A}^*(\Theta) \setminus \{i^*\}} \frac{c\Delta_i(\theta^*) \log n}{\inf_{\theta \in \Theta_i^*} \max_{j \in \mathcal{A}_i^*} |\mu_j(\theta^*) - \mu_j(\theta)|^2} + 2|\mathcal{A}^*(\Theta)|$$

- Scales with number of arms that are optimal in at least one model

Main Result

Theorem (Regret bound for SAE)

Let $\mathcal{A}_i^*$ be the set of arms that can be guaranteed to be active in the phase in which i is discarded, then

$$R_n^{\mathrm{SAE}}(\theta^*, \Theta) \leq \sum_{i \in \mathcal{A}^*(\Theta) \setminus \{i^*\}} \frac{c\Delta_i(\theta^*) \log n}{\inf_{\theta \in \Theta_i^*} \max_{j \in \mathcal{A}_i^*} |\mu_j(\theta^*) - \mu_j(\theta)|^2} + 2|\mathcal{A}^*(\Theta)|.$$

- Scales with number of arms that are optimal in at least one model
- Scales with **model gaps** instead of arm gaps

Main Result

Theorem (Regret bound for SAE)

Let $\mathcal{A}_i^*$ be the set of arms that can be guaranteed to be active in the phase in which i is discarded, then

$$R_n^{\mathrm{SAE}}(\theta^*, \Theta) \leq \sum_{i \in \mathcal{A}^*(\Theta) \setminus \{i^*\}} \frac{c\Delta_i(\theta^*) \log n}{\inf_{\theta \in \Theta_i^*} \max_{j \in \mathcal{A}_i^*} |\mu_j(\theta^*) - \mu_j(\theta)|^2} + 2|\mathcal{A}^*(\Theta)|.$$

- Scales with number of arms that are optimal in at least one model
- Scales with **model gaps** instead of arm gaps
- In order to figure out that arm i is sub-optimal

Main Result

Theorem (Regret bound for SAE)

Let $\mathcal{A}_i^*$ be the set of arms that can be guaranteed to be active in the phase in which i is discarded, then

$$R_n^{\mathrm{SAE}}(\theta^*, \Theta) \leq \sum_{i \in \mathcal{A}^*(\Theta) \setminus \{i^*\}} \frac{c\Delta_i(\theta^*) \log n}{\inf_{\theta \in \Theta_i^*} \max_{j \in \mathcal{A}_i^*} |\mu_j(\theta^*) - \mu_j(\theta)|^2} + 2|\mathcal{A}^*(\Theta)|.$$

- Scales with number of arms that are optimal in at least one model
- Scales with **model gaps** instead of arm gaps
- In order to figure out that arm i is sub-optimal
  - Discard all models in which i is optimal...

Main Result

Theorem (Regret bound for SAE)

Let $\mathcal{A}_i^*$ be the set of arms that can be guaranteed to be active in the phase in which i is discarded, then

$$R_n^{\mathrm{SAE}}(\theta^*, \Theta) \leq \sum_{i \in \mathcal{A}^*(\Theta) \setminus \{i^*\}} \frac{c\Delta_i(\theta^*) \log n}{\inf_{\theta \in \Theta_i^*} \max_{j \in \mathcal{A}_i^*} |\mu_j(\theta^*) - \mu_j(\theta)|^2} + 2|\mathcal{A}^*(\Theta)|.$$

- Scales with number of arms that are optimal in at least one model
- Scales with **model gaps** instead of arm gaps
- In order to figure out that arm i is sub-optimal
  - Discard all models in which i is optimal...
  - ...by pulling the **most informative** active arm (the one with largest model gap)

Comparison with Unstructured Algorithms

Theorem (SAE is sub-UCB)

There exist constant $c, c' > 0$ such that

$$R_n^{\mathrm{SAE}}(\theta^*, \Theta) \leq \sum_{i \in \mathcal{A} \setminus \{i^*\}} \frac{c \log n}{\Delta_i(\theta^*)} + c'$$

Comparison with Unstructured Algorithms

Theorem (SAE is sub-UCB)

There exist constant $c, c' > 0$ such that

$$R_n^{\mathrm{SAE}}(\theta^*, \Theta) \leq \sum_{i \in \mathcal{A} \setminus \{i^*\}} \frac{c \log n}{\Delta_i(\theta^*)} + c'$$

- SAE reduces to Improved UCB [Auer and Ortner, 2010] in the unstructured case...

Comparison with Unstructured Algorithms

Theorem (SAE is sub-UCB)

There exist constant $c, c' > 0$ such that

$$R_n^{\mathrm{SAE}}(\theta^*, \Theta) \leq \sum_{i \in \mathcal{A} \setminus \{i^*\}} \frac{c \log n}{\Delta_i(\theta^*)} + c'$$

- SAE reduces to Improved UCB [Auer and Ortner, 2010] in the unstructured case...
- ...but SAE is not optimistic in general

Comparison with Structured UCB

- The most related algorithm is *Structured UCB* (SUCB)
  [Azar et al., 2013, Lattimore and Munos, 2014]

Comparison with Structured UCB

- The most related algorithm is *Structured UCB* (SUCB)
  [Azar et al., 2013, Lattimore and Munos, 2014]
- Regret analysis of SUCB reveals that the algorithm pulls a sub-optimal arm proportionally to its model gaps

Comparison with Structured UCB

- The most related algorithm is *Structured UCB* (SUCB)
  [Azar et al., 2013, Lattimore and Munos, 2014]
- Regret analysis of SUCB reveals that the algorithm pulls a sub-optimal arm proportionally to its model gaps
- SAE is **sub-SUCB** in many structures

Towards Constant Regret

- An anytime version of SAE (ASAE) suffers constant regret in certain structures

Towards Constant Regret

- An anytime version of SAE (ASAE) suffers constant regret in certain structures

Assumption ([Lattimore and Munos, 2014])

The structure Θ is such that

$$\Gamma_* := \inf_{\theta \in \Theta \setminus \Theta^*_{i^*}} |\mu_{i^*}(\theta^*) - \mu_{i^*}(\theta)| > 0$$

Constant Regret

Theorem (Constant-regret Bound)

$$R_n^{\mathrm{ASAE}}(\theta^*, \Theta) \leq \sum_{i \in \mathcal{A}^*(\Theta) \setminus \{i^*\}} \frac{c\Delta_i(\theta^*)\log(1/\Gamma_*)}{\inf_{\theta \in \Theta_i^*} \max_{j \in \{i, i^*\}} |\mu_j(\theta^*) - \mu_j(\theta)|^2} + 9|\mathcal{A}^*(\Theta)|.$$

Constant Regret

Theorem (Constant-regret Bound)

$$R_n^{\mathrm{ASAE}}(\theta^*, \Theta) \leq \sum_{i \in \mathcal{A}^*(\Theta) \setminus \{i^*\}} \frac{c\Delta_i(\theta^*)\log(1/\Gamma_*)}{\inf_{\theta \in \Theta_i^*} \max_{j \in \{i, i^*\}} |\mu_j(\theta^*) - \mu_j(\theta)|^2} + 9|\mathcal{A}^*(\Theta)|.$$

- Finite-time lower bound reveals that $\mathcal{O}(\log(1/\Gamma_*))$ dependence is tight

Constant Regret

Theorem (Constant-regret Bound)

$$R_n^{\mathrm{ASAE}}(\theta^*, \Theta) \leq \sum_{i \in \mathcal{A}^*(\Theta) \setminus \{i^*\}} \frac{c\Delta_i(\theta^*)\log(1/\Gamma_*)}{\inf_{\theta \in \Theta_i^*} \max_{j \in \{i, i^*\}} |\mu_j(\theta^*) - \mu_j(\theta)|^2} + 9|\mathcal{A}^*(\Theta)|.$$

- Finite-time lower bound reveals that $\mathcal{O}(\log(1/\Gamma_*))$ dependence is tight
- SUCB suffers $\mathcal{O}(\log(1/\min\{\Gamma_*, \Delta_{\min}\}))$

Constant Regret

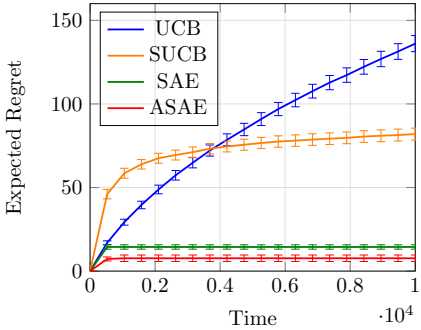
Theorem (Constant-regret Bound)

$$R_n^{\mathrm{ASAE}}(\theta^*, \Theta) \leq \sum_{i \in \mathcal{A}^*(\Theta) \setminus \{i^*\}} \frac{c\Delta_i(\theta^*)\log(1/\Gamma_*)}{\inf_{\theta \in \Theta_i^*} \max_{j \in \{i, i^*\}} |\mu_j(\theta^*) - \mu_j(\theta)|^2} + 9|\mathcal{A}^*(\Theta)|.$$

- Finite-time lower bound reveals that $\mathcal{O}(\log(1/\Gamma_*))$ dependence is tight
- SUCB suffers $\mathcal{O}(\log(1/\min\{\Gamma_*, \Delta_{\min}\}))$
- Asymptotically-optimal algorithms based on forced-exploration suffer $\mathcal{O}(1/\Gamma_*)$

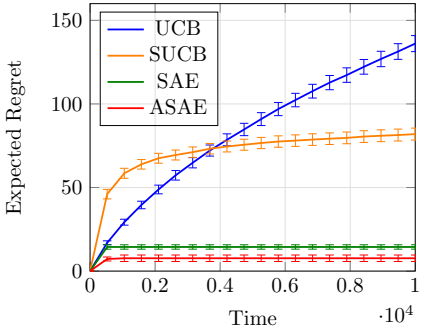
Experiments

Non-optimistic informative arm

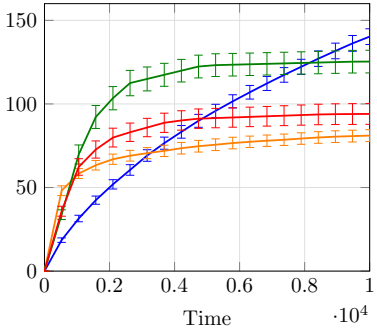


Experiments

Non-optimistic informative arm



Optimism is optimal



Conclusion

- SAE confirms that simple confidence-based strategies can be designed to exploit general structures with good finite-time performance

Conclusion

- SAE confirms that simple confidence-based strategies can be designed to exploit general structures with good finite-time performance
- Our analysis provides insights on the potential benefits of pulling informative arms

Conclusion

- SAE confirms that simple confidence-based strategies can be designed to exploit general structures with good finite-time performance
- Our analysis provides insights on the potential benefits of pulling informative arms
- SAE is not optimistic, a key step towards building finite-time optimal strategies

Thank you!

andrea.tirinzoni@polimi.it

https://andreatirinzoni.github.io

References

Abbasi-Yadkori, Y., Pál, D., and Szepesvári, C. (2011).
Improved algorithms for linear stochastic bandits.
In *Advances in Neural Information Processing Systems*, pages 2312–2320.

Auer, P. and Ortner, R. (2010).
Ucb revisited: Improved regret bounds for the stochastic multi-armed bandit problem.
Periodica Mathematica Hungarica, 61(1-2):55–65.

Azar, M., Lazaric, A., and Brunskill, E. (2013).
Sequential transfer in multi-armed bandit with finite set of models.
In Burges, C. J. C., Bottou, L., Welling, M., Ghahramani, Z., and Weinberger, K. Q., editors, *Advances in Neural Information Processing Systems 26*, pages 2220–2228.

Cesa-Bianchi, N. and Lugosi, G. (2012).
Combinatorial bandits.
Journal of Computer and System Sciences, 78(5):1404–1422.

References (cont.)

Chu, W., Li, L., Reyzin, L., and Schapire, R. (2011).
Contextual bandits with linear payoff functions.
In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pages 208–214.

Dani, V., Hayes, T. P., and Kakade, S. M. (2008).
Stochastic linear optimization under bandit feedback.

Lattimore, T. and Munos, R. (2014).
Bounded regret for finite-armed structured bandits.
In *Advances in Neural Information Processing Systems*, pages 550–558.

Lattimore, T. and Szepesvari, C. (2017).
The end of optimism? an asymptotic analysis of finite-armed linear bandits.
In *Artificial Intelligence and Statistics*, pages 728–737.

Magureanu, S., Combes, R., and Proutiere, A. (2014).
Lipschitz bandits: Regret lower bounds and optimal algorithms.
arXiv preprint arXiv:1405.4758.