Multi-Agent Thompson Sampling for Bandits with Sparse Neighborhood Structures

Timothy Verstraeten  Eugenio Bargiacchi  Pieter Libin  Jan Helsen  Diederik Roijers  Ann Nowé

Multi-Agent Multi-Armed Bandits
➢ Groups of agents
➢ Factored reward function
➢ Coordination problem

Multi-Agent Thompson Sampling (MATS)

Agents          Local joint actions          Sampling          Action selection

1) Variable Elimination

\[ a^* = \arg\max_{a} \mu_{ag}^g + \mu_{ah}^h + \ldots \]

2) Pull global joint action \( a^* \)

e.g., 0 1 0/0 1 ...

Theorem: For subgaussian rewards with bounded means, Bayesian regret of MATS is

\[ O\left(\sqrt{\hat{A}T \log(\hat{A}T)}\right) \]

where \( \hat{A} \) is the number of local joint actions.

For sparse graphs, regret is low-order polynomial in single agent’s actions.

Experiments
➢ Three synthetic benchmarks
➢ Wind farm control task:
   ➢ Align rotors to maximize power production

Contact
➢ E-mail: tiverstr@vub.be
➢ Page: https://ai.vub.ac.be/team/timothy-verstraeten/
➢ Affiliation:
  Artificial Intelligence Lab, Vrije Universiteit Brussel, Brussels, Belgium

Artificial icons created by Deemak Daksina from the Noun Project