Directed exploration with Bayes-adaptive replay

Georgy Antonov
Department of Computational Neuroscience
Max Planck Institute for Biological Cybernetics
72076 Tübingen, Germany

Graduate Training Centre of Neuroscience
International Max Planck Research School
University of Tübingen
72076 Tübingen, Germany
georgy.antonov@tuebingen.mpg.de

Peter Dayan
Department of Computational Neuroscience
Max Planck Institute for Biological Cybernetics
72076 Tübingen, Germany

University of Tübingen
72074 Tübingen, Germany
dayan@tue.mpg.de

Abstract

The *title* should be a maximum of 100 characters.

The *abstract* should be a maximum of 2000 characters of text, including spaces (no figure is allowed). You will be asked to copy this into a text-only box; and it will appear as such in the conference booklet. Use 11 point type, with a vertical spacing of 12 points. The word **Abstract** must be centered, bold, and in point size 12. Two line spaces precede the abstract.

Keywords: Reinforcement learning, DYNA, planning, exploration, replay

Acknowledgements

Put something here.

1 Plan

- Introduction
  - What is replay
  - DYNA; prioritised sweeping; M&D – replay as planning?
  - Planning and explore-exploit; the lack of exploration in M&D
  - Exploration in Bayesian bandits; Gittins indices
  - MCTS/BAMCP; Prioritised sweeping in belief trees, Bayes-adaptive replay
- Results
  - Prioritised sweeping in Bayesian bandits
    - Expected value of a backup – probabilistic need
    - Tree policy and empirical convergence to optimal values
  - Bayes-adaptive optimised replay
    - Expected value of a backup – joint belief and state dynamics (information states)
    - We should say something about forgetting here
    - (Estimated) information gain is subsumed by M&D's Gain
    - How information-augmented Gain interacts with Need (example replay in a maze)
- Discussion