Incremental Policy Iteration with Guaranteed Escape from Local Optima in POMDP Planning

Marek Grzes
School of Computing, University of Kent, UK
Pascal Poupart
David Cheriton School of Computer Science, University of Waterloo, Canada

AAMAS-15

Motivation and Contribution
- Finite-state controllers (FSCs) are the most energy efficient POMDP policies (see Grzes et al. “Energy Efficient Execution of POMDP Policies.”) which shows their suitability for mobile applications
- Efficient and robust algorithms that compute small policies/controllers become desirable
- We investigate incremental methods that guarantee the escape from local optima
- We push the understanding and the performance of policy iteration for POMDPs to the point that for the first time they are competitive with the state-of-the-art point-based methods

Finite-state Controllers for POMDPs

Node Improvement in Bounded Policy Iteration (BPI)

The Need to Escape Local Optima

Find a New Way to Compute the Exact DP Update for POMDPs

Optimal Solution to the Escape Problem

Theorem
There always exists an optimal solution to the quadratic problem shown in Fig. 3 that is integral \( V(n',a) \), i.e., there exists an optimal solution that corresponds to a deterministic node.

Our Algorithm

![Figure 4](http://www.cs.kent.ac.uk/people/staff/mg483/)
Thanks to the above theorem, McCormick relaxation finds an optimal, deterministic node
- MILP is intractable, but we don’t need the optimal solution
- Even a linear relaxation of our MILP can be sufficient (see the paper for interesting properties)

Practical Implementation with Fast Heuristics

Results

![Figure 5](http://www.cs.kent.ac.uk/people/staff/mg483/)
- A new view on principled methods for policy iteration in POMDPs
- A new efficient method for improving individual nodes
- An intuitive explanation of local optima and challenges in escaping it
- A guaranteed method for escape that facilitates fast, anytime execution
- The best node for escape is deterministic
- Heuristic methods analysed (with new connections identified) and used in a practical and well-justified manner

Acknowledgements
This research was sponsored by NSERC and MITACS.

e-mail: m.grzes@kent.ac.uk and ppoupart@waterloo.ca

www: http://www.cs.kent.ac.uk/people/staff/mg483/