Online Skill Discovery using Graph-based Clustering

Jan Hendrik Metzen
Universität Bremen, AG Robotik

Contact: jhm@informatik.uni-bremen.de
More information: http://www.informatik.uni-bremen.de/~jhm/
Motivation: Skill Discovery

- Hierarchical RL aims at
  - **decomposing** a complex problem into simpler subproblems
  - learn solutions (skills) for these subproblems
  - and learn to solve the greater problem using the skills

---

**Online Skill Discovery using Graph-based Clustering**

July 1st, 2012 (EWRL 2012)
Motivation: Skill Discovery

- Hierarchical RL aims at
  - **decomposing** a complex problem into simpler subproblems
  - learn solutions (**skills**) for these subproblems
  - and learn to solve the greater problem using the skills

- Skills can allow to
  - focus **exploration**
  - increase **reusability/transferability** between related tasks
  - increase **representability** of value functions/policies

[Konidaris and Barto, 2009]
Motivation: Skill Discovery

- Hierarchical RL aims at
  - decomposing a complex problem into simpler subproblems
  - learn solutions (skills) for these subproblems
  - and learn to solve the greater problem using the skills

- Skills can allow to
  - focus exploration
  - increase reusability/transferability between related tasks
  - increase representability of value functions/policies
    [Konidaris and Barto, 2009]

- Agent should identify useful skills autonomously while acting (skill discovery)

- Approaches based on the transition graph (representing $P_{ss'}^a$) allow to identify skills based on domain characteristics
Prior work: Dynamic abstraction in reinforcement learning via clustering [Mannor et al., 2004]

- Estimate transition graph from observed transitions
- Cluster graph nodes with agglomerative hierarchical clustering
- Specific linkage criterion s.t. cluster correspond to densely connected subgraphs
- Learn skills that transition from one cluster to an adjacent one
Prior work: Dynamic abstraction in reinforcement learning via clustering [Mannor et al., 2004]

- Estimate transition graph from observed transitions
- Cluster graph nodes with agglomerative hierarchical clustering
- Specific linkage criterion s.t. cluster correspond to densely connected subgraphs
- Learn skills that transition from one cluster to an adjacent one
- Disadvantage: Can be conducted only once during learning.
Online Graph-based Agglomerative Hierarchical Clustering

- OGAHC allows to perform skill discovery incrementally, i.e. several times during learning

Smoothing: Add pseudo transitions in under-explored parts of the graph. "Dense local connectivity in the face of uncertainty"

Clusterings have to obey constraints to be consistent with prior clusterings.

True Transition Graph (ψ = −0.15)
Sample Transition Graph (m = 1000)
Smoothed Sample Transition Graph (ρ = 5)
Smoothed Sample Transition Graph (ρ = 15)
Online Graph-based Agglomerative Hierarchical Clustering

- OGAHC allows to perform skill discovery incrementally, i.e. several times during learning
- Smoothing: Add pseudo transitions in under-explored parts of the graph. “Dense local connectivity in the face of uncertainty”
Online Graph-based Agglomerative Hierarchical Clustering

- OGAHC allows to perform skill discovery incrementally, i.e. several times during learning

- Smoothing: Add pseudo transitions in under-explored parts of the graph. “Dense local connectivity in the face of uncertainty”

- Clusterings have to obey constraints to be consistent with prior clusterings
Results: Online vs. Offline Clustering

- Experiment on 50 random graphs with 400 states and a ground-truth clustering consisting of 7 clusters
- Accordance ratio: agreement of determined clusters with ground truth
- Large values of $\rho$ result in more added pseudo transitions
Results: Comparison to Related Work

- Experiment in $23 \times 23$ maze world consisting of 4 rooms and 12 tasks
- Baselines: No skill discovery and skill prototypes based on ground-truth clustering
- Comparison to Offline Clustering [Mannor et al., 2004] and Local Graph Partitioning [Şimşek et al., 2005]
Results: Comparison to Related Work

- Experiment in $23 \times 23$ maze world consisting of 4 rooms and 12 tasks
- Baselines: No skill discovery and skill prototypes based on ground-truth clustering
- Comparison to Offline Clustering [Mannor et al., 2004] and Local Graph Partitioning [Şimşek et al., 2005]
- OGAHC superior to other approaches and on a par with predefined skills for large range of $\rho$
Summary and Future Work

- Summary: We have presented a novel graph-based skill discovery method that allows to discover skills incrementally during learning

Future work:
1. Other smoothing heuristics, other graph clustering approaches
2. Extension to domains with large and/or continuous state space
3. Analysis of computational complexity

Thank you for your attention! Do you have questions?
Summary and Future Work

- **Summary:** We have presented a novel graph-based skill discovery method that allows to discover skills **incrementally** during learning.

- Furthermore in the paper: proposal of a novel linkage criterion which is off-policy and can deal with stochastic environments.

Future work:
1. Other smoothing heuristics, other graph clustering approaches
2. Extension to domains with large and/or continuous state space
3. Analysis of computational complexity

Thank you for your attention!
Summary and Future Work

Summary: We have presented a novel graph-based skill discovery method that allows to discover skills incrementally during learning.

Furthermore in the paper: proposal of a novel linkage criterion which is off-policy and can deal with stochastic environments.

Future work:

1. Other smoothing heuristics, other graph clustering approaches
2. Extension to domains with large and/or continuous state space
3. Analysis of computational complexity
Summary and Future Work

- **Summary**: We have presented a novel graph-based skill discovery method that allows to discover skills *incrementally* during learning.
- **Furthermore in the paper**: proposal of a novel linkage criterion which is off-policy and can deal with stochastic environments.
- **Future work**:
  1. Other smoothing heuristics, other graph clustering approaches
  2. Extension to domains with large and/or continuous state space
  3. Analysis of computational complexity

Thank you for your attention!
Do you have questions?
References


Scenario: Random Graphs
Results: Linkage Criteria

\[ c_x(A, B) = \sum_{e \in E \cap (A \times B)} w_x(e) \]

\[ M(A, B) = \frac{\text{min}\{|A|, |B|\} \log(\text{max}\{|A|, |B|\})}{c_{\text{uni}}(A, B) + c_{\text{uni}}(B, A)} \], [Mannor et al., 2004]

\[ \hat{N}_{\text{cut}}(A, B) = \frac{c(A, B) + c(B, A)}{c(A, V) + c(B, A)} + \frac{c(B, A) + c(A, B)}{c(B, V) + c(A, B)} \], [Şimşek et al., 2005]
Results: OGAHC

![Graph showing transitions and num clusters vs. offline ratio for different values of ρ.](image)
Scenario: Multi-task Maze
Results: Learning curves

1: Comparison

Method
- OGAHC ($\rho = 10$)
- LGP ($t_0 = 15$)
- Offline (Steps w/o novelty: 30000)
- No
- Predefined

2: Comparison

3: OGAHC

4: LGP

5: Offline

Steps without novelty
- 01000
- 30000
- 50000