Discretising Continuous Action Spaces for Optimal Decision Trees

Verifiable Policies for Continuous Environments in Reinforcement Learning

1. Background

- Desire for verifiability and interpretability in reinforcement learning (RL)
- Decision trees (DT) have these attributes
- Broccoli is a state-of-the-art search algorithm for finding optimal decision trees (ODT)
- Only works for discrete action spaces (DAS).
- ODTs for continuous action spaces (CAS) are still part of the research gap.
- This study focuses on discretising CAS to use the Broccoli algorithm for continuous environments.

Research Question

*How do different **discretisation** techniques of **continuous action** spaces affect the performance of the **Broccoli** algorithm with deterministic **black-box** environments?*

2. Methodology

- The Broccoli algorithm will be used to evaluate the discretisations.
- The environments used will be the classic control environments from the Gymnasium benchmark.
- Each run will be done with the same parameters for the specific environment.

3. Experiments

3.1 Uniform Discretisation

- CAS discretisation depends on the ability to approximate CAS.
- More actions result in a better approximation.
- Not yet investigated as Broccoli was used on constant action spaces.
- Multiple uniform action sets of varying sizes are compared
- Results in Fig. 1 show that pruning performs well under increasing action space size.
- Despite effectiveness, runtime still scales exponentially.

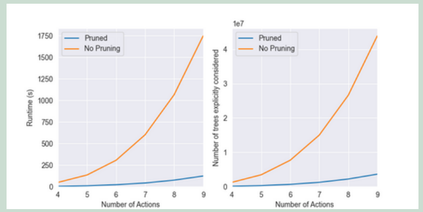

Figure 1: Searching with and without the trace-based pruning

3.2 Geometric Discretisation

- Geometric discretisation is used in poker-playing agents
- Possible to investigate the effect of action values on the pruning technique due to skewness

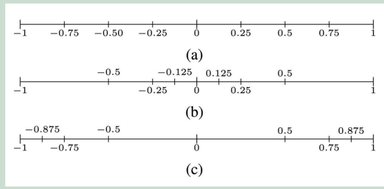

Figure 2: The Uniform (a), Geometric Progression (b), and Geometric Recession (c) discretisation.

- No consistent winner
- Possibly due to large overlap
- Uniform Discretisation can be considered a stable baseline
- Shows evidence for the effect of action values on pruning technique

3.3 Cheap Tree Sampling

- Uses cheaper sample trees to heuristically find good actions.
- Cheaper trees either consider fewer actions, fewer predicates, or less depth.
- Using these actions provide lower bound on performance of tree.

$$r_{t_{optimal}} \geqslant max_{t \in T_{sample}} r_t$$

Equation 1: Lower bound for final optimal tree given sample tree actions

- The sampling method increased actions considered without exponential search space increase.
- Action space decrease is around 70-80% by filtering out sub-optimal actions.
- No great performance gain on sample trees but outstanding final performance.

3.4 Twin-Delayed DDPG

- Twin-Delayed Deep Deterministic Policy Gradient (TD3) was taken as a teacher model
- Actions were sampled from the policy in the environment

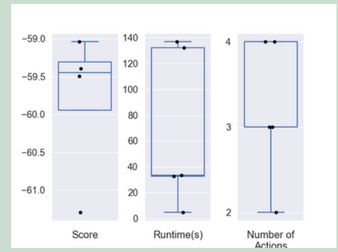

Figure 3: TD3-sampling tree performance for the Pendulum environment

- Results for Mountain Car Continuous (MCC) and Cart Pole Continuous were not good compared to other experiments
- Reduced runtime a lot for Pendulum environment and still had great perfomance.

Table 1: Best results for the MCC environment. Environment differed for the TD3 discretisation

Discretisation	Runtime (s)	Score
Static	122.6	98.58
Cheap sampling	448.2	**98.79**
TD3 (No training)	**14.8**	94.56

4. Conclusion

- Effect of increasing action spaces for Broccoli was analysed.
- Evidence for the effect of action values on trace-based pruning is shown.
- A lower bound was given for the performance of the cheap sampling discretisation
- Methods for both runtime reduction and performance increase were proposed.
- Small, well-performing ODTs were found for CAS RL environments.

5. Limitations and Future Work

Limitations

- Discretisation artificially symmetrical.
- TD3 hyperparameters violated the black-box constraint.
- MCC TD3 did not use the same environment as the other experiments.

Future work

- Sub-optimal tree sampling.
- Predicate and action space size trade-off.
- Tree sampling with different trees
- Warm-start searching using sample trees

MART VAN DER KUIL (505769) M.M.J.VANDERKUIL@STUDENT.TUDELFT.NL