Jonathan Simon Wols
j.s.wols@student.tudelft.nl
Supervisor: Anna Lukina

Interpretability and performance comparisons of decision tree surrogate models produced by AggreVaTe

1. Introduction

Imitation learning trains a policy which attempts to imitate the expert. When using imitation learning in critical decision-making processes, it is necessary for these policies to be interpretable.

We can train an interpretable surrogate model using imitation learning from the expert black boxes to obtain the required policies.

How do **decision trees,** produced by the **AggreVaTe**[1] algorithm, compare to a **Behavioral Cloning** baseline**,** and other **imitation learning** algorithms in terms of **interpretability** and **performance**?

The other imitation learning algorithms are:
GAIL[2], **Viper**[3] and **DAgger**[4]

2. Methodology

- AggreVaTe is an imitation learning algorithm that uses cost-to-go to obtain better long-term data
- The expert calculates, for each possible action, the cost to get to the finish when performing that action

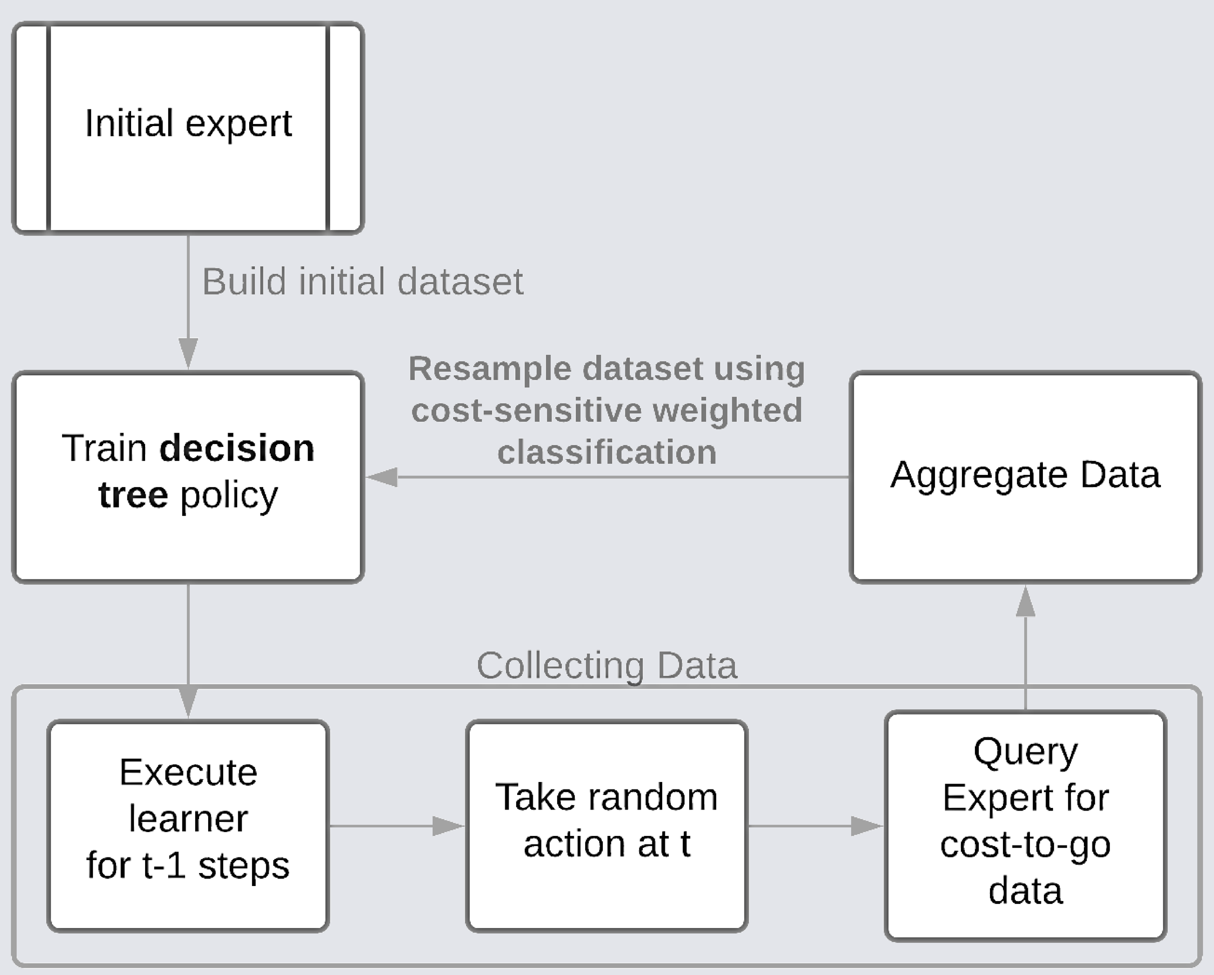

Figure 1: AggreVaTe algorithm learning process

- AggreVaTe requires a high number of training iterations or data collection rollouts due to obtaining only one datapoint per rollout
- The AggreVaTe algorithm has been modified to train decision tree policies to make it possible to compare on interpretability

3. Experiments

Setup

- 3 Open AI Gym Environments
- Fixed parameters
  - max depth of tree
  - number of iterations
  - rollouts per iteration

Policy

- A decision tree model (figure 2)
- Shows all decisions made
- Usable to find an explanation for difference in performance

Results

Comparison using the metrics:
- Average reward
- Standard deviation
- Number of nodes

The outcomes can be found in table 1, 2 and 3, where the best are in bold.

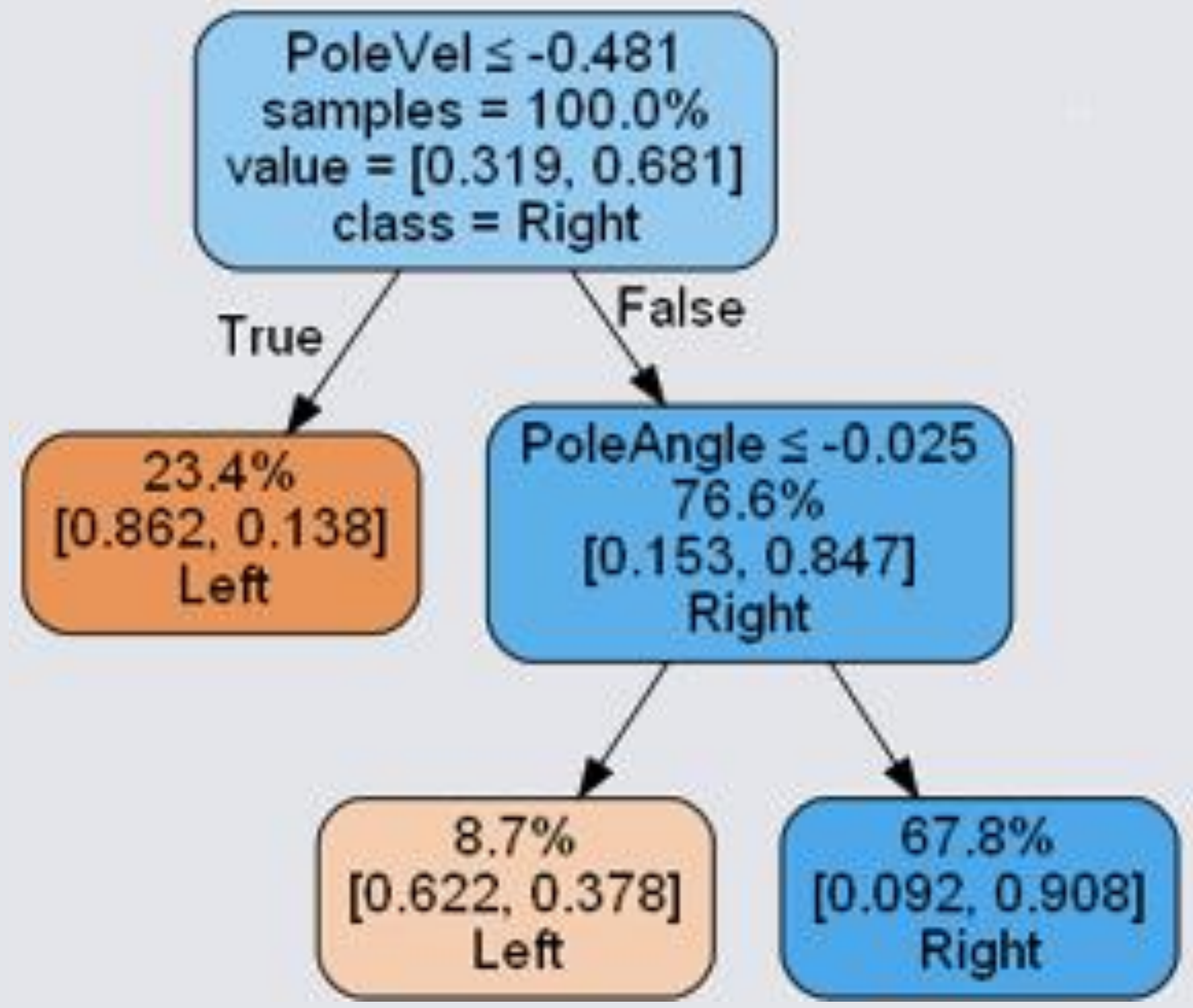

Figure 2: AggreVaTe algorithm on Cartpole

Mountain Car	BC	DAgger	GAIl	Viper	AggreVaTe
Average Reward	-120.8	-120.3	**-119.2**	-119.6	-120.4
Standard deviation	4.0	**3.6**	3.8	3.7	4.5
Nodes	3	3	3	3	3

Table 1: Results for mountain car with depth 1

Cartpole	BC	DAgger	GAIl	Viper	AggreVaTe
Average Reward	207.2	499.8	498.2	**500**	488.4
Standard deviation	44.2	0.7	14.12	**0**	22.0
Nodes	9	7	11	15(9)	**5**

Table 2: Results for cartpole with depth 3

Acrobot	BC	DAgger	GAIl	Viper	AggreVaTe
Average Reward	-94.5	-83.5	**-83.0**	-84.4	-85.5
Standard deviation	37.3	**10.8**	14.9	21.0	24.5
Nodes	5	5	7	7	**3**

Table 3: Results for acrobot with depth 2

4.Conclusion

In terms of interpretability, AggreVaTe performs equal or better than all other imitation learning algorithm.

In terms of performance, AggreVaTe performs only slightly worse than GAIL, Viper and DAgger, however it performs better than Behavioral Cloning.

This is consistent with expectations since fewer data points lead to more possible failing paths that have not yet been explored, but it also leads to a decision tree less prone to overfitting.

5. Limitations

- Experiments were on simple environments
- Experiments were performed with different experts
- The high number of rollouts required could have caused overfitting of the other algorithms.

References

[1] Stephane Ross and J Andrew Bagnell. Reinforcement and imitation learning via interactive. no-regret learning. arXiv preprint arXiv:1406.5979, 2014.
[2] Jonathan Ho and Stefano Ermon. Generative adversarial imitation learning. Advances in neural information processing systems, 29:4565–4573, 2016.
[3] Osbert Bastani, Yewen Pu, and Armando Solar-Lezama. Verifiable reinforcement learning via policy extraction. arXiv preprint arXiv:1805.08328, 2018.
[4] Stephane Ross, Geoffrey Gordon, and Drew Bagnell. A reduction of imitation learning and structured prediction to no-regret online learning. In Proceedings of the fourteenth international conference on artificial intelligence and statistics, pages 627–635. JMLR Workshop and Conference Proceedings, 2011.