Sequential Classical Control

Final Project CS 780

Brendan Crowe

TABLE OF CONTENTS

© Brendan Crowe

Imitation Learning

General Definition: Imitation learning (IL) techniques aim to mimic human behavior in a given task. An agent (a learning machine) is trained to perform a task from demonstrations by learning a mapping between observations (states) and actions.[3]

© Brendan Crowe

Imitation Learning Paradigms

- Behavior Cloning (BC):
  Methods learn a mapping from states to actions as a supervised learning problem [5]

- Inverse Reinforcement Learning (IRL):
  Attempt to recover the reward function the agent is trying to optimize. Then optimize that reward function.

© Brendan Crowe

Motivation

- Training robots or control systems
- Less need for domain knowledge
- Some tasks are very challenging to program, but can be learned

© Brendan Crowe

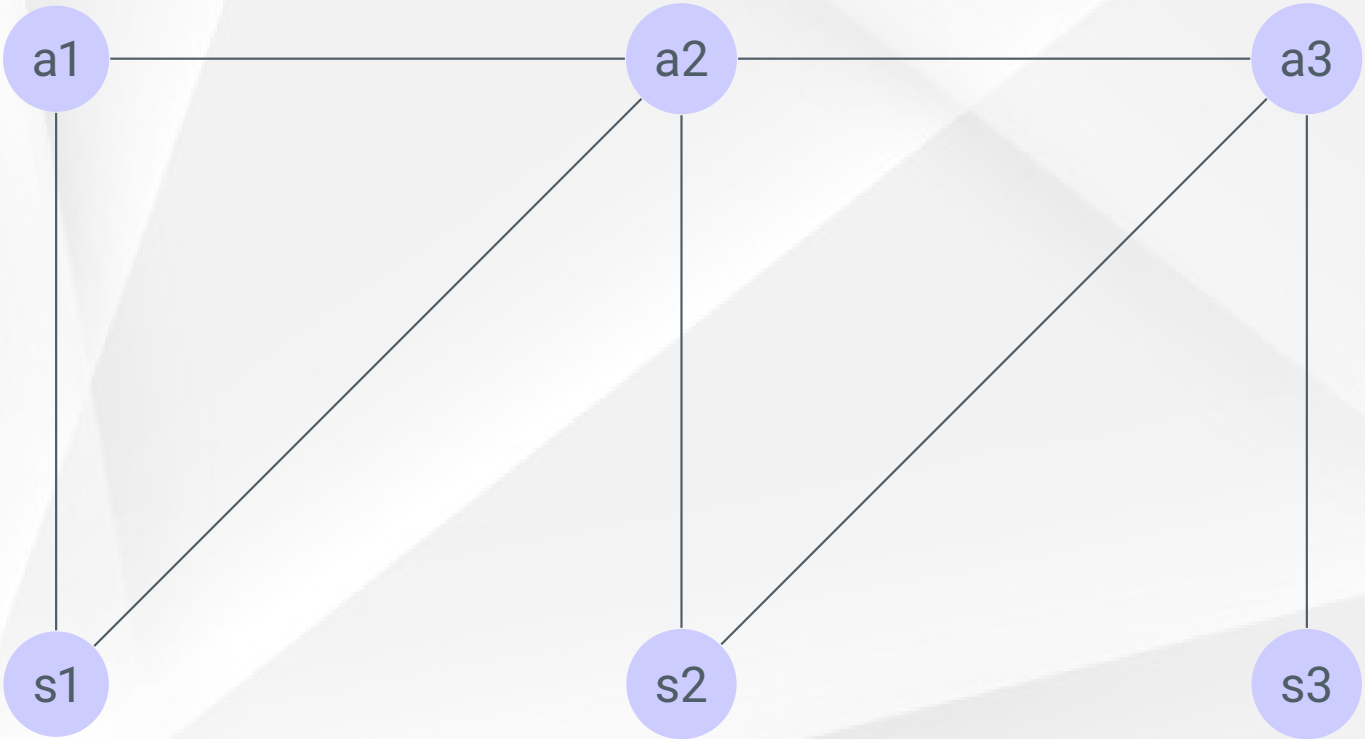
Environments

- Mountain Car [1]
- Cart Pole
- Acrobot

© Brendan Crowe

CRF

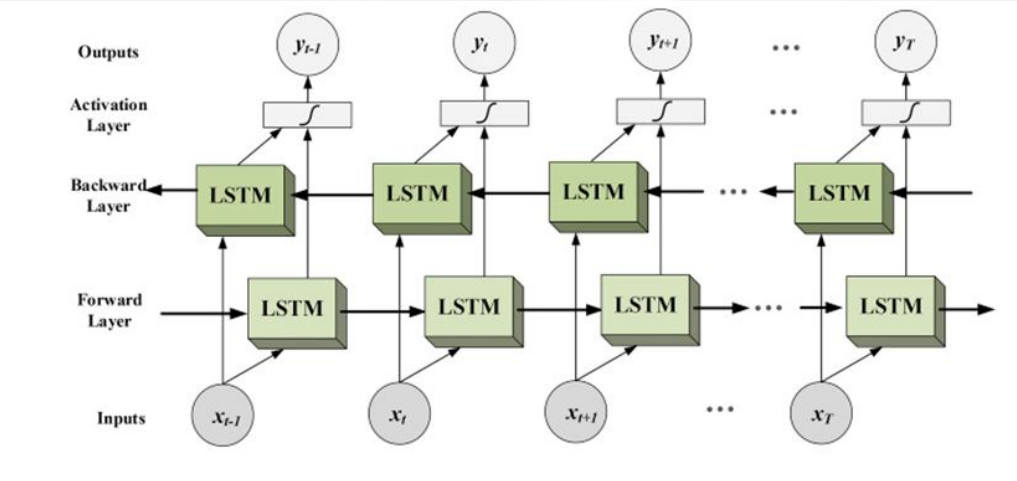
- Linear Chain CRF
- Infinite state space
- Probabilistic re-weighting of sequences.



© Brendan Crowe

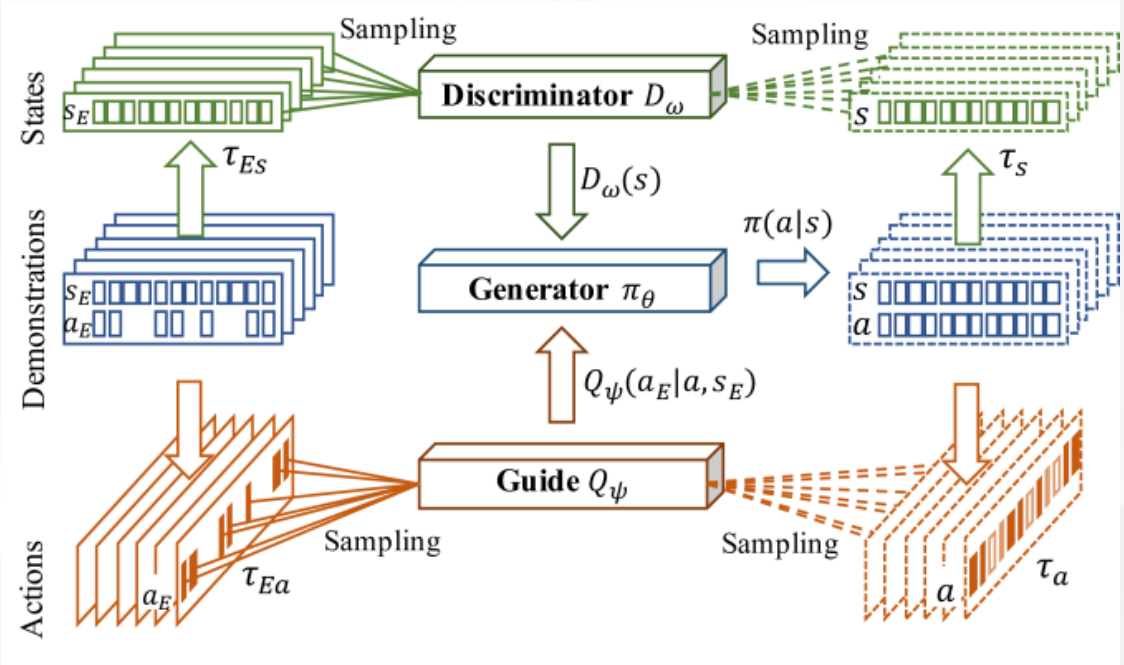
Bidirectional LSTM

- Data augmentation
- Breaks with adversarial data



GAIL

- Generative Adversarial Imitation Learning [2]
- Based upon GANs
- One of the most advanced frameworks in IL today
- Generalizes well and is scale-able



© Brendan Crowe

Results

Table: Mountain Car

Metric	Accuracy	Reward
Expert	1.00 ± 0.00	-99 ± 5.92
CRF	$\mathbf{0.97 \pm 0.013}$	$\mathbf{-108.5 \pm 6.47}$
LSTM	0.978 ± 0.003	-106.7 ± 7.78
GAIL	$\mathbf{0.995 \pm 0.005}$	-105.25 ± 7.73
Random	NA	$\mathbf{-200 \pm 0}$

© Brendan Crowe

Results

Table: Cart Pole

Metric	Accuracy	Reward
Expert	1.00 ± 0.00	200 ± 0.00
CRF	0.99 ± 0.024	200 ± 0.00
LSTM	0.999 ± 0.0008	200 ± 0.00
GAIL	$\mathbf{0.99 \pm 0.0013}$	200 ± 0.00
Random	NA	$\mathbf{26.45 \pm 5.25}$

© Brendan Crowe

Results

Table: Acrobot

Metric	Accuracy	Reward
Expert	1.00 ± 0.00	-103.1 ± 27.53
CRF	$\mathbf{0.976 \pm 0.013}$	-91.95 ± 16.69
LSTM	$\mathbf{0.99 \pm 0.0036}$	-85.5 ± 10.34
GAIL	$\mathbf{0.9957 \pm 0.0031}$	-96 ± 23.65
Random	NA	$\mathbf{-500 \pm 0}$

© Brendan Crowe

Going Forward

- Results with adversarial data
- Comparison of run-time CRF vs InfoGAIL [4]

© Brendan Crowe

Citations

[1] Greg Brockman et al. *OpenAI Gym*. 2016. eprint: arXiv:1606.01540.

[2] Jonathan Ho and Stefano Ermon. *Generative Adversarial Imitation Learning*. 2016. arXiv: 1606.03476 [cs.LG].

[3] Ahmed Hussein et al. "Imitation Learning: A Survey of Learning Methods". In: *ACM Comput. Surv.* 50.2 (Apr. 2017). ISSN: 0360-0300. DOI: 10.1145/3054912. URL: https://doi.org/10.1145/3054912.

[4] Yunzhu Li, Jiaming Song, and Stefano Ermon. *InfoGAIL: Interpretable Imitation Learning from Visual Demonstrations*. 2017. arXiv: 1703.08840 [cs.LG].

[5] Dean A Pomerleau. "Efficient training of artificial neural networks for autonomous navigation". In: *Neural computation* 3.1 (1991), pp. 88–97.

© Brendan Crowe