FeaTREL
Transfer Reinforcement Learning using features
Application to the ball-in-cup problem

Guillaume Doquet

Supervisors: Michèle Sebag, David Filliat

LRI, U. Paris-Sud; U2IS, ENSTA

12 Avril 2016

Motivations

Plan of this talk

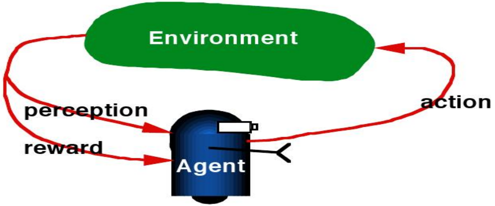
Formal background

- Reinforcement learning
- Transfer learning

Feature-based Transfer for RL

- FeaTREL, overview
- Empirical validation

Reinforcement Learning



Notations

State space $\mathcal{S}$, Action space $\mathcal{A}$
Policy $\pi : \mathcal{S} \mapsto \mathcal{A}$
Goal : Find $\pi^* = \text{argmax } \mathcal{F}(\pi)$

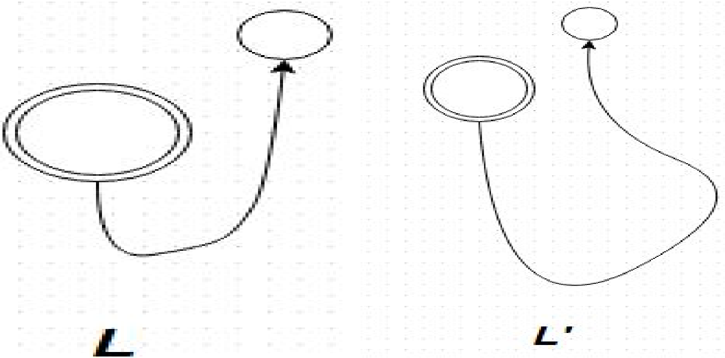
Parameterized policy

$$\text{weight vector of a NN} : W \mapsto \pi_W$$

Some Issues: $\mathcal{F}$ is an expectation

- noisy optimization problem
- expensive optimization

Transfer Learning



- Goal : Use π^*_L to accelerate learning of π'^*_L

FeaTREL 1/4

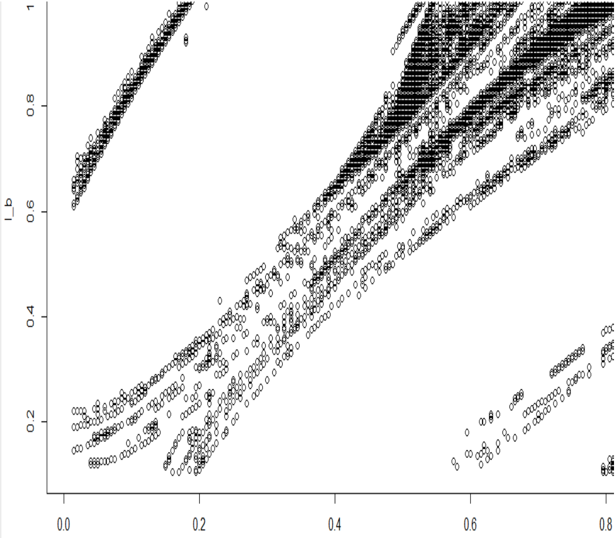
Core ideas

- Observe the trajectory with given policy parameters θ and rope length L
- Extract features $\phi(L, \theta)$ from trajectory (e.g. number of rebounds of the ball)
- Use features to find optimal parameters θ' for rope length L'

FeaTREL 2/4

First milestone

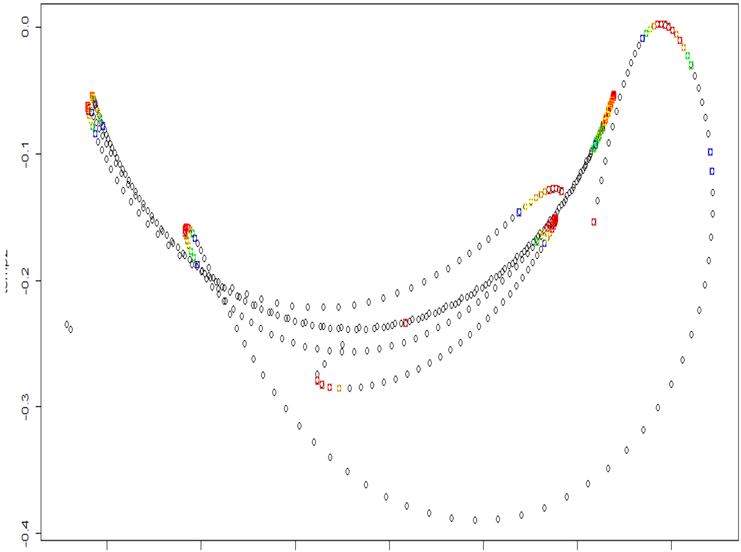
- Find/learn mapping $\phi(L, \theta)$ from (L, θ) unto features



FeaTREL 3/4

Second milestone

- Observe features $\phi(L', \theta)$ for (θ, L')



FeaTREL 4/4

Third milestone

- Start from θ_0
- Modify θ: $\theta_1 = \theta_0 + \Delta\theta$ where $\Delta\theta$ s.t. :

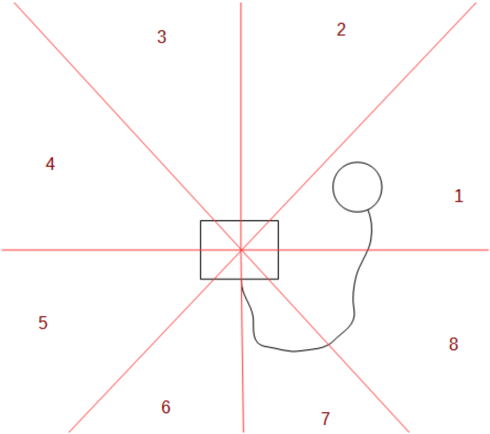
$$\frac{\partial \phi}{\partial L}.\Delta L = -\frac{\partial \phi}{\partial \theta}.\Delta\theta$$

- Iterate. Stopping criterion : θ_n optimal or max number of iterations reached.

Empirical validation 1/3

Implementation

- $\theta \in \mathbb{R}^2$: Major and minor axes of the cup's elliptic motion
- Features $\phi(L, \theta) \in \mathbb{R}^{16}$: Histogram of distance and angles between ball and cup during the trajectory



Empirical validation 2/3

Implementation

- Ball-in-cup simulator
- Online learning of $\Delta\theta$ done by a Neural Network

Experimental setting

- $L = 30$ cm , $L' = 32$, 35 or 60 cm
- Success if ball ends in cup
- Performance indicators :
  - Accuracy % on whole dataset (circa 150 trajectories)
  - Average number of iterations to succeed
  - Average distance (in θ) between solution and initial point
- All data is averaged on 10 runs

Empirical validation 3/3

L	Max iter	Accuracy (avg and std) FeaTREL & RandomWalk & CMA-ES						Avg θ dist FeaTREL & RandomWalk & CMA-ES			Avg iter FeaTREL & RandomWalk	
0.32	8	0.692	0.030	0.545	0.024	0.231	0.024	0.059	0.317	0.719	3.34	3.31
0.32	25	0.923	0.059	0.811	0.061	0.531	0.057	0.202	0.582	1.902	8.55	7.65
0.35	8	0.804	0.012	0.461	0.021	0.119	0.045	0.141	0.236	0.458	5.21	3.30
0.35	25	0.973	0.059	0.755	0.072	0.406	0.064	0.278	0.513	1.088	8.35	8.12
0.60	8	0.140	0.076	0.154	0.027	0.203	0.052	0.032	0.105	0.599	4.70	4.36
0.60	25	0.245	0.081	0.560	0.058	0.417	0.049	0.082	0.469	1.554	7.11	12.26

Conclusions

FeaTREL PROs

- Outperforms direct policy search

FeaTREL CONs

- Requires good features;
- Requires $\phi(L, \theta)$ to be learned
- Valid when target is sufficiently close to source

Perspectives

- Continuous optimization objective $\mathcal{F}$
- More complex parametric spaces
- *In situ* validation

Questions ?