Démarche

Mehdi Mounsif

16 mars 2018

1 Récap : The day before

- Etablissement d'une feuille de route
- Préparation pour IK
- Préparation pour GAN

2 Résultats de la réunion

Intégrer plus de robotique. Permettra de publier rapidos (et d'obtenir plus de liberté).

Feuille de route : Robotique

- Apprendre modèle géométrique direct
- Modèle géométrique indirect
- Déplacer le robot avec ces modèles analytiques
- Injecter du bruit dans les capteurs et le gérer avec RL

Feuille de route : IA

- DDPG
- IRL
- GAN
- Model-based

Feuille de route : Misc

- Tester PPO sur Atari
- Tester sur SNESx9
- Animation IK

3 IK

Pas d'activité hier. Regarder la conférence sur model-based, voir si quelque chose est faisable. Lecture de **Guided Cost Learning** [1], papier référence dans la conférence d'IA. Fait intervenir Inverse Optimal Control. Permet d'obtenir simultanément une approximation de la fonction de récompense et une politique calquée sur celle de l'expert.

4 GANs

Premiers tests sur Pokémon. Résultats loins de la perfection, mais on remarque que des formes de plus en plus ressemblantes sont générées. Malheureusement, THCudaCheck raised an error. Fin de l'apprentissage et corruption des sauvegardes. Par conséquent, pas d'images à montrer.

Tester GANs sur des problèmes différents.

5 DDPG

Reprise de l'implémentation. Pour l'instant, pas de target networks, ni de replay memory. A tester.

6 GAIL : Generative Adversarial Imitation Learning

Permet d'apprendre une politique IRL-style sans s'encombrer de la fonction de récompense [2]. Semble, à priori, plus approprié que [1].

En pratique, les auteurs introduisent une fonction de régularisation ψ qui cherche à minimiser la distance entre ce qu'ils appellent **occupancy measure** de l'expert et celle de la politique en cours d'apprentissage. Cette notion représente la distribution des paires état-actions rencontrés suivant la politique.

On a :

$$\psi_{GA} = \begin{cases} \mathbb{E}_{\pi_E}[g(c(s,a))] \text{ si } c < 0 \\ +\infty \text{ sinon} \end{cases} \qquad where \quad g(x) = \begin{cases} -x - \log(1 - e^x) \text{ si } x < 0 \\ +\infty \text{ sinon} \end{cases} \tag{1}$$

Références

[1] Finn, C., Levine, S., and Abbeel, P. Guided cost learning : Deep inverse optimal control via policy optimization. *CoRR abs/1603.00448* (2016).

[2] Ho, J., and Ermon, S. Generative adversarial imitation learning. *CoRR abs/1606.03476* (2016).