

APPRENTISSAGE ADAPTATIF DE COMPORTEMENTS ÉTHIQUES

Rémy Chaput ¹ Olivier Boissier ² Mathieu Guillermin ³ Salima Hassas ¹

30 Juin 2020

¹Univ. Lyon, Université Lyon 1, LIRIS, UMR5205, F-69622, LYON, France

²Univ. Lyon, IMT Mines Saint-Étienne, CNRS, Laboratoire Hubert Curien UMR 5516

³GEEST, UR Confluence, Sciences et Humanités, Université Catholique de Lyon

Travaux financés par la Région Auvergne Rhône-Alpes (Pack Ambition Recherche), dans le cadre du projet Ethics.AI.

PLAN DE LA PRÉSENTATION

Introduction

Contributions

Cas d'application

Expérimentations

Conclusion

INTRODUCTION

- Les systèmes d'IA ont de plus en plus d'**impact** sur les humains
 - (conséquences positives ou négatives)
- La société exprime le besoin d'intégrer des **considérations éthiques** [OCD20]
- De nombreuses approches ont été proposées [Yu+18], mais la question de l'**adaptation** reste ouverte

Notre objectif

Proposer un **modèle d'apprentissage** capable d'apprendre des **comportements "éthiques"**¹ et de **s'adapter** face aux changements dans la fonction de récompense.

¹Comportement pouvant être jugé par un humain comme correspondant à ses valeurs ou principes éthiques

CONTRIBUTIONS

MODÈLE D'APPRENTISSAGE

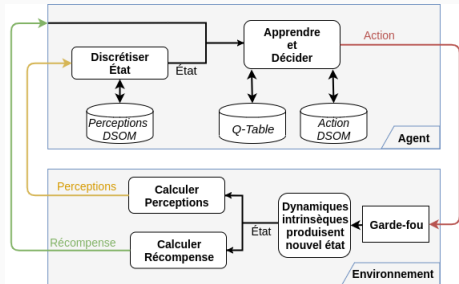


Figure 1: Cycle d'apprentissage et décision

- Modèle s'inspirant de l'approche de Smith [Smi02]
 - Q-Learning [WD92]
 - 2 Dynamic Self-Organizing Maps [RB11]
- Récompense différenciées [YT14] (Credit Assignment Problem)
 - $R_i = V(\text{Actions}) - V(\text{Actions} \setminus \{\text{Actions}_i\})$
- Éthique par conception [Dig18]
 - (un peu d'Éthique dans la conception également)

CAS D'APPLICATION

Répartition de l'énergie ; maximiser confort et intérêt des autres

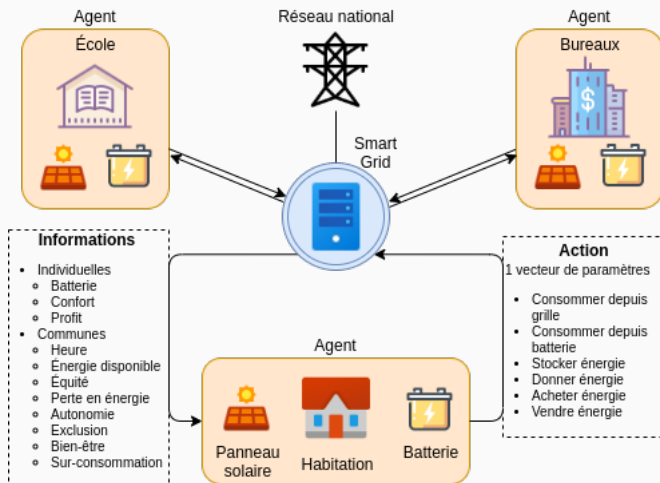


Figure 2: Schéma d'une Smart Grid (dans notre simulateur).

EXPÉRIMENTATIONS

- 4 Modèles
 - Aléatoire
 - DDPG [Lil+19] (Deep Deterministic Policy Gradient)
 - Q-SOM [Smi02] (modèle initial de Smith)
 - Q-DSOM (notre modèle)
- 6 Fonctions de récompense
 - Équité : minimiser dispersion des comforts
 - Sur-consommation : minimiser l'énergie sur-consommée
 - Multi-Objectif Somme : minimiser sur-consommation et maximiser confort personnel
 - Multi-Objectif Produit
 - Adaptabilité¹ :
 - D'abord minimiser sur-consommation
 - Puis, minimiser sur-consommation et maximiser équité
 - Adaptabilité² :
 - D'abord minimiser sur-consommation
 - Puis, minimiser sur-consommation et maximiser équité
 - Puis, minimiser sur-consommation, maximiser équité et confort personnel

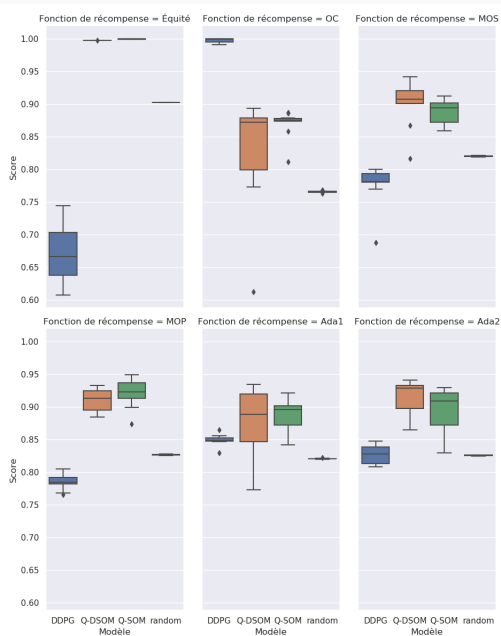


Figure 3: Résultats pour chaque modèle et fonction de récompense.

Score de simulation = moyenne du score "global" (tous agents) sur l'ensemble des pas de temps.

CONCLUSION

- Améliorations techniques :
 - Q-DSOM et Q-SOM ont des résultats assez proches
 - Apprentissage des actions (multi-dimensionnelles)
 - Question de la **non-stationarité**
- Scénarios différents :
 - **Profils de consommation annuels au lieu de journaliers**
 - Périodes de "pénuries" (moins d'énergie dans la grille)
 - Différentes tailles de sociétés multi-agents
- Questions éthiques :
 - Les fonctions de récompense sont-elles adaptées ?
 - Possibilité d'avoir des fonctions différentes selon les agents

MERCI DE VOTRE ATTENTION !

QUESTIONS ?

REFERENCES



Virginia Dignum. “Ethics in Artificial Intelligence: Introduction to the Special Issue”. en. In: *Ethics and Information Technology* 20.1 (Mar. 2018), pp. 1–3.



Timothy P. Lillicrap et al. “Continuous Control with Deep Reinforcement Learning”. en. In: *arXiv:1509.02971 [cs, stat]* (July 2019).



Leila Ouchchy, Allen Coin, and Veljko Dubljević. “AI in the Headlines: The Portrayal of the Ethical Issues of Artificial Intelligence in the Media”. en. In: *AI & SOCIETY* (Mar. 2020).

BIBLIOGRAPHIE II



Nicolas Rougier and Yann Boniface. “Dynamic Self-Organising Map”. en. In: *Neurocomputing* 74.11 (May 2011), pp. 1840–1847.



Andrew James Smith. “Applications of the Self-Organising Map to Reinforcement Learning”. en. In: *Neural Networks* 15.8-9 (Oct. 2002), pp. 1107–1124.



Christopher J. C. H. Watkins and Peter Dayan. “Q-Learning”. en. In: *Machine Learning* 8.3 (May 1992), pp. 279–292.



Logan Yliniemi and Kagan Tumer. “Multi-Objective Multiagent Credit Assignment Through Difference Rewards in Reinforcement Learning”. en. In: *Simulated Evolution and Learning*. Ed. by Grant Dick et al. Lecture Notes in Computer Science. Springer International Publishing, 2014, pp. 407–418.



Han Yu et al. “Building Ethics into Artificial Intelligence”. In: *Proceedings of the 27th International Joint Conference on Artificial Intelligence*. IJCAI’18. Stockholm, Sweden: AAAI Press, July 2018, pp. 5527–5533.