

Apprentissage évolutif de comportements éthiques

Application aux *Smart Grids*

Rémy Chaput¹ Olivier Boissier² Mathieu Guillermin³ Salima Hassas¹

¹LIRIS-CNRS Université de Lyon

²IMT, Mines Saint-Étienne, LaHC

³Lyon Catholic University

Objectifs

Proposer un modèle d'apprentissage, permettant de :

- construire une représentation du monde à partir de mesures objectives et calculables
- apprendre un comportement compatible avec des valeurs humaines
- apprendre selon la contribution de chaque agent
- apprendre plusieurs valeurs éthiques en même temps

Introduction

- Croissance rapide de l'utilisation d'algorithmes ayant un **impact potentiel sur les humains**
- Ex : *trading* automatique, conduite assistée, allocation de ressources
- **Gestion de l'énergie dans les *Smart Grids***
- Demande de plus en plus forte pour algorithmes exhibant un **comportement éthique**
- Impact de groupes de travaux
 - IEEE Working Group P7000™
 - AI High-Level Expert Group de l'Union Européenne

Cas d'étude

- Ensemble d'agents **producteurs-consommateurs**
- Représentent un gestionnaire de l'énergie d'un bâtiment (Hôpital ou Habitation) relié à la grille
- Doivent maximiser leur **confort** tout en considérant l'**intérêt de la grille**
- Se répartissent une quantité d'énergie disponible au sein de la grille

Approches

- Plusieurs agents apprenant en même temps ; la récompense doit correspondre à leur contribution.
 - *Multi-Agent Credit Assignment Problem*
 - utilisation de *Difference Rewards* :

$$DR_m = V(\text{Action}(i)|i \in \text{Agents}) - V(\text{Action}(i)|i \in \text{Agents} \setminus \{m\})$$
- Apprentissage de plusieurs valeurs ; sous-récompenses multiples associées.
 - *Multi-Objective Rewards*
 - utilisation de *Multi-Objective Difference Rewards* [2] :

$$R_+ = \sum_{o \in O} p_o f_o \quad R_\lambda = \prod_{o \in O} f_o$$
- Perceptions et actions sont des vecteurs de réels ; les agents doivent apprendre à discrétiser ces vecteurs pour apprendre une politique.
 - *Continuous Reinforcement Learning*
 - *Q-Learning* avec Cartes Auto-Organisatrices de Kohonen pour discrétiser les vecteurs (selon [1]).

Conclusion

- Proposition d'un **modèle d'apprentissage** basé sur des **métriques objectives**.
- Récompense **multi-objectifs** et **adaptée à leur contribution**.
- Comportements exhibent des **valeurs humaines**
- Création d'un **simulateur multi-agent générique** ; instantiation sur le problème des *Smart Grids*.
- Avec certaines fonctions de récompense, **adaptation** à un changement de situation.

Modèle

Utilisation de métriques objectives avec portée éthique (e.g. équité)

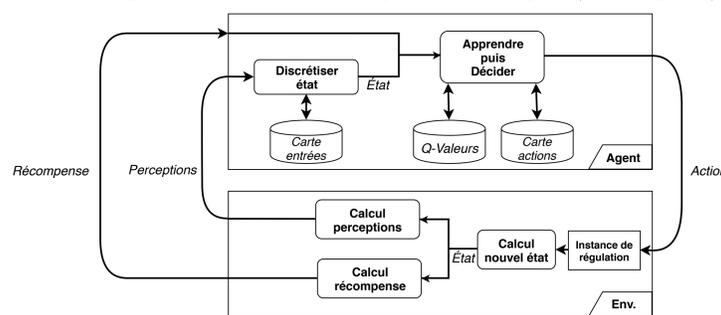


FIGURE 1 – Schéma du modèle d'apprentissage.

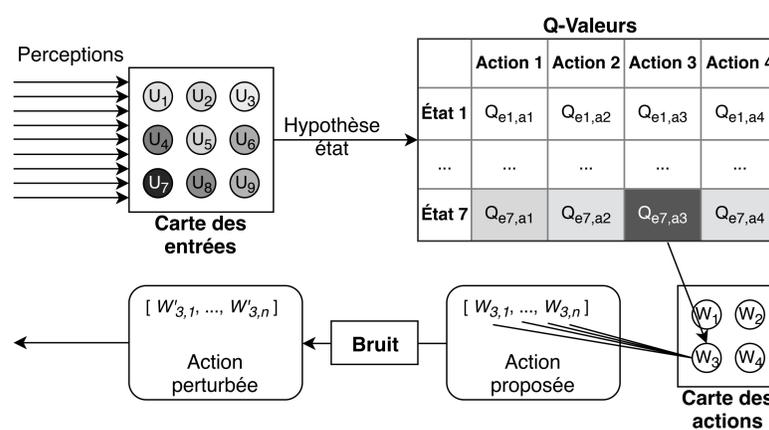


FIGURE 2 – Processus de décision détaillé.

Résultats

	Aléatoire	Appr.
équité	0.76	0.99
équité 2	0.76	0.99
surconsommation	0.18	0.56
MO-produit	0.57	0.58
MO-somme	0.33	0.32
adaptabilité 1	0.46	0.52
adaptabilité 2	0.53	0.67
adaptabilité 3	0.59	0.72

TABLE 1 – Score moyen pour chaque fonction de récompense, avec et sans apprentissage.

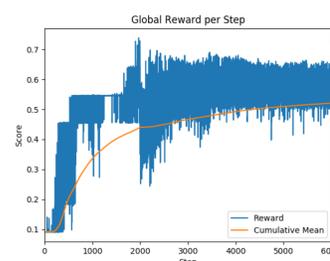


FIGURE 3 – Score global, adaptation des agents après changement de la récompense à t=2000.

Perspectives

- Améliorer la capacité d'adaptation
- Améliorer la discrétisation (topologie de l'espace d'entrée et exploration de l'espace des actions)

Références

- [1] A. J. Smith. Applications of the self-organising map to reinforcement learning.
- [2] L. M. Ylioniemi and K. Tumer. Multi-objective multiagent credit assignment through difference rewards in reinforcement learning.

Remerciements

Ce travail a été financé par la région Auvergne-Rhône-Alpes Pack Ambition Recherche dans le cadre du projet *Ethics.AI*.

Contact

- remy.chaput@liris.cnrs.fr
- Projet *Ethics.AI* (Artificial constructivist agents that learn ETHICS in human-involved co-construction)
- ethicsai@liris.cnrs.fr

La Région Auvergne-Rhône-Alpes

LIRIS

LABORATOIRE HUBERT CURIE

UCLY LYON CATHOLIC UNIVERSITY