Rapport Intermédiaire de Thèse en Informatique à l'Université de Lorraine (École doctorale IAEM)
Découverte automatique d'attributs pour l'imitation

Edouard KLEIN

5 septembre 2012

Sujet de thèse

L'adoption de systèmes automatisés autonomes peut réduire les coûts et les risques. Cependant, la complexité de leur mise au point gêne leur adoption généralisée. Des approches permettant à ces systèmes de prendre en charge des environnements de plus en plus difficiles font apparaître de nouveaux défis. L'un de ceux-ci est la définition de la consigne. A mesure que les tâches confiées aux systèmes automatiques se complexifient, la définition de ces dernières devient moins aisée.

L'*apprentissage par renforcement inverse* (ARI), cadre dans lequel s'inscrit cette thèse, a pour objet de contourner les difficultés évoquées : de la même manière que les jeunes gens n'apprennent pas à conduire en lisant le manuel de leur auto mais en observant leurs parents et leurs moniteurs de conduite puis en se mettant derrière le volant, nous comptons apprendre la tâche à effectuer en observant un *expert* la réaliser. Cette démarche exploite la capacité humaine à résoudre intuitivement et rapidement des conflits qu'il serait difficile d'analyser sur papier. Pour reprendre l'exemple précédent, un automobiliste saura après un rapide coup d'œil dans son rétroviseur s'il vaut mieux qu'il pile ou qu'il déboîte en urgence et effectuera sa manœuvre dans la foulée.

Notre but est de dériver du comportement d'un expert effectuant une tâche une description de cette tâche sous la forme d'une fonction de récompense, ce qui permet ensuite l'utilisation des outils d'apprentissage par renforcement pour apprendre cette tâche à un agent. Cela ouvrirait le champ d'application de l'apprentissage par renforcement à des tâches encore inaccessibles car trop complexes pour être "expliquées".
L'intitulé de la thèse, *Découverte automatique d'attributs pour l'imitation*, isole une partie du problème : il s'agit d'extraire de la description de la suite d'*états* traversés par l'expert les informations pertinentes pour l'expression de la récompense.

Contributions au domaine

Il faut attendre 2000 pour qu'une publication pose formellement le problème. La contribution centrale intervient en 2004 et introduit la notion d'*attribut moyen*, que suggérait déjà l'analyse de 2000. Après 2004, plusieurs travaux apparaissent qui utilisent cette notion d'attribut moyen.
L'attribut moyen est une mesure liée à la distribution des états que traverse un agent en suivant sa politique dans l'environnement. Cette mesure tient une place centrale dans nombre d'algorithmes existants, car deux agents ayant des attributs moyens similaires rempliront une certaine tâche (*i.e.*, optimiseront une certaine récompense) de manière similaire.

Notre première contribution au domaine consiste en un mécanisme de calcul de cet attribut moyen [KGP11b]. Inspiré d'algorithmes existants pour l'approximation de fonction de valeur, thème central en apprentissage par renforcement, cette contribution apporte une méthode de calcul permettant l'évaluation *off-policy* de l'attribut moyen d'une politique. Sans l'anglicisme, cela signifie que l'on peut évaluer une grandeur relative à une politique en observant *une autre politique* (comme par exemple celle de l'expert). Testée en l'injectant dans l'approche centrale de 2004, cette idée a permis de résoudre les problème de l'apprentissage par renforcement inverse sur un problème jouet simple à partir des seules données issues de l'expert, mais n'a pas permis de complètement lever les obstacles imposés par la structure des algorithmes existants. Ceux-ci nécessitent en effet très majoritairement de résoudre le problème direct (celui de l'apprentissage par renforcement) de manière répétée. Cela n'est pas possible uniquement avec les données de l'expert à part sur certains problèmes jouet. De manière générale, les algorithmes de la littérature ont besoin d'une autre source d'information que les données de l'expert.
Cette contribution permet néanmoins de s'affranchir d'un simulateur pour estimer une grandeur centrale dans la majorité des approches du domaine. Cela permet d'éviter les soucis liés à la modélisation de l'environnement ou à un trop grand besoin en données (coûteuses à produire). Une autre contribution de plus faible envergure consiste en un apport sur la définition mathématique formelle du problème [KGP11d] permettant de réduire l'espace dans lequel on doit chercher les solutions.
Nous avons poursuivi, lors de la deuxième année de thèse, en proposant deux nouveaux algorithmes qui permettent la résolution de l'ARI dans des domaines où les algorithmes existants ne fonctionnent pas. Les différentes contributions autour d'un de ces algorithmes a donné lieu à trois publications, dans une conférence francophone [KPGP12a] ainsi que deux conférences internationales [KPGP12b, KPGP12c]. Le travail sur l'autre algorithme se poursuit.

J'ai pu participé à cinq conférences en temps qu'auteur, j'ai pu rencontrer la communauté *Machine Learning* au sens large à IJCAI [1], ainsi qu'à ECML [2] qui précédait EWRL [3] où j'ai pu présenter mes travaux à un public plus directement concerné par l'ARI du fait du sujet précis de ce *workshop*. Enfin j'ai rencontré la communauté francophone à JFPDA [4] puis l'année suivante à CAp [5]. J'ai à nouveau participé à EWRL en 2012. Les échanges formels et informels permis par ces déplacements ont enrichi ma culture et ma reflexion.

1 Perspectives

Outre la rédaction de la thèse, qui sera commencée aux alentours de février pour une remise du manuscrit en juin et une soutenance en septembre, cette année sera occupée par l'application de nos nouveaux algorithmes sur des problèmes complexes et tirés de cas concrets afin d'évaluer empiriquement leur performance.

2 Mes publications

[KGP11a] Edouard Klein, Matthieu Geist, and Olivier Pietquin. Apprentissage par imitation étendu au cas batch, off-policy et sans modèle. In *Sixièmes Journées Francophones de Planification, Décision et Apprentissage pour la conduite de systèmes (JFPDA 2011)*, page 9 pages, Rouen (France), June 2011.

[KGP11b] Edouard Klein, Matthieu Geist, and Olivier Pietquin. Batch, Off-policy and Model-free Apprenticeship Learning. In *Proceedings of the European Workshop on Reinforcement Learning (EWRL 2011)*, Lecture Notes in Computer Science (LNCS), page 12 pages, Athens (Greece), september 2011. Springer Verlag - Heidelberg Berlin.

1. http ://ijcai-11.iiia.csic.es/
2. http ://www.ecmlpkdd2011.org/
3. http ://ewrl.wordpress.com/past-ewrl/ewrl9-2011/
4. https ://zanuttini.users.greyc.fr/jfpda2011/
5. http ://cap2012.loria.fr/

[KGP11c] Edouard Klein, Matthieu Geist, and Olivier Pietquin. Batch, Off-policy and Model-Free Apprenticeship Learning. In *IJCAI Workshop on Agents Learning Interactively from Human Teachers (ALIHT 2011)*, Barcelona (Spain), July 2011. 6 pages.

[KGP11d] Edouard Klein, Matthieu Geist, and Olivier Pietquin. Reducing the dimentionality of the reward space in the Inverse Reinforcement Learning problem. In *Proceedings of the IEEE Workshop on Machine Learning Algorithms, Systems and Applications (MLASA 2011)*, page 4 pages, Honolulu (USA), December 2011.

[KPGP12a] Edouard Klein, Bilal PIOT, Matthieu Geist, and Olivier Pietquin. Classification structurée pour l'apprentissage par renforcement inverse. In *Actes de la Conférence Francophone sur l'Apprentissage Automatique (CAp 2012)*, Nancy, France, 2012.

[KPGP12b] Edouard Klein, Bilal Piot, Matthieu Geist, and Olivier Pietquin. Structured Classification for Inverse Reinforcement Learning. In *European Workshop on Reinforcement Learning (EWRL 2012)*, Edinburgh (UK), 2012.

[KPGP12c] Edouard Klein, Bilal Piot, Matthieu Geist, and Olivier Pietquin. Structured Classification for Inverse Reinforcement Learning. In *NIPS*, Lake Tahoe (USA), 2012.