Apprentissage par renforcement et différences convexes

Objectifs

- Maîtrise des concepts d'apprentissage par renforcement
- Maîtrise des concepts d'optimisation de différences de fonctions convexes
- Application à la détermination d'une politique de contrôle optimal à partir de données

Description

L'apprentissage par renforcement est une catégorie d'apprentissage automatique qui se différencie des autres par le fait qu'elle a pour objectif l'optimisation d'une séquence de décisions, prenant en compte l'aspect temporel et surtout dirigé par un but du comportement. Cette méthode, d'inspiration biologique, est fondée sur l'accumulation par la machine de récompenses numériques distribuées après chaque décision. Le comportement appris est celui qui maximise, sur le long terme, l'accumulation de récompenses, menant à une séquence de décisions optimale. Le but est donc d'apprendre un contrôleur optimal pour une tâche complexe (difficile à modéliser) à partir de données. Le cas générique est celui de l'interaction dans un environnement stochastique, c'est à dire dont les réponses aux actions ne sont pas déterministes. On parle donc de contrôle de systèmes dynamiques stochastiques. Les applications sont nombreuses comme par exemple l'optimisation de la consommation d'énergie dans des grands parc industriels, la gestion d'interactions homme-machine, l'assistance à la conduite etc.

Il existe un grand nombre d'algorithmes dans la littérature pour tenter de résoudre ce problème [1]. Toutefois, elles passent mal à l'échelle et des méthodes d'approximation sont alors nécessaires. Ainsi, l'optimalité n'est plus garantie. Il y a plusieurs raisons à cela, parmi lesquelles le caractère non-convexe et non-différentiable de la fonction de coût qu'optimisent la plupart de ces algorithmes.

Depuis une trentaine d'années, une classe particulière de problèmes d'optimisation non-convexe a vu le jour. Il s'agit de l'optimisation de différences de fonctions convexes [2]. On parle aussi de programmation DC ou DCA (DC algorithms). Ce paradigme profite des propriétés intéressantes des classes de fonctions pouvant se mettre sous la forme de différences de fonctions convexes et particulièrement de leur propriétés invariantes face aux opérations fréquemment rencontrées en optimisation (par exemple, la fonction max).

Le but de ce projet est donc de reformuler le problème d'apprentissage par renforcement sous la forme d'un problème d'optimisation DC et d'ensuite trouver l'algorithme DC le plus approprié pour résoudre le problème d'optimisation.

Références

[1] R.S. Sutton and A.G. Barto. *Reinforcement learning : An introduction*. The MIT press, 1998.

[2] Pham Dinh Tao et al. The dc (difference of convex functions) programming and dca revisited with dc models of real world nonconvex optimization problems. *Annals of Operations Research*, 133(1-4) :23–46, 2005.

Encadrants

Olivier Pietquin : `olivier.pietquin@lifl.fr` (Prof. Lille 1 - équipe SequeL)