Approximations de l'Algorithme Itérations sur les Politiques Modifié

Bruno Scherrer[1], Victor Gabillon[2], Mohammad Ghavamzadeh[2], Matthieu Geist[3]

[1] INRIA Nancy, LORIA, Equipe-Projet MAIA
615 rue du Jardin Botanique, 54600 Villers-lès-Nancy
Bruno.Scherrer@inria.fr

[2] INRIA Lille, Equipe-Projet SEQUEL
Parc scientifique de la Haute Borne, 40, avenue Halley - Bât A - Park Plaza, 59650 Villeneuve d'Ascq
Victor.Gabillon@inria.fr
Mohammad.Ghavamzadeh@inria.fr

[3] Supélec Campus de Metz, Equipe IMS
2 rue Edouard Belin, 57070 Metz
Matthieu.Geist@supelec.fr

Résumé :
Itérations sur les politiques modifié (MPI) est un algorithme de programmation dynamique qui généralise les deux algorithmes célèbres *Itérations sur les valeurs* (VI) et *sur les politiques* (PI). Malgré sa généralité, cet algorithme – et particulièremet sa mise en œuvre approchée qui est utilisée lorsque les espaces d'états/actions sont très grands – n'a pas encore été l'objet d'une analyse approfondie. Nous proposons ici trois implémentations approchées de MPI (AMPI) qui sont des extensions d'algorithmes de la littérature (*Fitted Value Iteration*, *Fitted Q-Iteration* et *Classification Based Policy Iteration*). Nous développons une analyse de la propagation d'erreur qui unifie celles développées indépendemment pour VI et PI dans la littérature. Nous fournissons enfin une analyse en échantillons finis pour le dernier algorithme basé sur un classifieur de politiques, qui est en quelque sorte le plus général. Une observation intéressante est que la paramètre principal de MPI permet de contrôler, dans la borne de performance, l'équilibre entre les erreurs dans le calcul des valeurs et celles dans l'estimation de la politique gourmande.

Mots-clés : Apprentissage par renforcement, programmation dynamique approchée, analyse

Le corps de cet article est à paraître, en langue anglaise, dans ICML'2012.