Politiques de meilleur compromis dans les processus décisionnels de Markov multicritères

Paul Weng[1]

UPMC ; LIP6, DESIR ; 104 avenue du Président Kennedy, 75016 Paris, France
paul.weng@lip6.fr

Mots-Clés : *processus décisionnels de Markov, multicritère, meilleur compromis*

1 Présentation du problème

Les processus décisionnels de Markov (MDP) permettent de modéliser et résoudre des problèmes de décision séquentielle dans l'incertain [3]. Dans ce modèle, les préférences du décideur sont représentées par une fonction de récompenses scalaires réelles et la résolution du problème consiste à déterminer la séquence de décisions, appelée politique, qui maximise l'espérance de la somme des récompenses.

Dans les problèmes réels, pour évaluer une décision, il est souvent nécessaire de prendre en compte plusieurs dimensions, de nature généralement contradictoire (coût et qualité par exemple). De ce fait, le modèle des MDP a été étendu au modèle des MDP multicritères en considérant des fonctions de récompenses vectorielles. Dans ce cadre, on cherche alors généralement les politiques Pareto non-dominées [1, 5, 2, 7, 6]. L'inconvénient de cette approche est que l'espace des politiques non-dominées peut être très grand.

C'est pourquoi, parmi l'ensemble des politiques non-dominées, nous nous intéressons à la recherche de politiques particulières, qualifiées de meilleur compromis. En décision multicritère, une solution de meilleur compromis est une solution non-dominée dont la distance (au sens de Tchebychev) à une solution idéale (généralement non réalisable) est minimale. Cette solution particulière est intéressante car elle a généralement un profil équilibré sur chacune des composantes du vecteur de valuation.

Cette approche nous semble intéressante, que ce soit pour la prise de décision automatique ou pour l'aide à la décision. En effet, dans un système autonome, le calcul d'une solution de meilleur compromis est plus direct et plus rapide que l'approche en deux étapes qui consiste d'abord à calculer les solutions non-dominées et ensuite à choisir une parmi celles-ci. Par ailleurs, pour un système d'aide à la décision, il nous semble plus judicieux dans certains cas qu'une procédure de résolution retourne une solution particulière ayant des propriétés intéressantes plutôt que toutes les solutions non-dominées. Ainsi le décideur n'est pas noyé sous l'information, à charge pour lui de critiquer éventuellement la solution fournie. Puis par une procédure interactive, il pourrait alors explorer le front de Pareto pour atteindre la solution qu'il préfère.

2 Approche considérée

Dans les MDP classiques monocritères, il est possible de recourir à trois grandes familles de méthodes de résolution : itération de la valeur, itération de la politique et programmation linéaire. Les deux premières reposent sur la programmation dynamique. Dans notre cadre, l'exploitation de la distance de Tchebychev (induite par la norme sup) comme critère de décision introduit une non-linéarité qui interdit l'utilisation directe de ces deux méthodes. Nous nous sommes donc intéressés à l'approche fondée sur la programmation linéaire.

Viwanathan et al. [5] ont proposé un programme linéaire multi-objectif permettant de déterminer les politiques non-dominées pour un MDP multicritère. L'approche par programmation linéaire a l'avantage de pouvoir accepter de manière aisée l'introduction de nouvelles contraintes linéaires. Ainsi, nous exploitons le programme linéaire multi-objectif proposé par Viswanathan et al. et ajoutons des contraintes linéaires pour restreindre la recherche aux politiques de meilleur compromis.

Ces contraintes sont obtenues par de simples considérations géométriques sur l'emplacement de la solution de meilleur compromis dans l'espace des critères. Il est connu que cette solution est à l'intersection entre le front de Pareto et la droite passant par le point idéal et le point anti-idéal (ou mieux le point de Nadir s'il peut être obtenue simplement) [4]. Les contraintes à rajouter sont donc que la solution recherchée appartienne à la droite passant par ces deux points, ce qui s'exprime simplement avec l'équation paramétrique de cette droite.

Nous discuterons de la résolution du programme multi-objectif obtenu. Dans le cas général, il est possible d'avoir recours à différentes méthodes de résolution telles que la généralisation de l'algorithme du simplexe [8, 4]. Dans notre cas, les propriétés du problème permettent de se ramener à un programme linéaire classique mono-objectif.

Références

[1] N. Furukawa. Vector-valued markovian decision processes with countable state space. *Ann. Math. Stat.*, 36, 1965.

[2] N. Furukawa. Characterization of optimal policies in vector-valued markovian decision processes. *Mathematics of operations research*, 5(2) :271–279, May 1980.

[3] M.L. Puterman. *Markov Decision Processes - Discrete Stochastic Dynamic Programming*. John Wiley and Sons, 1994.

[4] R.E. Steuer. *Multiple criteria optimization*. John Wiley, 1986.

[5] B. Viswanathan, V.V. Aggarwal, and K.P.K. Nair. Multiple criteria markov decision processes. *TIMS Studies in the management sciences*, 6 :263–272, 1977.

[6] K. Wakuta. Vector-valued markov decision processes and the systems of linear inequalities. *Stochastic processes and their applications*, 56 :159–169, 1995.

[7] D.J. White. Multi-objective infinite-horizon discounted markov decision processes. *Journal of mathematical analysis and applications*, 89 :639–647, 1982.

[8] M. Zeleny. *Linear multiobjective programming*. Springer-Verlag, 1974.