Optimizing the **Value at Risk** in a Markov Decision Process with parameter uncertainty

Erick Delage and Shie Mannor

Although one never knows what the future holds, we often struggle to understand better the influence of our actions on how it will unfold in order to seize new opportunities (or avoid accidents). Markov decision processes (MDP) can serve this purpose in a large variety of contexts. To name only a few, MDPs have played a significant role in queue control, epidemics control, inventory management, product pricing, machine maintenance, etc. Unfortunately, one can happen to be disappointed by the performance achieved when implementing the proposed policy. In fact, while common practice suggests calibrating the parameters of this model based on historical data, the estimated values can be highly inaccurate when this data is limited. Resolving an optimal policy using the wrong parameters can often explain the difference between expected performance and what is observed during implementation.

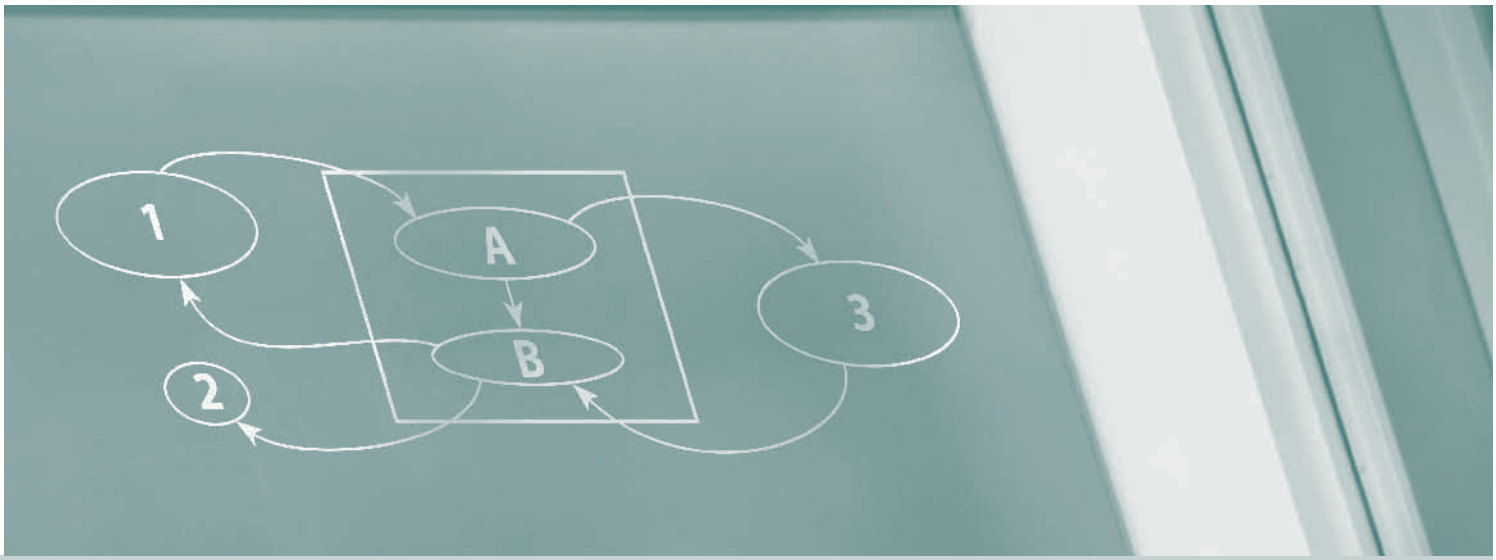
In order to address this problem, a recent approach, referred to as "robust", suggests formulating confidence intervals for each parameter of the MDP. One should then measure the performance of a policy using the worst model among the eligible ones. This approach is obviously pessimistic since it assumes that the future will only present a series of most unfortunate events. It is therefore not surprising that many consider that it leads to overly conservative policies. In this article, we suggest instead to characterize one uncertainty in the parameters using a distribution over the set of possible values. This allows one to quantify precisely for each eligible model the probability that it is the one from which the performance will actually be measured. In this context, it is natural to search for a policy that minimizes the value at risk (i.e. a percentile) of performances as measured on this distribution of eligible models. Unlike the robust approach, this value at risk criterion allows a better representation of one's risk aversion (i.e. level of conservatism). The article studies the numerical difficulties associated to the search of a policy that is optimal with respect to this criterion.

We first demonstrate that there are instances of this optimization problem that are very hard to solve; actually, some of them are NP-hard. This could justify the use of a simpler approach. Yet, we also show that if the uncertainty is limited to the cost parameters, then the problem can usually be solved efficiently. Specifically, we identify a list of distributions for the cost parameters that allows us to reformulate the problem as a conic program. When one also needs to deal with uncertainty about the dynamics of the Markov process, we present an approximation method which accuracy can be measured in terms of the number of observations. We finally illustrate the value of this criterion through experiments on a machine maintenance problem. In this context, the new approach is the most successful at choosing among different repair services for which we know little about the expertise and cost-effectiveness.

(*Operations Research*, 58(1), pp. 203-213, 2010. Original title: Percentile Optimization for Markov Decision Processes with Parameter Uncertainty)

***Erick Delage**, Department of Management Sciences, HEC Montréal*

***Shie Mannor**, Department of Electrical and Computer Engineering, McGill University and GERAD*



Optimisation de la **valeur à risque** d'un processus décisionnel de Markov avec incertitude paramétrique

Erick Delage et Shie Mannor

Alors que nul ne sait ce que l'avenir lui réserve, nous tentons constamment de mieux comprendre l'influence de nos actions sur son déroulement afin de reconnaître les opportunités (ou d'éviter les accidents) qui se présenteront. Les processus décisionnels de Markov sont utilisés à cette fin dans une multitude de domaines. Parmi la longue liste d'applications, nous pouvons compter la gestion de file d'attente, le contrôle d'épidémies, la gestion d'inventaire, la tarification de produits, la maintenance de machines, etc. Malheureusement, il arrive parfois que la performance atteinte lors de l'implémentation de la stratégie proposée soit décevante. En fait, alors que la pratique courante suggère le calibrage des paramètres du modèle à partir de données historiques, l'estimation de ces paramètres peut être erronée lorsque les données sont limitées. Un mauvais choix de paramètres peut souvent expliquer la différence entre la performance prévue par le modèle et celle qui est observée lors de l'implantation de la stratégie.

Afin de répondre à ce problème, une approche récente suggère de définir des intervalles de confiance pour chacun des paramètres pour ensuite mesurer la performance d'une stratégie par rapport au pire modèle parmi les modèles qui sont éligibles. Cette approche reflète une nature pessimiste et est souvent qualifiée de conservatrice. Dans cet article, nous suggérons plutôt de caractériser l'incertitude des paramètres sous la forme d'une distribution. Ceci permet de quantifier précisément pour chacun des modèles éligibles la probabilité qu'il soit le modèle à partir duquel la performance réelle sera mesurée. Dans ce contexte, il est donc naturel de rechercher une stratégie qui minimise la valeur à risque (ou quantile) des performances mesurées selon cette distribution de modèle. Contrairement à l'approche dite « robuste », le critère de valeur à risque permet une bonne représentation du niveau d'aversion au risque (conservatisme) souhaité. L'article étudie les difficultés numériques reliées à la recherche d'une stratégie optimale selon ce critère.

Nous démontrons d'abord que certaines instances de ce problème d'optimisation sont très difficiles à résoudre exactement. En fait, ils sont potentiellement NP-difficiles à résoudre. Heureusement, si l'incertitude est limitée aux paramètres de coût, le problème peut typiquement être résolu de manière efficace et parfois même sous la forme d'un programme conique. Autrement, dans le cas où le modèle est dérivé par inférence Bayesienne, nous identifions une méthode d'approximation dont la précision dépend du nombre d'observations recueillies. Nous illustrons finalement la valeur de ce critère à l'aide d'expériences sur un problème de maintenance d'une machine. En minimisant la valeur à risque, cette nouvelle approche permet de mieux choisir parmi différents services de réparations sur lesquels nous n'avons accès qu'à peu d'informations.

(*Operations Research*, 58(1), pp. 203-213, 2010. Titre original : Percentile Optimization for Markov Decision Processes with Parameter Uncertainty)

***Erick Delage**, Service de l'enseignement des méthodes quantitatives de gestion, HEC Montréal*

***Shie Mannor**, Département de génie électrique et informatique, Université McGill et GERAD*