

Apport de l'apprentissage par renforcement aux problèmes de maintenance optimale : application aux constellations de satellites

André Cabarbaye*, Frédéric Garcia** ,

Linda Tomasini*

*CNES, Département Sûreté de Fonctionnement,
18, avenue Ed. Belin, 31401 Toulouse Cedex 4
Email {Andre.Cabarbaye,Linda.Tomasini}@cnes.fr

**INRA, Unité de Biométrie et Intelligence Artificielle
B.P. 27, 31326 Castanet-Tolosan
Email fgarcia@toulouse.inra.fr

Mots-clés : apprentissage par renforcement, maintenance optimale, fiabilité, décision dans l'incertain, optimisation combinatoire.

Longtemps négligée, l'optimisation de la maintenance est une problématique qui intéresse de plus en plus les grands donneurs d'ordres (armement, espace, nucléaire...) depuis qu'ils ont pris conscience que le coût du maintien opérationnel de leurs systèmes est généralement supérieur au coût de ces derniers. Les **problèmes de maintenance optimale** sont des problèmes de décisions séquentielles prises pour répondre à des aléas (actions de maintenance curative) ou pour limiter leur occurrence (maintenance préventive) [ZWIN96]. Suivant la nature des systèmes, l'objectif recherché peut être de minimiser le coût des interventions plus celui induit par les indisponibilités de service, le coût à disponibilité donnée, l'indisponibilité à coût donné ou un ratio entre ces différents paramètres.

Plusieurs méthodes existent pour formaliser et résoudre ces problèmes, la plupart consistant à effectuer une modélisation du système intégrant une politique de maintenance définie a priori mais paramétrable. De nombreuses méthodes mathématiques (théorie des files d'attente, programmation dynamique, etc.), permettent d'obtenir analytiquement les paramètres optimaux de la politique, mais au prix d'hypothèses contraignantes sur le nombre d'unités du système, l'interaction entre les pannes, la structure des politiques, etc... Un autre type de résolution consiste à optimiser numériquement les paramètres de la politique à partir d'évaluations calculées ou simulées du modèle aléatoire considéré. Cette dernière solution offre alors plus de liberté quant à la définition des politiques de maintenance considérées.

Ces deux approches peuvent toutefois être conciliées dans le cadre de l'**apprentissage par renforcement** [SUTT98], ou *neuro-dynamic programming* [BERT96], qui est une méthode d'optimisation pour les problèmes de décisions séquentielles dans l'incertain basée sur une modélisation mathématique sous forme de processus décisionnels de Markov (PDM). L'apprentissage par renforcement (A/R) consiste ainsi à rechercher l'action optimale à prendre dans chacun des états du système, en exploitant pour cela un modèle de simulation de ce système. Cet apprentissage diffère de la programmation dynamique stochastique par cet utilisation de la simulation, et par l'emploi de fonctions d'approximation qui permettent de traiter des problèmes de décisions séquentielles de taille bien plus importante : de nombreuses applications inabordables directement par les algorithmes classiques de programmation dynamique ont pu être récemment résolus par apprentissage par renforce-

ment [BERT96].

Dans cette communication, on se propose d'illustrer l'apport de l'apprentissage par renforcement aux problèmes de maintenance sur un exemple concret de maintenance issu du domaine spatial, l'**entretien d'une constellation de satellites**. Pour assurer une bonne couverture de la sphère terrestre et rendre ainsi des services de télécommunication ou d'aide à la navigation, les industriels du spatial développent aujourd'hui des constellations, qui sont constituées de plusieurs satellites (de 24 pour le *GPS* à environ 200 pour *Teledesic*) répartis sur des plans orbitaux. Il est certain que sur un tel nombre de satellites utilisés pour des durées de mission de plusieurs années, les opérateurs auront à traiter des pertes de satellites dues à des défaillances des équipements ou à des fins de vie dimensionnées par les quantités d'ergol embarquées.

L'objectif étant de maximiser la disponibilité du système à moindre coût, la gestion des pannes et des redondances d'un tel système est un problème complexe de part la taille du système, mais aussi de par la diversité des actions de maintenance possibles sur ce système. En effet, pour effectuer un remplacement de satellite défaillant, on pourra utiliser un satellite en redondance (ou *spare*) déjà en orbite, ce qui pose le problème du dimensionnement de l'ensemble de ces *spares*, ou bien effectuer des lancements par grappe de satellites avec des lanceurs de capacité, coût, fiabilité et délai de lancement différents. Ces lancements par grappe pourront viser à remplacer des *spares* utilisés, des satellites défaillants ou bien arrivant en fin de vie prévisibles.

Le choix d'une politique de maintenance d'un tel système peut avoir des conséquences opérationnelles et financières très importantes. Pour l'instant, il n'existe pas encore de stratégies de maintenance complètement satisfaisantes, et la recherche d'une telle solution est un problème décisionnel complexe. La démarche que nous suivons ici est de modéliser ce problème comme un problème décisionnel de Markov, et de rechercher une politique optimale approchée par des techniques d'apprentissage par renforcement. Ce problème est intéressant pour la discipline de l'apprentissage par renforcement, car suffisamment complexe pour illustrer de manière convaincante l'intérêt de l'A/R dans la résolution de problèmes décisionnels de grande taille.

Lors de la communication, nous présenterons le problème de maintenance considéré et sa formalisation en termes de PDM. Nous développerons ensuite la démarche algorithmique propre à l'apprentissage par renforcement, en présentant et comparant différents algorithmes et différents choix d'architectures d'approximation de la fonction de valeur du problème.

[BERT96] D.P. Bertsekas et J.N. Tsitsiklis. *Neuro-Dynamic Programming*. Athena Scientific 1996.

[SUTT98] R.S. Sutton et A.G. Barto. *Reinforcement Learning : An Introduction*. MIT Press 1998.

[ZWIN96] G. Zwingelstein. *La maintenance basée sur la fiabilité*. Hermès 1996.