

# Réflexion sur la notion d'agent markovien dans des réseaux de transports

Guillaume Lozenguez

Guillaume Desquesnes

Arnaud Doniec

Éric Duviella

Mines-Douai, Institut Mines-Télécom, FRANCE,

prenom.nom@mines-douai.fr

## 1 Introduction

Plusieurs objectifs peuvent justifier la modélisation d'un réseau de transport : *comprendre* - reproduire des observations réelles ? ; *prédire* - estimer le comportement probable du système ; *anticiper* - estimer l'évolution du système en introduisant des éléments particuliers ; *planifier* - prédéfinir une succession de contrôle à appliquer.

Cependant du fait de la complexité des systèmes de transport, il est impossible de modéliser simplement son comportement global. Sur la base de système multi-agent, le système de transport est décomposé en entités autonomes. Dès lors, la modélisation s'appuie sur une décomposition en comportements individuels.

L'approche présentée ici vise à définir un framework composé d'un modèle de comportement individuellement cohérent et de mécanisme de coopération. L'idée consiste à modéliser les capacités d'actions individuelles via des automates probabilistes couplés à des protocoles de coordination définissant ainsi des processus de décision markovien distribués.

La modélisation de l'environnement s'appuie sur une définition topologique des déplacements possibles. Les nœuds identifient des positions clés interconnectées par des arcs auxquels il est possible d'associer des capacités, des probabilités d'acheminement ou de déviation, etc. Nous nous intéressons alors à modéliser les comportements soit d'agents mobiles dans un réseau soit d'agents de régulation. Cette présentation s'appuie sur deux exemples, le multi-voyageur de commerce et la gestion de la ressource en eaux dans un réseau de canaux.

## 2 Agent markovien

L'approche s'appuie sur des comportements identifiés pour des agents et modélisés sous forme de processus de décision markovien. Le comportement multi-agent dépend alors de la synergie des actions découlant des processus de décision individuels.

### 2.1 Agent dans un réseau de transport

Le paradigme des systèmes multi-agent permet d'appréhender un système complexe en le modélisant et en l'étudiant sous forme d'un ensemble d'entités (agents) en interaction. Un agent sera une entité autonome capable de

percevoir et d'agir dans son environnement.

Dans le cadre de système de transport tel qu'abordé dans cette approche, l'environnement est modélisé par un graphe de flux avec certaines capacités, les agents modélisent soit des éléments mobiles qui se déplacent d'un nœud à un autre si une connexion (arc) existe (exemple : multi-voyageur de commerce) ; soit des éléments de régulation fixe qui vont jouer sur les capacités des arcs (exemple : gestion de réseau fluvial) ( il est cependant possible d'imaginer la combinaison des deux possibilités, des agents mobiles et avec des capacités de régulation).

Un agent aura au minimum une perception locale de son environnement, le nœud sur lequel il est situé (ou les nœuds dont il a le contrôle) et l'ensemble des arcs connectés à ce ou ces nœud(s). Les actions définissent la capacité de se déplacer sur un arc ou de modifier la structure du graphe.

Enfin, ce travail s'intéresse aux agents délibératifs selon un modèle BDI (Belive, Desire, Intension) [7]. C'est à dire, un agent capable de se représenter son état dans l'environnement (Belive state) et capable de planifier ses actions (Intension) en fonction de ses objectifs (Desire).

L'état de l'agent sera à minima défini sur sa perception immédiate avec une combinatoire sur la profondeur de perception (le nombre de nœuds et les capacités des arcs qui les connectent). Au besoin, l'état de l'agent peut être enrichi d'une mémoire de certains événements passés (mais de nouveaux, avec une explosion combinatoire du nombre d'états). Par exemple, les agents ont à mémoriser les positions visitées dans le cadre du multi-voyageur de commerce.

Nos travaux partent du postulat qu'il existe une relation de probabilité sur l'évolution de l'état de l'agent en fonction des actions qu'il accomplit. Plus formellement, il est possible de modéliser le contrôle qu'a un agent sur le système, sur la base de processus de décision markovien (Markov Decision Process MDP) [3]. Le comportement de l'agent découle alors de la définition de ses désirs sous forme d'une fonction de récompense et d'une planification à horizon plus ou moins lointain.

### 2.2 Processus de décision Markovien

Les MDPs permettent de définir un système et ses possibilités d'évolution en fonction de la succession d'actions qui sera réalisée. Un MDP est défini par un tuple  $\langle S, A, t, r \rangle$

[[1, 6]] représentant respectivement les ensembles des états et des actions, les fonctions de transition et de récompense. La fonction de transition  $t$  est définie par  $t : S \times A \times S \rightarrow [0, 1]$  et donne la probabilité  $t(s, a, s')$  d'atteindre  $s'$  depuis  $s$  en exécutant l'action  $a$ . La fonction de coût/récompense  $r$  est définie par  $r : S \times A \rightarrow \mathbb{R}$ , où  $r(s, a)$  retourne la récompense obtenue en exécutant l'action  $a$  depuis  $s$  si  $r(s, a)$  est positif ou retourne le coût si  $r(s, a)$  est négatif. Trouver une solution optimale à un MDP consiste à chercher la politique optimale  $\pi^* : S \rightarrow A$  qui maximise les gains espérés sur les récompenses. Bellman [[1]] a proposé une fonction permettant une évaluation réursive des états par rapport à  $\pi$  ( $V^\pi : S \rightarrow \mathbb{R}$ ). La valeur d'un état  $V^\pi(s)$  est alors définie comme une somme pondérée de sa récompense immédiate après réalisation de l'action  $\pi(s)$  et des valeurs des états atteignables proportionnellement aux probabilités de les atteindre :

$$V^\pi(s) = r(s, \pi(s)) + \gamma \sum_{s' \in S} t(s, \pi(s), s') V^\pi(s') \quad (1)$$

Le paramètre  $\gamma \in [0, 1]$  pondère l'importance entre les récompenses immédiates et futures.

### 2.3 Modéliser un système de transport

Le modèle des MDPs permet une représentation discrète de l'évolution d'un système et s'applique aux systèmes multi-agents. Cependant, le modèle et sa résolution sont nativement centralisés. Les états et les actions expriment les conjonctions d'états et d'actions individuelles.

Le modèle le plus complet issu du formalisme des MDP est le Dec-POMDP pour Decentralized Partially Observable Markov Decision Process [2]. Les DecPOMDPs permettent d'exprimer un problème planification où l'état du système n'est pas directement observable et avec des observations et des actions qui s'effectuent de façon distribuée. Cependant, cette approche idéale, mais qui suppose une connaissance fine du système, n'est pas solvable pour des applications de taille réelle telle que pour des graphes de déplacement.

L'approche défendue ici, suppose d'effectuer le mécanisme inverse : Partir des capacités perceptives et d'action de chaque agent pour tenter de construire des MDPs simples et individuels, et compléter la solution avec des protocoles de coordinations.

### 2.4 Nos travaux récents

Concernant le multi-voyageur de commerce, les MDPs individuels permettent à chaque voyageur de planifier son trajet sur la base d'un environnement modélisé sous la forme d'un graphe de déplacement. Cette planification s'effectue sur la base d'un ensemble des positions que le voyageur sait devoir visiter ou pourrait ajouter à sa liste de destinations. Un protocole de vente aux enchères permet aux voyageurs de se répartir les destinations à visiter.

Cette approche a été appliquée avec succès en robotique mobile dans des missions d'explorations [5]. L'évaluation

des enchères découle directement des différences observées sur l'équation de Bellman en fonction des attributions possibles de destinations à visiter. Un mécanisme d'"altruisme" permet aussi à chaque agent de viser une répartition équitable.

Avec une telle approche, l'optimalité de la solution ne peut plus être garantie, cependant les ressources nécessaires pour la planification permettent à un groupe de voyageurs de commerce de replanifier à la volé, collectivement et ponctuellement leur mission. Sur les scénarios testés, les comportements du groupe ont démontré l'intérêt de l'approche et sa capacité à monter en charge.

Dans des travaux débutés plus récents, nous cherchons à appliquer des MDPs distribués pour la gestion de la ressource en eau dans un réseau fluviale. Ici, le problème consiste à planifier les échanges de volume d'eau contrôlables entre différents canaux de navigation d'un réseau [4]. La modélisation basée-agent la plus cohérente associe la responsabilité d'un ou plusieurs points de transfert à chaque agent. Chaque agent construit alors un MDPs individuel modélisant l'évolution possible de l'ensemble des canaux connecté à ses points de transfert.

Cette approche permet une modélisation cohérente de la gestion de flux dans un réseau tout en cassant la combinatoire qui serait liée à une modélisation centralisée sur un MDP unique. L'objectif à court terme vise à finaliser le protocole de coordination pour démontrer sa validité (convergence vers une solution) et la qualité des solutions produites.

## 3 Perspectives

Nous nous intéressons ici à la modélisation de réseau de transport. L'approche présentée est à l'intersection de la modélisation basée-agent et des processus de décision markovien et défini une notion d'agent markovien. C'est-à-dire des agents dont le comportement découle de tel processus. Le problème de la coordination d'agent markovien dans des réseaux de transport est abordé à travers le multi-voyageur de commerce et la gestion distribuée d'un graphe de flot. Les processus individuels s'appuient alors sur des protocoles de coordination traditionnels : vente aux enchères, vote, etc. pour coordonner les actions des agents.

L'objectif à plus long terme vise à dépasser les cadres applicatifs définis pour proposer un framework plus générique. Nous souhaiterions aussi adresser d'autre problématique comme la construction de connaissance individuelle par des processus collectifs, toujours sous le prisme des processus de décision markovien distribué.

## Références

- [1] R. Bellman. A Markovian Decision Process. *Journal of Mathematics and Mechanics*, 6 :679–684, 1957.
- [2] D. S. Bernstein, S. Zilberstein, et N. Immerman. The Complexity of Decentralized Control of Markov Decision Processes. In *16th Conference on Uncertainty in Artificial Intelligence*, 2000.

- [3] O. Buffet et O. Sigaud. *Processus décisionnels de Markov en intelligence artificielle*. Lavoisier, hermes science edition, 2008.
- [4] G. Desquesnes, G. Lozenguez, A. Doniec, et E. Duvieux. MDP à grande échelle, étude de cas des voies navigables. In *Advances on Practical Applications of Agents and Multiagent Systems*, Advances in Intelligent and Soft Computing. Springer Berlin / Heidelberg, 2016.
- [5] G. Lozenguez, L. Adouane, A. Beynier, A.-I. Mouadib, et P. Martinet. Punctual versus continuous auction coordination for multi-robot and multi-task topological navigation. *Autonomous Robots*, 40(4) :599–613, 2015.
- [6] M. L. Puterman. *Markov Decision Processes : Discrete Stochastic Dynamic Programming*. John Wiley & Sons, Inc., 1994.
- [7] A. S. Rao, M. P. Georgeff, et al. Bdi agents : From theory to practice. In *ICMAS*, volume 95, pages 312–319, 1995.