

Processus décisionnel partiellement observable pour le contrôle des activités d'un robot en présence d'un humain

Partially observable decision process for controlling robots near humans

Abir-Béatrice Karami

Laurent Jeanpierre
GREYC-CNRS/Université de Caen
Boulevard Maréchal Juin, BP5186
14032 Caen cedex

Abdel-Ilah Mouaddib

{akarami@etu.info.unicaen.fr, {mouaddib,laurent}@info.unicaen.fr}

Résumé

Nous présentons dans cet article une nouvelle architecture de contrôle d'un robot d'assistance en présence d'un humain. Le robot doit réaliser des activités de service en interaction avec l'homme. La réalisation de ces activités devient complexe car les informations que possède le robot sur les intentions de l'homme sont incomplètes. Nous considérons dans un premier temps une interaction via l'environnement où le robot observe l'homme et construit des croyances sur ses intentions. À partir de ces croyances, il construit des politiques d'actions. Nous avons formalisé ce problème par un Processus Décisionnel de Markov Partiellement Observable (POMDP) pour la réalisation de quelques tâches simples comme se diriger vers les lieux souhaités par l'homme ou distribuer des objets à des personnes en s'adaptant à leurs réactions et en révisant les croyances qu'on a sur leurs intentions. Le robot doit prendre des décisions concordantes avec le comportement de l'humain. Il n'existe aucun plan partagé entre le robot et l'humain mais seulement des croyances qu'a le robot sur l'homme. L'état du robot encapsule à la fois son état interne que nous supposons complètement observable et un état de croyance sur les intentions de l'homme. À partir de cette représentation, nous proposons un POMDP augmenté qui permet de dériver une politique autorisant le robot à se coordonner avec l'homme. Un scénario expérimental qui consiste en l'aide proposée par un robot à transporter des objets vers un lieu connu en présence d'un homme est mis en place. Le robot ne connaît pas les objets que va porter l'humain mais s'adapte à son mouvement. Le résultat est prometteur et nous ouvre des perspectives intéressantes comme la construction de décisions jointes entre homme et robot. Aussi, nous mettrons en oeuvre ce modèle dans le simulateur OpenHRP du robot humanoïde HRP2.

Mots Clef

Processus décisionnel partiellement observable, Interaction Homme-Robot

Abstract

We present in this paper a new architecture for controlling a robot in the presence of a Human. The operation of the robot is performed in interaction with the human. The robot-human interaction makes this decision process more complex because of partial information the robot can have on the skills, state and intention of the human. We formalize this problem with a decision process with incomplete information that leads to Partially Observable MDP (POMDP) controller. The local decision of the robot should consider its effects on the relationship with the human. In this paper, we consider only some simple situations where the robot has to adapt his behavior to the human being's actions. There is no explicit shared plan between the human and the robot but they have a common mission to accomplish. Human and robot do not share in advance tasks to complete but the robot has to develop a policy allowing him to make the most appropriate decision according to the behavior of the human. The robot observes, in a permanent way, the behavior of the human, constructs beliefs and it makes a decision based on its belief state. The state of the robot is a factored state representing a fully observable internal state of the robot (position, resources, ...) and a partially observable state on the skills and intention of the human. The augmented POMDP we propose allows the robot to derive a policy which provides a behavior of the robot coordinated with the human. Experiments show promising results which open interesting perspectives such as joint decision making of humans and robots. Perspectives of this work are to consider other situations such as having a joint plan and developing a joint policy. We also expect to evaluate our model in OpenHRP simulator of the robot HRP-2.

Keywords

POMDP, Interaction Human-Robot

1 Introduction

L'utilisation avec succès des robots dans plusieurs domaines comme le transport, la surveillance, l'exploration et autres [13, 8, 3] permet de développer de nouvelles techniques pour aborder de nouveaux secteurs. Ainsi, les secteurs de la robotique en plein essor concernent la robotique de service, de loisir ainsi que l'assistance. Ces derniers se caractérisent par l'introduction de robots dans des environnements où l'homme est présent. La présence de l'homme dans l'environnement où évolue le robot ajoute une contrainte supplémentaire aux contraintes classiques du processus de décision d'un robot. Ainsi, la planification de tâches d'un robot doit tenir compte non seulement des contraintes classiques de l'environnement mais aussi des informations sur l'homme et son environnement. Ce problème devient crucial lorsque le robot et l'homme doivent se partager les tâches à réaliser ou réaliser une tâche en synergie. Dans de tels contextes, le robot doit décider quelle tâche réaliser et comment la réaliser de façon coopérative avec l'homme. Plusieurs approches ont été proposées pour la mise en place d'une architecture d'interaction homme-robot avec différentes hypothèses [1, 7, 10]. La plupart de ces approches proposent différentes techniques de planification ou de modalités d'interaction. Notre approche complètera ces travaux en proposant un modèle décisionnel avec information incomplète en tenant compte de la présence de l'homme et en coordination avec ses activités. Pour cela, nous proposons un modèle qui permet au robot d'observer l'homme, de construire un modèle de ses intentions et de prendre la décision la plus appropriée.

Pour cela, nous présentons un système décisionnel avec informations incomplètes qui permet au robot de construire des décisions adéquates et compatibles avec les intentions supposées de l'homme, mais aussi satisfaisantes pour le robot. Ce système se fonde sur les informations récoltées par diverses modalités, que nous réduisons dans notre article à de simples observations sur l'environnement, à partir desquelles le robot construit un réseau de croyances sur l'intention de l'homme afin de prendre une décision.

Nous proposons pour cela un formalisme fondé sur les processus décisionnels partiellement observables (POMDP) [6, 2] où l'état du robot encapsule son état interne, complètement observable, et un état de croyance sur l'homme. Le modèle proposé peut ainsi être résolu par l'un des algorithmes classiques de résolution des POMDP comme les algorithmes dans [11, 4, 12]. Ce modèle présente ainsi un premier cadre pour l'interaction homme-robot mais aussi pour un groupe de robots en présence des humains. En effet, des formalismes comme ceux présentés dans [9] doivent être adaptés et augmentés en se fondant sur notre modèle. Une autre caractéristique de ce problème est le fait que l'environnement change à cause des actions de l'homme. Cette contrainte nous a conduit à considérer des POMDPs sur des horizons à très court terme afin de pouvoir s'adapter aux évolutions de l'homme et de l'environnement. Ainsi, notre approche se fonde sur la construction

d'un POMDP à court terme, en ligne, suivie d'une prise de décision. Ce processus sera répété jusqu'à la réalisation de toutes les tâches.

Nous illustrons ce formalisme dans des exemples simples de travail collaboratif entre le robot et l'humain. Dans l'article nous considérons l'exemple comprenant un robot et un humain qui transportent des objets répartis dans l'espace pour les déposer dans un entrepôt. Cependant, le robot ne connaît ni les objets que l'humain va porter, ni sa stratégie de mouvement, ce qui complexifie la décision du robot pour savoir vers quel objet se diriger et quel objet transporter. Cet exemple illustratif sera détaillé ainsi que sa formalisation par notre modèle.

La suite de l'article contient quatre sections. Dans la section 2, nous décrivons l'exemple illustratif et ses propriétés. Dans la section 3 sera présenté le modèle proposé. Quelques résultats préliminaires seront présentés et discutés dans la section 4 et enfin nous allons conclure dans la section 5.

2 Un exemple introductif

Dans cette section nous décrivons un exemple pour illustrer une interaction homme-robot dans un environnement partiellement observable.

Cet exemple inclut un humain, un robot, un certain nombre d'objets et une boîte [FIG.1]. L'objectif est de ranger ces objets dans la boîte, chaque acteur (le robot et l'humain) ne pouvant porter qu'un seul objet à la fois.

L'objectif du robot est d'apporter une aide à l'humain sans devenir une gêne pour celui-ci. On suppose que l'humain désire toujours déplacer l'objet le plus proche de lui. Lorsque le robot doit prendre une décision, il doit éviter tout conflit et aider l'être humain autant que possible.

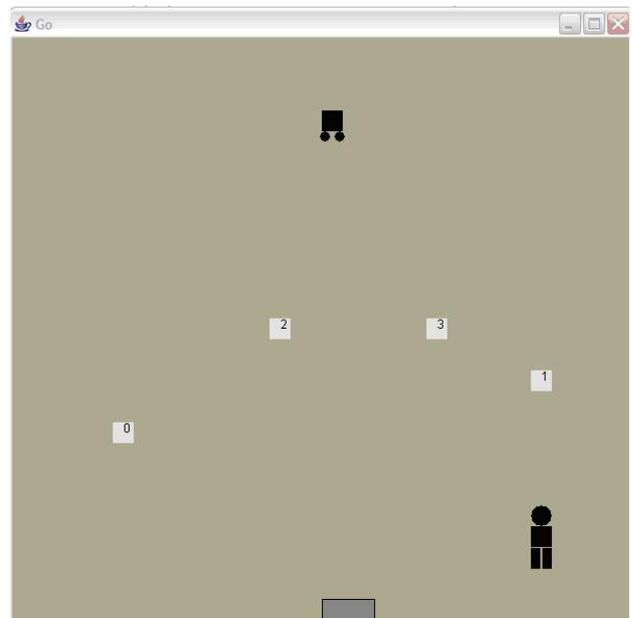


FIG. 1: L'environnement considéré pour l'exemple : un humain, un robot, 4 objets et une boîte.

Pour y arriver, il décidera donc de déplacer l'objet le plus loin de l'humain. Dans cet exemple, nous considérons que le robot et l'homme ne se partagent pas les tâches (porter des objets) au préalable. Le robot doit donc développer des croyances, à chaque instant, sur les objets que l'humain va porter et prendre une décision pour aider au mieux l'humain à accomplir leur tâche commune.

Cet exemple a pour intérêt de voir comment ce type de problème où l'homme et le robot se partagent un environnement et doivent accomplir une mission ensemble peut être formalisé. On peut envisager ce type de scénario dans plusieurs applications réelles comme un robot déménageur avec des hommes, un robot domestique, un robot apportant une assistance à une personne (personne âgée), ...

Afin de proposer un modèle à ce problème, nous essayons de définir les éléments essentiels à sa mise en place. Pour cela, nous avons décidé de représenter le monde par l'état de croyance représenté dans la figure [FIG.2] qui contient à la fois des éléments sur l'état interne du robot mais aussi les croyances sur les intentions de l'homme. Ces croyances sont mises à jour à chaque étape du mouvement, en fonction de la dernière action effectuée par le robot et conformément à ses observations.

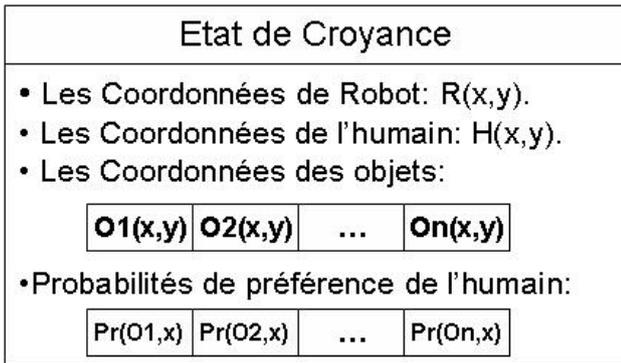


FIG. 2: État de croyance

Les intentions supposées de l'humain sont représentées par une distribution de probabilités sur les objets qui peuvent être choisis et déplacés par l'homme à l'instant suivant. Dans notre cas, nous considérons un modèle simple de ces probabilités. Ceci n'affecte pas la généralité du modèle. Ces probabilités sont estimées par le robot en se basant sur la distance entre l'homme et chaque objet. De cette façon, nous considérons que l'humain préfère toujours déplacer l'objet le plus proche de lui. Dès que l'humain est proche de plusieurs objets, les croyances du robot sont une distribution de probabilités sur ces objets. Cette supposition pourrait être erronée, mais la mise à jour fréquente de l'état de croyance permet au robot d'adapter ses probabilités de manière rapide si l'homme ne se comporte pas comme prévu. Cette révision de croyance sera fondée sur une simple règle de Bayes. Le robot observe uniquement les mouvements de l'être humain, mais il connaît la position de chaque objet à l'avance. À chaque instant, le robot

observe, met à jour ses croyances et, selon son nouvel état de croyance, il choisit quel objet il doit déplacer : l'objet qui a la plus basse probabilité d'être choisi par l'humain. Afin de formaliser ces idées intuitives autour d'un modèle tout en offrant une certaine robustesse, nous proposons une formalisation fondée sur les processus décisionnels de Markov partiellement observables (POMDP). La section suivante introduira ce formalisme et comment il a été utilisé dans notre problème.

3 Le modèle formel de décision, à base de POMDP

Dans cette section nous présenterons une modélisation de ce type de problèmes à l'aide du modèle des processus décisionnels de Markov partiellement observables, ou POMDP [4] (Partially Observable Markov Decision Processes.)

Pour cela, nous commencerons par un rappel rapide du formalisme, avant de présenter notre modèle. Nous terminerons par la modélisation de l'exemple présenté dans la section précédente.

3.1 Le formalisme général

Un POMDP est un tuple $\langle S, A, \Omega, T, O, R \rangle$:

S Un ensemble d'états s_i décrivant les configurations possibles du système modélisé.

A Un ensemble d'actions a_i permettant de modifier l'évolution du système.

Ω Un ensemble d'observations ω_i décrivant ce que l'on peut percevoir du système.

T Une fonction de transition $S \times A \rightarrow \mathcal{P}(S)$ indiquant la distribution de probabilités des états atteignables lorsque l'action a_j est employée dans l'état s_i .

O Une fonction d'observation $S \times A \times S \rightarrow \mathcal{P}(\Omega)$ indiquant la distribution de probabilités d'observer chaque symbole lorsque le système est dans l'état s_i et que l'action a_j est employée.

R Une fonction de récompense $S \times A \times S \rightarrow \mathbb{R}$ qui rétribue l'utilisation de l'action a_j dans l'état s_i lorsque son effet est d'arriver dans l'état s_k .

Ce modèle repose sur l'hypothèse de Markov, que l'on peut résumer par "Le futur ne dépend que de l'état présent". Mathématiquement, cela se traduit par les deux relations suivantes :

$$P(s_{t+1}|s_t, a_t, s_{t-1}, a_{t-1}, \dots, s_0, a_0) = P(s_{t+1}|s_t, a_t)$$

$$P(\omega_{t+1}|s_t, a_t, s_{t-1}, a_{t-1}, \dots, s_0, a_0) = P(\omega_{t+1}|s_t, a_t)$$

3.2 Notre modèle

Maintenant que le modèle générique des POMDP a été présenté, nous allons nous intéresser à son utilisation pour modéliser les problèmes de collaboration Homme-Robot auxquels nous nous intéressons plus particulièrement dans cet article.

Nous rappelons que la contrainte principale pour notre robot consiste à apporter son aide à un être humain sans devenir une gêne pour ce dernier.

A cette fin, il est nécessaire d'anticiper les réactions de l'homme, afin de décider d'une action non-conflictuelle aidant à la résolution de leur but commun. Pour obtenir cet effet sans augmenter démesurément la complexité du calcul, nous proposons de subdiviser l'espace d'états en deux parties : l'état de l'environnement S_E et l'intention supposée de l'homme B_H .

Afin de simplifier le modèle, nous supposons que l'état de l'environnement est observable. Par contre, l'intention de l'être humain n'est pas observable directement. On ne peut que faire des suppositions basées sur ce que l'on observe de son comportement. Nous supposons de plus que l'évolution des intentions de l'être humain sont indépendantes des actions du robot.

Les intentions de l'humain n'étant pas connues avec exactitude, nous introduisons la notion de *belief-state* qui représente un état de croyance sur ses intentions. Ce belief-state est une distribution de probabilités discrète, telle que chacun de ses éléments représente une intention possible pour l'homme. Comme toute distribution de probabilités, la somme de tous ses éléments doit être égale à 100%.

Notre modèle d'interaction sera donc finalement composé du tuple $\langle S = (S_E, B_H), A, \Omega, \Theta, (T_E, T_H), O, R \rangle$:

Θ L'ensemble des intentions θ_i possibles pour l'homme.

S_E Les configurations possibles de l'environnement.

B_H Un état de croyance $\mathcal{P}(\Theta)$.

A Un ensemble d'actions a_i .

Ω Un ensemble d'observations ω_j .

T_E Une fonction de transition $S \times A \rightarrow \mathcal{P}(S_E)$.

T_H Une fonction de transition $S \rightarrow \mathcal{P}(\Theta)$.

O Une fonction d'observation $\Theta \times \mathcal{P}(\Omega)$.

R Une fonction de récompense $S \times A \times S \rightarrow \mathbb{R}$.

Dans ce modèle, la fonction de récompense va donc être une combinaison de deux choses : Atteindre l'objectif fixé et ne pas gêner l'être humain. Dans le cas où le robot déciderait d'une action conflictuelle (en fonction de sa croyance sur les intentions de l'homme), cette récompense doit donc être négative. Par contre, lorsque le robot ou l'homme réalisent une partie de la tâche qui leur est assignée, la récompense doit être positive.

La fonction de transition T_E est une fonction classique de la robotique, puisqu'elle représente la dynamique de l'environnement en réponse aux actions effectuées. Il est à noter que l'être humain agissant de façon indépendante, l'environnement peut être modifié sans rapport direct avec les actions du robot. Cette fonction dépend donc de l'action choisie par le robot, mais aussi des intentions supposées de son partenaire humain. Par exemple, si l'homme a deux possibilités, aller au nord et à l'est, deux états résultants différents sont possibles pour une même action du robot. Dans cette situation, la distribution de probabilité fournie

par T_E devra couvrir ces deux cas. Leurs probabilités respectives dépendent des intentions supposées de l'homme.

La fonction de transition T_H représente l'évolution "normale" des intentions de l'homme. C'est donc certainement la partie la plus difficile à évaluer, puisque ce dernier peut changer d'avis à tout moment. Cependant, on peut espérer qu'il conserve une certaine constance, et qu'il agisse avec logique. Par exemple, s'il est sur le point d'accomplir une partie de l'objectif, on peut supposer qu'il le fera.

Enfin, la fonction d'observation indique les observations qui sont attendues, en supposant que les intentions de l'homme aient été correctement devinées. Dans notre cas, cela se traduit le plus souvent par des actions qu'il entreprend.

À partir de ces fonctions, il est possible de mettre à jour le belief-state à partir des observations effectuées en utilisant la loi de Bayes :

$$\begin{aligned} P(\theta^t | \omega^t, s^{t-1}) &= \frac{P(\omega^t | \theta^t) P(\theta^t)}{P(\omega^t)} \\ &= \frac{P(\omega^t | \theta^t) \sum_i P(\theta^t | \theta_i^{t-1}) P(\theta_i^{t-1})}{\sum_j P(\omega | \theta_j)} \\ &= \frac{O(\theta^t, \omega^t) \sum_i T_H(s^{t-1}, \theta^t)}{\sum_j O(\theta_j, \omega^t)} \cdot B_H^{t-1}(\theta_i^{t-1}) \end{aligned}$$

Cette équation permet de calculer le belief-state $B_H(\theta)$. À partir de cette croyance, nous pouvons calculer la fonction de valeur optimale $V^*(s)$ en utilisant l'équation de Bellman :

$$\begin{aligned} V_a^*(s) &= \sum_{s'} B_H(\theta | s) T(s, a, s') (R(s, a, s') + \gamma V^*(s)) \\ V^*(s) &= \max_{a \in A} V_a^*(s) \end{aligned}$$

La politique optimale du robot est alors donnée par :

$$\pi^* = \operatorname{argmax}_{a \in A} V_a^*(s)$$

Plusieurs algorithmes dans la littérature [11, 4, 12] existent qui permettent de dériver cette politique et que l'on peut utiliser immédiatement, même si notre perspective est de fournir un autre algorithme plus efficace pour mieux tenir compte de la spécificité du problème abordé.

3.3 Exemple : Le rangement d'objets

Afin d'explicitier notre modèle, nous allons reprendre et détailler l'exemple dans lequel un homme et un robot doivent ranger des objets dans une boîte.

Afin de simplifier la présentation, on se limitera au rangement de deux objets.

Dans ce cas, notre modèle sera le suivant :

$\Theta = O_1, O_2$ où chaque O_i représente pour l'homme l'intention de transporter l'objet i .

$S_E = \{P_r, P_h, P_b, P_1, P_2\}$ où les P_* correspondent respectivement à la position du robot, de l'homme, de la boîte, du premier objet et du second objet. Chacune de ces variables peut correspondre à un jeu de coordonnées dans un espace continu de positions, ou au numéro d'une telle position dans un espace discret, un damier ou un pavage hexagonal par exemple.

$B_H = \mathcal{P}(\Theta)$ correspondant aux désirs estimés de s'emparer des différents objets.

$A = \{D_1, D_2, D_b\}$ où le robot se dirigera respectivement vers l'objet 1, l'objet 2 ou la boîte de rangement.

$\Omega = \{O_1, O_2, O_b\}$ avec O_1, O_2 et O_b observés respectivement lorsque l'homme se dirige vers l'objet 1, l'objet 2 ou la boîte de rangement.

T_E Cette fonction sera définie par cas :

- Si le mouvement doit entraîner une collision avec l'homme, rester sur place.
- Si le robot est seul, modifier P_r selon l'action qui a été choisie.
- Si le robot est sur un objet, modifier la position de l'objet et du robot selon l'action choisie.
- Si le mouvement amène un objet dans la boîte, retirer l'objet du monde (on utilisera par exemple une position non atteignable.)
- Si l'humain est seul, modifier P_h en direction de l'objet désiré, avec une faible probabilité de se diriger dans une autre direction.
- Si l'humain transporte un objet, modifier P_h et la position de l'objet transporté vers P_b , avec une faible probabilité d'aller dans une autre direction.

T_H Cette fonction sera définie par cas également :

- Si l'homme ne transporte pas d'objet, équilibrer légèrement le belief-state courant afin de prendre en compte un éventuel changement d'avis. Par exemple, on passera de (30%; 70%) à (40%; 60%).
- Si l'homme transporte un objet, mettre la probabilité de cet objet à 90% et la probabilité de choisir l'autre objet à 10%. Bien sûr, si l'humain est connu pour être lunatique, on peut équilibrer ces probabilités un peu plus.

O Cette fonction indiquera le comportement normal d'une personne en fonction de ses intentions :

- Si l'homme désire transporter un objet particulier, il est probable (90%) qu'il se dirige vers celui-ci.
- Si l'homme transporte un objet, il est probable (90%) qu'il se dirige vers la boîte.
- Dans 10% des cas, l'homme se déplacera aléatoirement.

Ainsi, si par exemple les intentions supposées sont de 70% pour le premier objet, et de 30% pour le second : $P(O_1) = 65%$, $P(O_2) = 30%$ et $P(O_b) = 5%$.

R Cette fonction sera elle aussi définie par cas :

- Si un objet arrive dans la boîte, $R = 100$.
- S'il y a collision avec l'homme, $R = -1000$.

- Si le robot transporte l'objet désiré par l'homme, $R = -10$.

Bien sûr, ce modèle n'est donné qu'à titre d'exemple ; on peut ajuster les probabilités à volonté. Par exemple, on pourrait augmenter ou réduire la pénalité de -10 afin de modifier le "désir" du robot de ne pas déplaire à l'homme.

4 Résultats préliminaires

Dans cette section, nous présentons comment le modèle introduit dans la section 3.3 a été utilisé dans la production des résultats ci-dessous montrés dans la figure 5 sur l'exemple de la section 2. En effet, le modèle $\langle S_E, B_H, A, \Omega \rangle$ a été utilisé comme introduit dans la section 3.3.

Cependant, quelques adaptations et simplifications ont été faites sur le modèle : T_H et O n'ont pas été utilisées, car une fonction ad hoc les remplace en jouant le même rôle, puisqu'il n'y a aucune action "informative", c'est à dire modifiant la fonction d'observation, et pouvant apporter des connaissances nouvelles. La fonction R de récompense a été définie et utilisée d'une manière implicite. Enfin, seule une étape de Value Iteration a été calculée.

Nous présentons ci-dessous, dans les figures FIG.3 et FIG.4, quelques exemples des états de croyance du robot et la décision prise, telle que l'indique la politique générée.

Etat de croyance
Robot (100,100) Humain(0,0) Objets : $O_1(50, 75), O_2(50, 25)$
Préférences de l'humain : $Pr(O_1) = 38.38\%, Pr(O_2) = 61.62\%$
Décision : déplacer l'objet 2

FIG. 3: L'état de croyance du robot et sa décision selon ses observations pour une bonne coordination avec l'humain.

Etat de croyance
Robot (100,100) Humain(0,100) Objets : $O_1(50, 75), O_2(50, 25)$
Préférences de l'humain : $Pr(O_1) = 61.72\%, Pr(O_2) = 31.28\%$
Décision : déplacer l'objet 1

FIG. 4: L'état de croyance du robot et sa décision selon ses observations pour une bonne coordination avec l'humain.

Nous montrons maintenant un exemple décrivant la mise à jour des croyances en utilisant les observations du robot ; finalement nous montrerons les résultats obtenus avec le modèle simplifié expliqué ci-dessus. Quelques copies d'écrans de la démonstration (Figure [FIG. 5]) montrent le comportement obtenu par notre modèle pour le robot, et sa capacité à s'adapter à l'homme. Les étapes sont les suivantes :

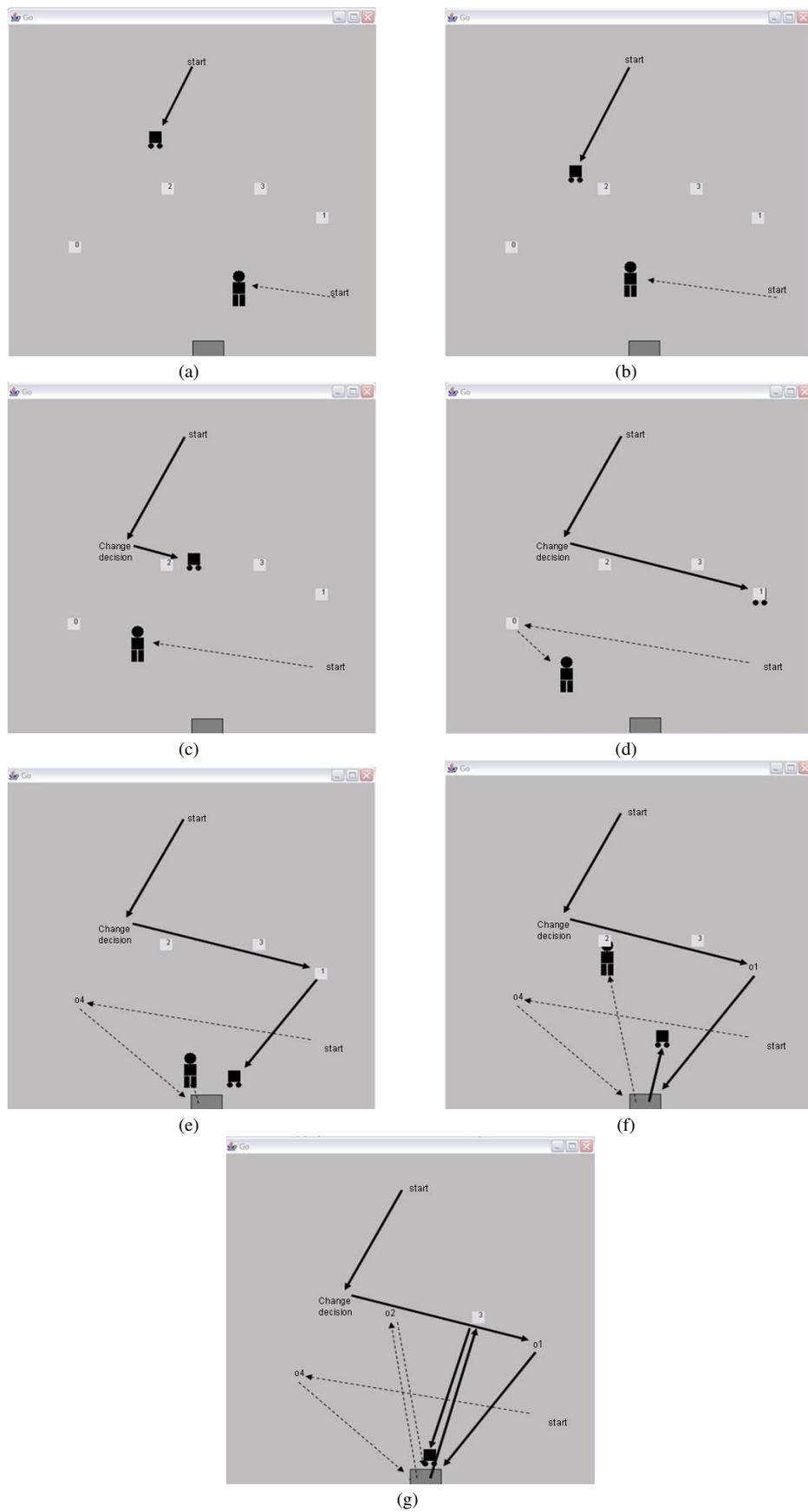


FIG. 5: Un exemple illustratif du modèle dans la résolution d'une tâche collaborative entre robot et humain

- **étapes (a,b)** : La croyance du robot est que l’homme va se diriger vers l’objet le plus proche (les objets 3 ou 1). Il prend donc la décision de se diriger vers l’objet 0 qui est le plus loin de l’homme. On signale au passage que le robot a pour objectif de rendre la tâche la moins pénible possible pour l’homme. Il se dirige donc toujours vers l’objet le plus loin pour l’homme.
- **étapes (c,d,e)** : Le robot observe que l’homme est plus proche des objets 0 et 2 et donc ses croyances deviennent que l’homme prendra l’objet 0 ou 2. Il change donc sa décision d’aller vers l’objet 0 pour aller vers l’objet 1 (le plus loin). Le robot s’adapte en ligne au comportement de l’homme. Pour qu’il y ait convergence, nous considérons que l’homme a un comportement rationnel, ou au minimum qu’il ne change pas d’avis continuellement.
- **étapes (f,g)** : Le robot observe l’homme qui se dirige vers l’objet 2 et construit une croyance sur le fait que l’homme va prendre l’objet 2 et donc prend la décision d’aller vers l’objet 3.

Cet exemple nous montre comment le robot est capable de s’adapter au comportement de l’humain en mettant à jour en permanence ses croyances sur les intentions de l’homme et en construisant la décision adéquate. Le robot est capable aussi de changer de décision si ses croyances évoluent comme dans l’étape (c) de la figure [FIG. 5].

Dans cet exemple, nous pouvons constater que le robot s’adapte très bien au comportement de l’humain. Il adapte à chaque fois son aide par rapport à ce que fait l’humain. Cependant, des hypothèses simples faites sur cet exemple doivent être relâchées comme l’évaluation du degré de satisfaction de l’aide apportée à l’homme puis une mise à jour de l’utilité de l’aide. Nous étudierons à ce moment là des politiques d’aide satisfaisantes et optimales ainsi qu’une comparaison entre les deux comportements, sachant que l’homme cherche plus à satisfaire qu’à optimiser. De plus, nous souhaitons considérer des tâches complexes où le robot et l’homme doivent agir simultanément. Pour cela, le formalisme doit être enrichi pour prendre en compte les décisions de l’homme et non seulement observer leurs conséquences comme c’est le cas actuellement.

5 Conclusions et perspectives

Dans cet article, nous avons abordé le problème de la prise de décision d’un robot en présence d’un homme en considérant des situations simples d’interaction. En effet, nous considérons que seul le robot est concerné par l’adaptation de son comportement à l’humain et que ce dernier se comporte d’une manière inconnue et indépendante de la présence du robot. Par conséquent, nous avons proposé un formalisme pour le processus décisionnel du robot qui ne dispose que d’informations incomplètes sur les intentions de l’homme. Le processus décisionnel du robot que nous avons considéré consiste en une phase d’observation qui permet au robot de mettre à jour ses croyances et de construire une décision.

Nous avons proposé un formalisme fondé sur un MDP aug-

menté par les croyances qu’a le robot sur l’homme. Ensuite, nous avons montré comment ce problème peut être formalisé. Le modèle obtenu est un POMDP que nous pouvons résoudre avec n’importe quel algorithme de résolution existant dans la littérature. Ce formalisme à été mis en oeuvre dans le cadre d’un scénario simple où l’humain et le robot doivent transporter des objets vers une caisse, sans expliciter un comportement global à l’avance. Le robot doit adapter son comportement (choix de l’objet à transporter) en fonction de ce que fait l’homme (l’objet qu’il va transporter). Le robot observe l’homme, développe une croyance sur l’objet que l’homme va prendre ensuite, et décide de l’objet qu’il va prendre en essayant d’apporter la meilleure aide possible à l’homme et de rendre sa tâche la moins pénible possible. Les expérimentations sont concluantes et plusieurs améliorations sont possibles, ouvrant des perspectives intéressantes. Dans un futur proche, nous poursuivrons ces expérimentations sur le simulateur du robot humanoïde HRP2, puis sur la plate-forme réelle installée à Toulouse pour une expérimentation grandeur nature.

Une des premières remarques que nous pouvons faire à partir des expérimentations est que le formalisme POMDP est approprié à ce problème mais que l’on retrouve les limites classiques de ce formalisme, qui résident dans son incapacité de travailler sur des exemples de très grande taille malgré les efforts récents dans l’amélioration des algorithmes de résolution. En effet, les algorithmes classiques poussent à discrétiser les variables d’état, ici la position de chaque agent et de chaque objet. Ceci se traduit donc naturellement par une augmentation du nombre d’états exponentielle en fonction du nombre de variables.

Une des voies explorées actuellement dans notre approche est de construire des POMDP à horizon réduit, voire à une seule étape de décision pour s’adapter à la fois au problème de la taille de l’espace d’états de croyances et à la dynamique de l’environnement, qui change assez fréquemment et donc qui peut rendre la politique précédente caduque. Actuellement, nous étudions un formalisme de construction dynamique de la politique sur la base de morceaux de politiques construits pour des situations simples [5].

Une autre approche permettant d’améliorer nos travaux s’intéresse à la notion d’observation. Alors que nous avons utilisé dans l’exemple développé ici la simple mesure de distance entre l’homme et ses sous-objectifs possibles, il semblerait avantageux de s’intéresser plutôt à la variation de celle-ci. Mieux encore, on pourrait évaluer la pertinence des observations vis à vis de diverses politiques possibles pour l’homme. Ceci fera l’objet de nos travaux futurs.

Une autre direction que nous souhaitons aborder est l’enrichissement du modèle pour prendre en compte différents types d’interactions entre l’homme et le robot. Nous souhaitons étudier et enrichir notre modèle afin de considérer des situations où une interaction explicite entre l’homme et le robot existe via différentes modalités, comme l’interaction langagière, émotionnelle ou via un terminal.

Alternativement, considérer des situations où l'homme s'adapte aux choix du robot semble très intéressant ; cependant ceci laisse entrevoir de gros problèmes, tels que ceux rencontrés dans le co-apprentissage d'un système multi-agents.

Différents domaines d'applications sont possibles, comme des tâches d'assistanat ou d'aide aux personnes âgées. À long terme, nous souhaitons mettre en oeuvre quelques scénarios d'aide aux personnes âgées ou de soutien aux aide-soignantes dans une maison de retraite.

Références

- [1] R. Alami, V. Sisbot R. Clodic, A. Montreuil, and Chatila. Task planning for human-robot interaction. In *Soc-EUSAI*, 2005.
- [2] C. Boutilier, T. Dean, and S. Hanks. Decision-theoretic planning : Structural assumptions and computational leverage. *Journal of Artificial Intelligence Research*, 1 :1–93, 1999.
- [3] J. Bresina, R. Dearden, N. Meuleau, S. Ramakrishnan, D. Smith, and R. Washington. Planning under continuous time and resource uncertainty : A challenge for AI. In *UAI*, 2002.
- [4] Anthony Cassandra, Michael L. Littman, and Nevin L. Zhang. Incremental pruning : A simple, fast, exact algorithm for partially observable Markov decision processes. In *UAI-97*, 1997.
- [5] G. Dibangoye, A.-I. Mouaddib, and B. Chaib-Draa. Periodic real-time resource allocation for teams of progressive processing agents. In *AAMAS*, pages 768–770, 2007.
- [6] Leslie Pack Kaelbling, Michael L. Littman, and Anthony R. Cassandra. Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, 101 :99–134, 1998.
- [7] Kulic and E. Croft. Safe planning for human-robot interaction. In *IEEE ICRA*, 2004.
- [8] A.-I. Mouaddib and S. Zilberstein. Optimal scheduling for dynamic progressive processing. In *ECAI-98*, pages 499–503, 1998.
- [9] D. Pynadath and M. Tambe. The communicative multiagent team decision problem : Analyzing teamwork theories and models. *Journal of Artificial Intelligence Research*, pages 389–423, 2002.
- [10] P. Scerri, D. Pynadath, L. Johnson, L. Paul, R. Schurr, N. Si, and M. Tambe. A prototype infrastructure for distributed robot-agent-person teams. In *AAMAS*, 2003.
- [11] E.J. Sondik. The optimal control of partially observable Markov decision process over finite horizon. In *Operational Research*, pages 24 :282–304, 1978.
- [12] W. Zhang and N.L. Zhang. Value iteration working with belief states. In *AAAI*, pages 307–403, 2002.
- [13] S. Zilberstein and A.-I. Mouaddib. Optimizing resource utilization in planetary rovers. In *2nd NASA Workshop on Planning and Scheduling for Space*, pages 163–168, 2000.