Un modèle de décision distribué pour la collecte d'information active multiagents

Jennifer Renoux, Abdel-Illah Mouaddib, Simon Le Gloannec

Un modèle de décision distribué pour la collecte d'information active multiagents

Jennifer Renoux[1] Abdel-Illah Mouaddib[1] Simon Le Gloannec[2]

[1] GREYC, Université de Caen Basse Normandie, France
[2] Airbus Defense and Space

{jennifer.renoux, abdel-illah.mouaddib}@unicaen.fr - simon.legloannec@cassidian.com

Résumé

Les systèmes multirobots ont permis de grandes avancées dans les domaines de l'exploration et de la surveillance. Néanmoins, les agents se contentent généralement de percevoir leur environnement passivement. Dans cet article, nous proposons un modèle permettant de récupérer l'information de manière active : les agents explorent, mettent à jour leurs croyances, évaluent la pertinence des informations qu'ils collectent et les communiquent aux autres agents. Ce modèle se base sur un processus de décision distribué où chaque robot met à jour une matrice représentant ses croyances sur l'environnement ainsi que sur les croyances des autres agents. Ce processus de décision utilise une fonction de récompense convexe basée sur l'évaluation de la pertinence d'une information ainsi que sur les divergences de croyances entre agents. Ce modèle permet de calculer une politique d'exploration visant à améliorer la recherche d'information en évitant les explorations redondantes et en communiquant. Un scénario expérimental a été développé pour des missions de recherche d'information en intérieur.

Mots Clef

Recherche d'information active, Pertinence orientée agent, Processus de décision décentralisé, Système multi-agents

Abstract

Multirobot systems have made tremendous progress for exploration and surveillance. However, information gathering tasks remains passive.In this paper, we present a model and an algorithm for active information gathering. In this model, robots explore, assess the relevance, update their believes and communicate the appropriate information to relevant robots. To do so, we propose a distributed decision process where an agent maintains a beliefs matrix representing its beliefs and beliefs on the beliefs of the other agents. This decision process uses a convex reward function to assess the relevance of their beliefs and the divergence with each other. This model allows the derivation of a policy to improve active information gathering by avoiding redundant exploration and communicating. An experimental scenario has been developed for an indoor information gathering mission.

Keywords

Active Information Gathering, Information Relevance, Decentrelized Decision Process, Multiagent System

1 Introduction

Les systèmes robotisés sont de plus en plus utilisés dans les applications de surveillance et d'exploration. Dans le futur, les robots pourront assister voire remplacer les opérateurs humains dans les zones les plus dangereuses afin de collecter de l'information. Cette collecte d'information est alors le but principal du système : celui-ci doit créer une vue complète et précise de la situation. Ainsi, le système doit identifier l'information manquante et effectuer les actions nécessaires à sa collecte. Cependant, il n'est évidemment pas souhaitable que tous les robots du système collectent chacun toute l'information possible et créent ainsi une redondance. De même, dans des applications réelles, les robots ne peuvent généralement pas échanger à chaque instant toute l'information dont ils disposent pour des raisons physiques, telles qu'une bande passante réduite, ou de sécurité : moins les robots communiquent, moins leurs communications ont de chance d'être interceptées. Ils doivent donc sélectionner l'information qu'ils collectent ou communiquent en fonction de ce qu'ils savent déjà, mais également en fonction de ce que les autres robots savent. Ainsi, chaque robot doit intégrer une représentation de ses propres connaissances ainsi que de celles des autres robots, et une mesure de la pertinence d'une information par rapport à ces connaissances.

Le développement de méthodes permettant à des robots de décider des actions et communications à effectuer est un problème de décision sous incertitude. Les Processus de Markov Partiellement Observable (POMDPs) sont traditionnellement utilisés pour traiter ce genre de problèmes. Cependant, les POMDPs classiques ne sont pas conçus pour définir la collecte d'information comme but du système et donc ne sont pas adaptés à la perception active dans un système multiagents. Des extensions ont été développées permettant la perception active dans des systèmes monoagent, mais le lien avec les systèmes multiagents n'a

pas encore été fait. Nous proposons dans cet article un nouveau modèle dédié à la collecte d'information, permettant l'exploration active de l'environnement et la communication d'information pertinente.
La section 2 présente un état de l'art et des études pertinentes par rapport au problème considéré. Les sections 3 et 4 présentent notre modèle de perception active avec un système multiagents. Un degré de pertinence orienté agent y est défini ainsi que le Processus Décisionnel de Markov Partiellement Observable utilisé dans notre système. Enfin, la section 5 présente une expérimentation basée sur un problème de perception en intérieur.

2 Travaux connexes

Pertinence

Des agents situés dans un environnement reçoivent une quantité très importante de données qu'ils doivent analyser afin d'en extraire des caractéristiques d'un plus haut niveau d'abstraction. Cependant, l'intérêt d'une caractéristique pour un agent donné dépend d'un certain nombre de paramètres tels que la situation courante, le problème à traiter, le but poursuivi par l'agent... Or, il est contre-productif de communiquer des informations non intéressantes ainsi que d'agir dans le but de collecter ces informations. Ainsi, il est important pour un agent qu'il puisse quantifier l'importance d'une information par rapport aux paramètres précédemment évoqués. Ce degré d'importance d'une information est généralement appelé pertinence de l'information. Borlund [4] a défini deux types de pertinence : la pertinence orientée système (system-oriented relevance) et la pertinence orientée agent (agent-oriented relevance). La pertinence orientée système considère l'information par rapport à une requête : plus l'information correspond et répond à une requête explicitement exprimée, plus son degré de pertinence est élevé. Ce type de pertinence est principalement étudié et utilisé dans les systèmes de recherche d'information [2]. La pertinence orientée agent définit un lien entre une information donnée et le besoin en information d'un agent donné. Si l'information correspond à un besoin spécifique de cet agent, alors son degré de pertinence est élevé. Cependant, il est à noter que ce besoin n'est pas forcément explicité, contrairement à la pertinence orientée système et les requêtes associées. Floridi [6] a proposé une base pour la pertinence épistémique. Roussel et Cholvy [14] ont approfondi ces travaux dans le cas des agents BDI et de la logique multimodale. Cependant, ces études sont basées sur la logique propositionnelle et ne sont pas applicables dans le cadre du raisonnement sur des connaissances incertaines. Ainsi, nous proposons dans cet article une définition de la pertinence pouvant être utilisée dans le cas du raisonnement sur l'incertitude.

Perception active

En utilisant la notion de pertinence, un agent est capable de décider si une information est intéressante ou non. Il peut ainsi percevoir son environnement activement. La perception active désigne le comportement d'un agent agissant dans le but de collecter de l'information et non plus recevant uniquement ces informations passivement. Dans ce contexte, un agent doit prendre des décisions dans un environnement qu'il ne peut pas percevoir complètement. L'un des modèles les plus couramment utilisés dans ce type de problèmes est le Processus Décisionnels de Markov Partiellement Observable (POMDP). Des études ont été menées afin d'introduire la perception active dans les POMDPs. Ponzoni et al. [11] a proposé un critère mixte pour la perception active appliquée à la reconnaissance de véhicules. Ce critère est basé sur une mesure d'entropie et une fonction de récompense classique et permet de fusionner la perception active avec des buts plus classiques des POMDPs. Dans la même période, Araya-Lopez et al. [1] ont proposé une approche plus générale permettant l'utilisation d'une fonction de récompense basée sur les états de croyance des POMDPs. Ces deux études prouvent la faisabilité d'un système ayant la collecte d'information pour but. Cependant, elles sont toutes deux monoagent et ne sont pas applicables directement à un système multiagents. À notre connaissance, il n'existe actuellement pas de modèle pour un système multiagent faisant de la perception active.

Collecte d'information multiagents

La recherche d'information par un système multiagents est un problème relié au domaine de la décision multiagents sous incertitude : un ensemble d'agents doit contrôler simultanément un processus de décision pour collecter de l'information. Cependant, aucun agent n'a le contrôle total du processus. Un certain nombre d'extensions aux POMDPs ont été développés pour traiter ce type de problèmes [16]. Cependant, résoudre un POMDP multiagents est un problème NEXP-complet [3]. Bien que de nombreux algorithmes et heuristiques aient été développés afin de réduire cette complexité [7] [15], de type de modèle n'est généralement pas applicable sur des problèmes réels. Afin de surmonter cette limitation, Spaan et al [18] ont proposé un système basé sur les POMDP permettant à un système de robots de classifier certaines caractéristiques données en agissant pour récolter la meilleure information possible. Dans cette étude, les auteurs ont modélisé la perception active grâce à des actions de classification afin d'éviter d'utiliser une fonction de récompense basée sur une mesure d'entropie, ce qui aurait augmenté la difficulté de la planification. Cependant Araya-Lopez et al. [1] ont prouvé qu'il était possible de réutiliser les techniques standards des POMDPs avec une fonction de récompense basée sur les états de croyance, comme le serait une fonction basée sur l'entropie. De plus, dans le système proposé par Spaan et al., tous les agents doivent effectuer les actions de classification et construire une vue complète de l'environnement. Or, dans les applications classiques de la perception active (surveillance, exploration...), il n'est généralement pas nécessaire que chaque agent ait une vue complète de l'environnement à partir du moment où le système dans son

ensemble possède cette vue complète. Ainsi, il est nécessaire que les agents communiquent entre eux afin d'éviter une exploration répétitive.
Un autre limitation des POMDPs multiagents concerne la communication : celle-ci est généralement considérée comme gratuite et instantanée. Cependant, une telle hypothèse n'est pas possible pour des problèmes réels. Communiquer une information est une action qui a un coût et doit être décidée au même titre que n'importe quelle autre action. Roth et al. [13] ont proposé un algorithme permettant de prendre en compte le coût de la communication dans un POMDP multiagents. Dans cet article, la communication est considérée uniquement lors de la phase d'exécution et doit permettre d'améliorer les performances du système : un agent peut communiquer tout son historique d'observations si cela est utile pour le système. Cependant, aucune décision n'est prise concernant l'observation à communiquer. Dans cet article, la collecte d'information est vue comme une moyen d'atteindre un but donné et n'est pas le but en lui-même.

3 Pertinence orientée agent

Considérons un agent a_i situé dans un environnement $\mathcal{E}$. L'environnement est modélisé comme un ensemble de caractéristiques. Chaque caractéristique est décrite par une variable aléatoire X_k dont les valeurs possibles appartiennent à son domaine $DOM(X_k)$.

$$\mathcal{E} = \{X_k\}$$

L'agent a_i possède des croyances $\mathcal{B}_i^{\mathcal{E}}$ concernant les caractéristiques de l'environnement. Ces croyances sont modélisées par des distributions de probabilité sur les $X_k \in \mathcal{E}$.

$$\mathcal{B}_{i,t}^{\mathcal{E}} = \{b_{i,t}^k \forall X_k \in \mathcal{E}\}$$

où $b_{i,t}^k$ est la distribution de probabilité de l'agent a_i sur la variable X_k à l'instant t. Les agents reçoivent des observations o_m sur l'environnement. Lorsqu'il reçoit une nouvelle observation, l'agent a_i met à jour ses croyances en fonction de cette observation : $\mathcal{B}_{i,t+1}^{\mathcal{E}} = update(\mathcal{B}_{i,t}^{\mathcal{E}}, o_m)$ [5, 17]

Afin de définir le degré de pertinence d'une observation, nous considérons dans un premier temps que cette observation est vraie. En effet, une observation ne peut être pertinente si elle est fausse [6]. La mise en application de cette hypothèse dans le processus de décision sera discutée dans la section 4. Une observation o_m est considérée comme pertinente pour un agent a_i si elle répond aux critères suivants :

1. l'agent a_i est intéressé par le sujet de l'observation o_m
2. l'observation o_m est nouvelle pour l'agent a_i
3. si l'observation o_m n'est pas nouvelle, elle doit améliorer l'état de croyance de l'agent a_i, c'est à dire le rendre plus précis.

L'intérêt que porte un agent a_i pour le sujet une observation o_j (point numéro 1) dépend du modèle du monde que possède cet agent. Si l'arrivée de l'observation modifie les croyances d'un agent, c'est à dire $\mathcal{B}_{i,t+1}^{\mathcal{E}} \neq \mathcal{B}_{i,t}^{\mathcal{E}}$, cela signifie que cette observation peut modifier l'état interne de l'agent et donc correspond à un sujet d'intérêt pour cet agent.
Nous considérons qu'une observation o_m est nouvelle pour un agent a_i si elle modifie beaucoup ses croyances, c'est à dire si $\mathcal{B}_{i,t+1}^{\mathcal{E}}$ et $\mathcal{B}_{i,t}^{\mathcal{E}}$ sont distants l'un de l'autre. Cette distance est mesurée grâce au ratio de Kullback-Leibler.

Définition 1. *Une observation o_m est nouvelle pour un agent a_i à l'instant t si et seulement si*

$$D_{KL}(\mathcal{B}_{i,t}^{\mathcal{E}} || \mathcal{B}_{i,t+1}^{\mathcal{E}}) > \epsilon \quad (1)$$

où ϵ est un seuil arbitrairement fixé et $D_{KL}(\mathcal{B}_{i,t}^{\mathcal{E}} || \mathcal{B}_{i,t+1}^{\mathcal{E}})$ est le ratio de Kullback-Leibler défini par :

$$D_{KL}(\mathcal{B}_{i,t}^{\mathcal{E}} || \mathcal{B}_{i,t+1}^{\mathcal{E}}) = \sum_{X_k \in \mathcal{E}} \sum_{p=1}^{n} b_{i,t}^k(x_p) ln \frac{b_{i,t}^k(x_p)}{b_{i,t+1}^k(x_p)} \quad (2)$$

avec $b_{i,t}^k(x_p)$ la probabilité selon l'agent a_i que la variable X_k vaille x_p.

La précision d'un état de croyance $\mathcal{B}_{i,t}^{\mathcal{E}}$ est modélisée grâce à l'entropie de Shannon.

Définition 2. *L'état de croyance $\mathcal{B}_{i,t+1}^{\mathcal{E}}$ est plus précis que l'état de croyance $\mathcal{B}_{i,t}^{\mathcal{E}}$ si et seulement si*

$$H(\mathcal{B}_{i,t+1}^{\mathcal{E}}) < H(\mathcal{B}_{i,t}^{\mathcal{E}}) \quad (3)$$

où $H(\mathcal{B}_{i,t}^{\mathcal{E}}) = -\sum_{X_k \in \mathcal{E}} \sum_{p=1}^{n} b_{i,t}(x_p) log(b_{i,t}(x_p))$.

Étant donné les définitions précédentes, il est à présent possible de définir un degré de pertinence de la manière suivante :

Définition 3. *Le degré de pertinence d'une observation o_m pour un agent a_i, noté $rel_i(o_m)$ est donné par*

$$rel_i(o_m) = \alpha D_{KL}(\mathcal{B}_{i,t}^{\mathcal{E}} || \mathcal{B}_{i,t+1}^{\mathcal{E}}) - \beta H(\mathcal{B}_{i,t+1}^{\mathcal{E}}) \quad (4)$$

avec $\mathcal{B}_{i,t+1}^{\mathcal{E}} = update(\mathcal{B}_{i,t}^{\mathcal{E}}, o_m)$, α et β étant des poids arbitrairement choisis.

Théoreme 1. *Le degré de pertinence $rel_i(o_m)$ est une fonction convexe.*

Démonstration. Le ratio de Kullback-Leibler est une fonction convexe [10]. L'entropie de Shannon est une fonction concave, donc son opposé est convexe. Le degré de pertinence $rel_i(o_m)$ est la somme de deux fonctions convexes. Il est donc convexe. □

4 Processus de décision pour la perception active

Soit un système multiagents défini par un tuple $< \mathcal{E}, \mathcal{AG}, \mathcal{B}, \mathcal{D} >$ avec
– $\mathcal{E}$, l'environnement comme défini précédemment
– $\mathcal{AG}$, l'ensemble de tous les agents du système
– $\mathcal{D}$, l'ensemble des fonctions de décision de tous les agents

$\mathcal{D} = \{\mathcal{D}_i, \forall i \in \mathcal{AG}\}$ est l'ensemble de toutes les fonctions de décision des agents. Chaque $\mathcal{D}_i$ est représenté par un Processus de Markov Partiellement Observable Factorisé (FPOMDP)[8]. Les ensembles d'états et d'observations considérés ainsi que les fonctions de transition et d'observation sont similaires à celles définies dans le cadre des FPOMDPs. Nous décrirons dans la suite de l'article l'ensemble des actions, la façon dont l'état de croyances est maintenu ainsi que la fonction de récompense, qui eux diffèrent dans notre modèle.

Ensemble des actions

Nous considérons deux types d'actions : regarder la valeur d'une variable aléatoire donnée (action de type $Explore$) et communiquer une observation à un agent (action de type $Communicate$).

$$\begin{aligned}\mathcal{A} = \quad & \{Exp(X_k), \forall X_k \in \mathcal{E}\} \cup \\ & \{Comm(o_m, a_j), \forall o_m \in \mathcal{O}, \forall a_j \in \mathcal{AG}\}\end{aligned}$$

La cardinalité de l'ensemble des actions est :

$$\begin{aligned}|\mathcal{A}| \quad &= |\mathcal{A}_{Explore}| \quad &+|\mathcal{A}_{Communicate}| \\ &= |\mathcal{E}| \quad &+|\mathcal{O}| \times |\mathcal{AG}|\end{aligned} \tag{5}$$

Maintient de l'état de croyances

Un agent ne connaît pas exactement l'état courant du système : il ne reçoit que des observations le renseignant à ce sujet. Ainsi, cet agent doit maintenir un état de croyances sur l'état du système. Dans un contexte de perception active multiagents, un agent donné ne doit pas avoir des croyances uniquement sur l'état de l'environnement mais également sur l'état de croyances des autres agents. En effet, afin de coordonner l'exploration et d'éviter que les agents explorent les mêmes zones, chaque agent doit choisir les observations les plus pertinentes à communiquer. Sachant que la pertinence d'une observation donnée est dépendante de l'état de croyances de l'agent qui la reçoit, chaque agent du système doit également modéliser les croyances courantes des autres agents. On définit ainsi un *état de croyances étendu* de la manière suivante :

Définition 4. *Soit $\mathcal{B}_{i,t}$ l'état de croyance étendu de l'agent a_i à l'instant t.*

$$\mathcal{B}_{i,t} = \mathcal{B}^{\mathcal{E}}_{i,t} \cup \{\mathcal{B}^{j,\mathcal{E}}_{i,t}, \forall j \in \{\mathcal{AG} \backslash i\}\} \tag{6}$$

$\mathcal{B}^{\mathcal{E}}_{i,t} = \{b^{i,k}_{i,t}, \forall X_k \in \mathcal{E}\}$ étant les croyances de l'agent a_i sur l'environnement $\mathcal{E}$ et $\mathcal{B}^{j,\mathcal{E}}_{i,t} = \{b^{j,k}_{i,t}, \forall X_k \in \mathcal{E}\}$ étant les croyances de l'agent a_i sur les croyances de l'agent a_j sur l'environnement.

Il est à noter que $\mathcal{B}^{j,\mathcal{E}}_{i,t}$ est une approximation de $\mathcal{B}^{\mathcal{E}}_{j,t}$.
Afin de maintenir une représentation aussi précise que possible de l'état du système et des croyances des autres agents, un agent a_i doit mettre à jour ses propres croyances régulièrement. Nous considérons trois cas de mise à jour :

1. l'agent a_i reçoit une nouvelle observation après une action de type $Explore$. Il doit alors mettre à jour ses propres croyances concernant l'environnement : $\mathcal{B}^{\mathcal{E}}_{i,t+1}$.
2. l'agent a_i reçoit une nouvelle observation communiquée par l'agent a_j. Il met alors à jour ses propres croyances sur l'environnement $\mathcal{B}^{\mathcal{E}}_{i,t+1}$ mais également ses croyances sur les croyances de l'agent a_j : $\mathcal{B}^{j,\mathcal{E}}_{i,t+1}$.
3. l'agent a_i envoie une nouvelle observation à l'agent a_j par une action de type $Communicate$. Il met alors à jour ses croyances concernant les croyances de l'agent a_j : $\mathcal{B}^{j,\mathcal{E}}_{i,t+1}$.

Dans tous les cas, la mise à jour $\mathcal{B}^{x,\mathcal{E}}_{i,t+1} = update(\mathcal{B}^{x,\mathcal{E}}_{i,t}, o_m)$,$o_m$ étant l'observation, est effectuée comme habituellement dans les POMDPs[5, 17].

Fonction de récompense

Le choix de l'action à effectuer dans un état donné est donné par la politique de l'agent. La politique optimale est calculée en utilisant la fonction de récompense. Celle-ci définit la récompense qu'un agent reçoit en effectuant une certaine action a dans un certain état s. Cependant, dans un contexte de perception active, nous ne cherchons pas à atteindre un état particulier du système mais à collecter et échanger les observations les plus pertinentes. Ainsi, la fonction de récompense n'est pas définie sur les états réels du système mais sur les états de croyances des agents. Un agent est récompensé s'il collecte des observations pertinentes pour lui ou s'il communique des observations pertinentes pour les autres agents. Comme mentionné dans la section 3, une observation doit être vraie pour être pertinente. Sachant que les agents ne disposent que de croyances concernant le monde, il leur est impossible de déterminer si une observation est vraie ou fausse avec certitude. Cependant, les agents ne doivent pas échanger des observations qui renforcent des croyances existantes sans tenir compte de leur véracité. Pour cela, nous utilisons la croyance de l'agent concernant l'environnement et la fonction d'observation classique des POMDPs. Nous définissons donc la fonction de récompense du système de la manière suivante :

$$\begin{aligned}&R(\mathcal{B}_{i,t}, Exp(X_k)) = \\ &\textstyle\sum_{s \in \mathcal{S}} \sum_{o_m \in \mathcal{O}} \mathcal{B}_{i,t}(s) \omega(o_m, s, Exp(X_k)) rel_i(o_m) \\ &- C_{Exp(X_k)}\end{aligned}$$

$$\begin{aligned}&R(\mathcal{B}_{i,t}, Comm(o_m, a_j)) = \\ &\textstyle\sum_{s \in \mathcal{S}} \mathcal{B}_{i,t}(s) \omega(o_m, s, Comm(o_m, a_j)) rel_j(o_m) \\ &- C_{Comm(o_m, a_j)}\end{aligned}$$

$C_{Exp(X_k)}$ et $C_{Comm(o_m,a_j)}$ étant les coûts respectifs des actions $Explore$ et $Communicate$ et $\mathcal{B}_{i,t}(s)$ étant la croyance à un instant t de l'agent a_i que l'état s est l'état courant.

Résolution

Dans le POMDP décrit précédemment, les actions sont épistémiques : elles ne modifient pas l'état réel du système. Ainsi, il est possible de transformer ce POMDP en un MDP continu défini par un tuple $< \Delta, \mathcal{A}, \tau >$ où :

– Δ est le nouvel espace d'états. Il correspond à l'espace des états de croyances du POMDP précédent. $\Delta = \mathcal{B}_i$
– $\mathcal{A}$ est le même espace d'actions que précédemment
– τ est la nouvelle fonction de transition

La fonction de transition τ de ce MDP est alors définie comme suit :

$$\tau(\mathcal{B}_{i,t}, a, \mathcal{B}_{i,t+1}) =$$

$$\begin{cases} \sum_{s\in\mathcal{S}} \sum_{o_m \in U_t} \omega(o_m, s, a)\mathcal{B}_{i,t}(s) & si \quad U_t \neq \emptyset \\ 0 & sinon \end{cases}$$

$U_t = \{o_m \in \mathcal{O} \mid \mathcal{B}_{i,t+1} = update(\mathcal{B}_{i,t}, o_m)\}$ étant l'ensemble de toutes les observations permettant de passer de l'état $\mathcal{B}_{i,t}$ à l'état $\mathcal{B}_{i,t+1}$, $\omega(o_m, s_t, a)$ la fonction d'observation du POMDP et $\mathcal{B}_{i,t}(s_t)$ étant la croyance de l'agent a_i que l'état courant est s_t.

La fonction de valeur correspondant à ce MDP continu est :

$$V(\mathcal{B}_{i,t}) = \mathcal{R}(\mathcal{B}_{i,t}) + \max_{a\in\mathcal{A}} \int_{\mathcal{B}'_{i,t}} \tau(\mathcal{B}_{i,t}, a, \mathcal{B}'_{i,t}) V(\mathcal{B}'_{i,t}) \tag{7}$$

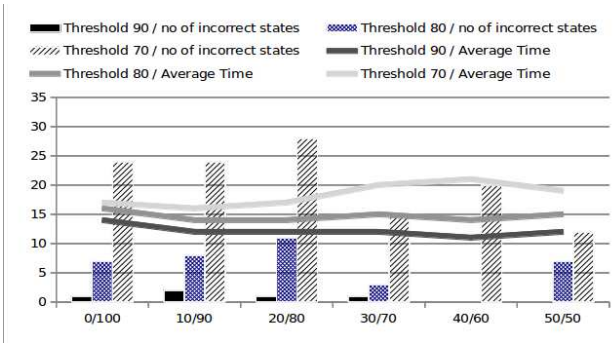
En discrétisant les distributions de probabilité dans l'espace d'états, on peut ainsi discrétiser l'espace d'états lui-même et transformer l'équation 7 en :

$$V(\mathcal{B}_{i,t}) = \mathcal{R}(\mathcal{B}_{i,t}) + \max_{a\in\mathcal{A}} \sum_{\mathcal{B}'_{i,t}\in Samples} \tau(\mathcal{B}_{i,t}, a, \mathcal{B}'_{i,t}) V(\mathcal{B}'_{i,t}) \tag{8}$$

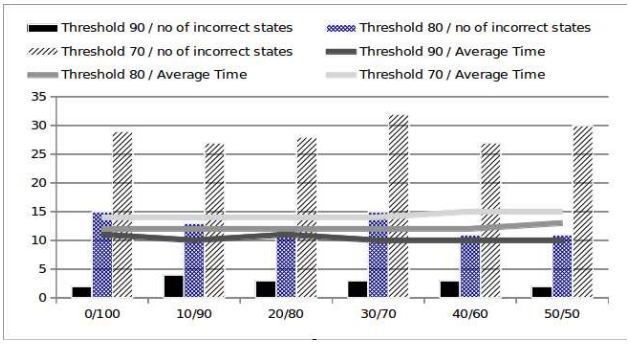
Il est alors possible d'utiliser n'importe quelle technique de la littérature pour résoudre ce MDP [12] [9].

5 Experiments

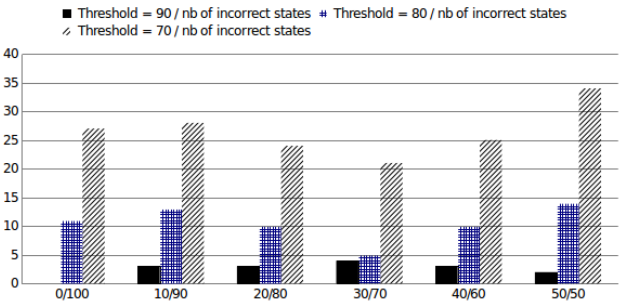
Le modèle présenté dans cet article a été implémenté sur un scénario simple de surveillance en intérieur. Deux robots doivent explorer un environnement composé de 4 zones connectées les unes aux autres. On considère deux observations pour chaque zone : *salleVide* et *salleNonVide*. La politique optimale a été calculée en utilisant différents poids de compromis entre le ratio de Kullback-Leibler et l'entropie, ainsi que différentes probabilités d'obtenir une fausse observation. La simulation a été lancée 50 fois pour chaque ensemble de paramètres. Afin d'évaluer la politique calculée, nous avons mesuré le nombre moyen de messages envoyés par les robots, le temps moyen nécessaire pour obtenir un état de croyances stable et le nombre d'états de croyances faux à la fin de l'exploration. Nous avons comparé ces mesures à celles obtenues avec un système multiagents sans communication et avec un système dans lequel chaque agent communique chaque observation reçue. Les résultats sont présentés sur la figure 1.



(a) Évaluation de la politique calculée avec communication d'observations pertinentes



(b) Évaluation de la politique calculée sans communication

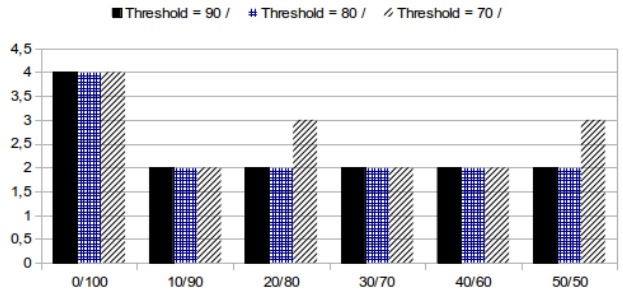


(c) Évaluation de la politique calculée avec communication totale

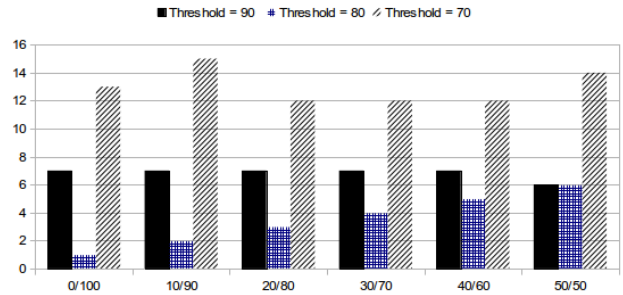
FIGURE 1 – Évaluation des trois politiques. L'axe des abscisses représente les différents ratios Kullback-Leible/Entropie utilisés. Les seuils présentés correspondent aux différentes probabilités d'obtenir une observation correcte après une action de type $Explore$

Les courbes présentes sur les graphiques 1a (Système avec communication pertinente) et 1b (Système sans communication) représentent le temps moyen nécessaire pour atteindre un état de croyance stable mesuré en nombre d'itérations. Une itération est constituée de trois étapes : l'exécution d'une action, la réception de l'observation associée et la réception d'une communication s'il y en a. Cette mesure n'a pas été effectuée pour le système avec communi-

cation totale car il ne nous a pas paru pertinent d'effectuer la comparaison avec ce système. En effet, nous cherchons à mesurer le rapport entre la perte de temps induite par l'introduction de la communication dans le système et le gain en efficacité du système. Le système avec communication totale quant à lui nous sert de référence quant à l'efficacité de notre système de communication. Les histogrammes représentent quant à eux le nombre moyen d'états finaux partiellement ou totalement incorrects. Nous notons tout d'abord que le système avec communication pertinente (Figure 1a) est en moyenne 18% plus long que le système sans communication (Figure 1b). Cependant, le nombre d'états faux est réduit dans les cas de communication pertinente comme totale. Dans le pire cas, c'est à dire avec une probabilité de 70% de recevoir une observation correcte, le système avec communication pertinente réduit de 28% le nombre d'états faux, tandis que le système avec communication totale ne le réduit que de 8%. Dans le cas moyen, c'est à dire une probabilité de 80% d'obtenir une observation correcte, le nombre d'états faux est réduit de 72% avec notre système à communication pertinente et de 19% avec le système à communication totale. De plus, la figure 2a montre que le nombre moyen de messages envoyés reste à peu près constant et - de façon évidente - bien moindre que dans le cas d'un système avec communication totale (Figure 2b).



(a) Système avec communication d'informations pertinentes



(b) Système avec communication totale

FIGURE 2 – Nombre moyen de messages envoyés

Ces résultats montrent que notre système avec communication pertinente est plus efficace qu'un système sans communication mais également plus efficace qu'un système avec une communication totale. Ceci peut être expliqué par le fait qu'un agent peut recevoir une observation fausse. Dans le cas d'un système avec communication totale, cette observation sera communiquée et propagera l'erreur dans tout le système. Dans le cas de notre système avec communication pertinente, cette observation ne perturbera que les croyances de l'agent qui la reçoit. En effet, si ses croyances ne sont pas suffisamment précises et ne sont pas en faveur de la véracité de l'observation , l'agent ne la communiquera pas. Ces expériences semblent donc valider l'hypothèse initiale : choisir des observations pertinentes à communiquer permet d'améliorer les performances d'un système d'exploration tout en réduisant le nombre de communications.

6 Conclusion et Perspectives

Nous avons présenté un nouveau modèle de pertinence orientée agent ainsi qu'un processus de décision permettant la perception active avec un système multiagents. Chaque agent calcule la pertinence d'une observation par rapport à ses croyances ou à celles d'un autre agent du système. Il peut par la suite décider s'il doit explorer une zone donnée pour récupérer cette observation ou s'il doit la communiquer. La pertinence d'une observation est un compromis entre sa nouveauté - modélisée grâce au ratio de Kullback-Leibler - et l'effet de confirmation qu'elle a sur les croyances de l'agent qui la reçoit - modélisé grâce à une mesure d'entropie. Ce modèle a été implémenté et testé en simulation. Les résultats montrent que cette approche est plus efficace qu'un modèle sans communication mais aussi qu'un modèle avec communication totale. Bien qu'encourageant, ces résultats ont été obtenus sur un exemple relativement restreint. Des heuristiques sont actuellement à l'étude afin de faire passer ce modèle à l'échelle et de permettre de résoudre des problèmes de taille réelle.
Dans le modèle présenté, un agent est capable de communiquer n'importe quelle observation de l'ensemble des observations si celle-ci est pertinente et si son état de croyance est suffisant pour la considérer comme vraie. Cependant, un agent peut ainsi communiquer une observation qu'il n'a jamais reçue directement. Des travaux futurs viseront à maintenir un historique des observations reçues et ne permettront à un agent que de communiquer des observations de cet historique. De plus, les croyances d'un agent sur les croyances d'un autre agent ne sont mises à jour qu'en cas de communication explicite. Nous prévoyons de travailler sur une méthode moins naïve. En effet, les agents utilisent les mêmes politiques optimales. Il est donc possible de mettre à jour les états de croyances en mettant des hypothèses sur les actions effectuées par les autres agents. Enfin, nous prévoyons également d'intégrer ce modèle à des POMDPs non épistémiques pour élargir les possibilités d'applications.

Références

[1] M. Araya, O. Buffet, V. Thomas, and F. Charpillet. A pomdp extension with belief-dependent rewards. In *Advances in Neural Information Processing Systems*, pages 64–72, 2010.

[2] R. Baeza-Yates, B. Ribeiro-Neto, et al. *Modern information retrieval*, volume 463. ACM press New York, 1999.

[3] D. S. Bernstein, R. Givan, N. Immerman, and S. Zilberstein. The complexity of decentralized control of markov decision processes. *Mathematics of operations research*, 27(4) :819–840, 2002.

[4] P. Borlund. The concept of relevance in ir. *Journal of the American Society for information Science and Technology*, 54(10) :913–925, 2003.

[5] A. R. Cassandra, L. P. Kaelbling, and M. L. Littman. Acting optimally in partially observable stochastic domains. In *AAAI*, volume 94, pages 1023–1028, 1994.

[6] L. Floridi. Understanding epistemic relevance. *Erkenntnis*, 69(1) :69–92, 2008.

[7] C. V. Goldman and S. Zilberstein. Decentralized control of cooperative systems : Categorization and complexity analysis. *J. Artif. Intell. Res.(JAIR)*, 22 :143–174, 2004.

[8] E. A. Hansen and Z. Feng. Dynamic programming for pomdps using a factored state representation. 2000.

[9] J. Hoey and P. Poupart. Solving pomdps with continuous or large discrete observation spaces. In *International Joint Conference on Artificial Intelligence*, volume 19, page 1332, 2005.

[10] M. Lea. Useful facts about the kullback-leibler discrimination distance. *Houston, Texas*, 2004.

[11] C. Ponzoni Carvalho Chanel, F. Teichteil-Königsbuch, and C. Lesire. Pomdp-based online target detection and recognition for autonomous uavs. In *ECAI*, pages 955–960, 2012.

[12] J. M. Porta, N. Vlassis, M. T. Spaan, and P. Poupart. Point-based value iteration for continuous pomdps. *The Journal of Machine Learning Research*, 7 :2329–2367, 2006.

[13] M. Roth, R. Simmons, and M. Veloso. Reasoning about joint beliefs for execution-time communication decisions. In *Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*, page 786–793, 2005.

[14] S. Roussel and L. Cholvy. Cooperative interpersonal communication and relevant information. In *ESSLLI Workshop on Logical Methods for Social Concepts, Bordeaux*. Citeseer, 2009.

[15] S. Seuken and S. Zilberstein. Memory-bounded dynamic programming for dec-pomdps. In *IJCAI*, pages 2009–2015, 2007.

[16] S. Seuken and S. Zilberstein. Formal models and algorithms for decentralized decision making under uncertainty. *Autonomous Agents and Multi-Agent Systems*, 17(2) :190–250, 2008.

[17] R. D. Smallwood and E. J. Sondik. The optimal control of partially observable markov processes over a finite horizon. *Operations Research*, 21(5) :1071–1088, 1973.

[18] M. T. Spaan, T. S. Veiga, and P. U. Lima. Active cooperative perception in network robot systems using POMDPs. In *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, page 4800–4805, 2010.