A Distributed Decision-Theoretic Model for Multiagent Active Information Gathering

Jennifer Renoux, Abdel-Illah Mouaddib, Simon Le Gloannec

A Distributed Decision-Theoretic Model for Multiagent Active Information Gathering

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

[1] GREYC, Universtiy of Caen
[2] Airbus Defence and Space

Abstract. Multirobot systems have made tremendous progress in exploration and surveillance. In that kind of problem, agents are not required to perform a given task but should gather as much information as possible. However, information gathering tasks usually remain passive. In this paper, we present a multirobot model for active information gathering. In this model, robots explore, assess the relevance, update their beliefs and communicate the appropriate information to relevant robots. To do so, we propose a distributed decision process where a robot maintains a belief matrix representing its beliefs and beliefs about the beliefs of the other robots. This decision process uses entropy and Kullback-Leibler in a reward function to access the relevance of their beliefs and the divergence with each other. This model allows the derivation of a policy for gathering information to make the entropy low and a communication policy to reduce the divergence. An experimental scenario has been developed for an indoor information gathering mission.

1 Introduction

Robotic systems are increasingly used in surveillance and exploration applications. In the future robots may assist humans and eventually replace them in dangerous areas. In these particular research fields the main goal of the system is not to reach a given goal but to gather information. The system needs to create a complete and accurate view of the situation. This built view may be used afterwards by some agents - human or artificial - to make some decisions and perform some actions. Therefore the information gathering system must be able to identify lacking information and take the necessary steps to collect it. However it is obviously not productive that all the robots in the system try to collect all possible information, just as it is not possible for the robots to communicate all the information they have all the time. They should select pieces of information to collect or to communicate depending on what they already know and what other agents already know. Developing methods to allow robots to decide how to act and what to communicate is a decision problem under uncertainty.

Partially Observable Markov Decision Processes (POMDPs) are traditionally used to deal with this kind of problem, more particularly Decentralized POMDPs (DEC-POMDPs) that are an extension of POMDPs for multiagent systems. However the classic POMDP framework is not designed to have information gathering as a target : information gathering is usually a means of reaching another goal. Some extensions have been developed for mono-agent systems but the plunge to multi-agent systems has not

been taken. We suggest in this paper a formal definition of the relevance of a piece of information as well as a new model dedicated to multiagent information gathering that is able to explore actively its environment and communicate relevant information.

Section 2 presents some background knowledge and other studies that are relevant to our problem. Section 3 presents the proposed model to do active sensing with a multi-robot system. It defines a agent-oriented relevance degree and describes the Partially Observable Markov Decision Process used in the system. Finally, section 4 presents an implementation of the model on a simple indoor sensing problem.

2 Background and Related Work

Relevance

Agents situated in an environment perceive a huge amount of data and they need to process those data to extract higher-level features. However the interest of a feature for an agent depends on several parameters such as the situation, the problem to be dealt etc. Since it is counterproductive to communicate neither to perform an action to collect non interesting information, agents need to quantify the importance of a piece of information according to the current situation. This degree of importance is the relevance of information. Borlund [1] defined two types of relevance : system-oriented relevance and agent-oriented relevance. System-oriented relevance analyzes relevance regarding a request. The better the match between information and the request, the greater the degree of relevance. System-oriented relevance is used and studied in several Information Retrieval Systems [3]. Agent-oriented relevance defines a link between information and an agent's needs. Information is relevant if it matches a specific need, if it is informative and useful for an agent which receives it. However the need may not be explicit. Floridi [4] suggested a base of epistemic relevance and Roussel and Cholvy [5] deepened this study in the case of BDI agents and multimodal logic. However those studies are based on propositional logic and are not applicable for reasoning with uncertain knowledge. Therefore we need to define a relevance theory that may be used in uncertain reasoning.

Active information gathering

Using relevance, an agent is able to decide if a piece of information is interesting or not. Therefore it is able to perform active information gathering. Active information gathering defines the fact that an agent will act voluntarily to gather information and not just perceive passively its environment. In this context the agent has to make decisions in an environment that it cannot perceive completely. One of the best and commonly used models to deal with that kind of problem is Partially Observable Markov Decision Process. Some studies have already been carried out to perform active perception using POMDPs. Ponzoni et al. [6] suggested a mixed criterion for active perception as a mission to recognize some vehicles. The criterion is based on an entropy function and a classical reward function to mix active perception and more classical goals of POMDPs. Meanwhile Araya-Lopez et al. [7] suggested a more general approach to use any reward function based on belief state in POMDPs. These two approaches proved the feasibility of such a system where information gathering is the goal. However they are both

mono-agent and are not applicable to a multiagent system. To our knowledge there is no model of multiagent system for active information gathering.It is obvious that information gathering would be more efficient if it is done by several agents instead of a single one. However, it is important that agents are able to coordinate themselves to make the gathering efficient.

Multiagent active information gathering

The problem of multiagent active information gathering relates to multiagent decision under uncertainty : a set of agents have to control together a decision process to collect information. However no agent is able to control the whole process. Different equivalent frameworks extending POMDPs have been developed to deal with multiagent decision problems under uncertainty [14].

Solving a multiagent POMDP is a problem NEXP-complete [15]. Even if algorithms and heuristics have be suggested to overcome this complexity [17], those frameworks are usually not applicable to real problems. To overcome this issue, Spaan et al. [18] suggested a system based on POMDP to enable a Network Robot System to classify particular features by acting to get the best information possible. In this study, the authors decided to model the active information gathering thanks to classifying actions in order to avoid using a reward function based on entropy, that would increase the complexity of planning. However Araya-Lopez et al. [7] proved that it is possible to reuse techniques from the standard POMDP literature with a reward function based on belief states as would be a reward function using entropy. On top of that, in the system suggested by Spaans et al., all the agents have to make the classification steps and build a complete view of the environment. However, in usual active information gathering problems, it not useful that each agent of the system has this complete view as long as the system view is complete. Therefore, agents need to communicate with each others in order to avoid repetitive exploration. Communication in multiagent POMDP framework is usually assumed to be free and instantaneous. However such assumption is not possible in real problems. Communication is an action that has a cost and must be decided. Roth et al. [21] presented an algorithm to take into account the communication cost in multiagent POMDPs. In this paper, the communication is considered only during execution and should improve the performance of the system : if it is useful for the system, an agent communicates all its history of observations. There is no decision concerning the observations to communicate. Information gathering is once again a means to reach a goal and not the goal in itself.

3 The model

3.1 Definition of an agent-based relevance

Let's assume an agent a_i situated in an environment $\mathcal{E}$. The environment is modeled as a set of features of interest. Each feature is described using a random variable X_k which can take values in its domain $DOM(X_k)$. The agent a_i has some beliefs $\mathcal{B}_i^{\mathcal{E}}$ concerning the features of interest modeled as probability distributions over the $X_k \in \mathcal{E}$.

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

with $b^k_{i,t}$ being the probability distribution of agent a_i over the variable X_k at time t. Let's assume an agent receives observations concerning the features of interest. Possible observations are the possible values of the random variables $X_k : o_k \in DOM(X_k)$

When receiving a new observation, agent a_i updates its beliefs according to it : $\mathcal{B}^{\mathcal{E}}_{i,t+1} = update(\mathcal{B}^{\mathcal{E}}_{i,t}, o_k)$ [22]

First of all we considered that observations received are true. As a matter of fact, an observation cannot be relevant if it is a false observation [4]. We discuss this assumption and the way it is used in the decision process in section 3.2. Considering this assumption, an observation o_k is relevant for an agent a_i if it matches the following criteria:

1. agent a_i is interested in the subject of the observation o_k, that is to say X_k
2. the observation o_k is new for agent a_i or
3. if the observation o_k is not new, it should render agent's a_i beliefs more accurate

The first point is dealt with the way we represent agent's beliefs : if agent a_i is interested in X_k then X_k is in agent's a_i beliefs. We assume that an observation o_k is new for agent a_i if beliefs $\mathcal{B}^{\mathcal{E}}_{i,t+1}$ and $\mathcal{B}^{\mathcal{E}}_{i,t}$ are distant from each other. The distance between two probability distributions is measured by the Kullback-Leibler ratio.

Definition 1. *An observation o_k is new for agent a_i at time t if and only if*

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

ϵ is a fixed threshold and $D_{KL}(\mathcal{B}^{\mathcal{E}}_{i,t}||\mathcal{B}^{\mathcal{E}}_{i,t+1})$ is the Kullback-Leibler ratio and defined by

$$D_{KL}(\mathcal{B}^{\mathcal{E}}_{i,t}||\mathcal{B}^{\mathcal{E}}_{i,t+1}) = \sum_{X_k \in \mathcal{E}} \sum_{k=1}^{n} b^k_{i,t}(x_k) ln \frac{b^k_{i,t}(x_k)}{b^k_{i,t+1}(x_k)} \tag{2}$$

where $b^k_{i,t}(x_k)$ is the belief of agent a_i that the random variable X_k is equals to x_k.

To model the accuracy of a belief $\mathcal{B}^{\mathcal{E}}_{i,t}$, we use an entropy measure.

Definition 2. *Belief $\mathcal{B}^{\mathcal{E}}_{i,t+1}$ is more precise than belief $\mathcal{B}^{\mathcal{E}}_{i,t}$ if and only if*

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

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

Given the previous definitions we may define the degree of relevance as shown below :

Definition 3. *The degree of relevance of an observation o_k concerning a random variable X_k for an agent a_i, noted $rel_i(o_k)$, is given by*

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

with $\mathcal{B}^{\mathcal{E}}_{i,t+1} = update(\mathcal{B}^{\mathcal{E}}_{i,t}, o_k)$, α and β being weights and δ being a translation factor to ensure the relevance is positive.

3.2 Decision Process for multiagent active information gathering

Let a multiagent system be defined as a tuple $< \mathcal{E}, \mathcal{AG}, \mathcal{B}, \mathcal{D} >$ with

- $\mathcal{E}$ being the environment as defined previously
- $\mathcal{AG}$ being the set of agents
- $\mathcal{D}$ being the set of all agents' decision functions

$\mathcal{D} = \{\mathcal{D}_i, \forall i \in \mathcal{AG}\}$ is the set of all agents' decision function. Each $\mathcal{D}_i$ is represented as a Factored Partially Observable Markov Decision Process (FPOMDP)[24].

Set of actions

We consider two type of actions : look for the value of a particular random variable ($Explore$-type actions) and communicate an observation to an agent ($Communicate$-type actions):

$$\begin{aligned}\mathcal{A} = &\{Exp(X_k), \forall X_k \in \mathcal{E}\} \cup \\ &\{Comm(o, Ag), \forall o \in \mathcal{O}, \forall a \in \mathcal{AG}\}\end{aligned}$$

The size of the action set is :

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

Set of observations

In a Partially Observable Markov Decision Process, an agent doesn't know exactly the current state of the system. It only receives observations when performing actions, which are only indications about the current state. So the agent may estimate the current state from the history of observations it received. When performing an $Explore$-type action, the agent receives an observation concerning the random variable it is trying to sense. Therefore the possible observations are the possible values the random variable may take, that is to say the domain of the random variable. Therefore the set of all possible observations from an $Explore$-type action is:

$$\mathcal{O}_{Explore} = DOM(\mathcal{E}) = \bigcup_{\forall X_k \in \mathcal{E}} DOM(X_k)$$

When performing a $Communicate$-type action the agent receives an observation stating that the message has been properly sent or not :

$$\mathcal{O}_{Communicate} = \{okMsg, nokMsg\}$$

So the entire set of possible observations is

$$\mathcal{O} = \mathcal{O}_{Explore} \cup \mathcal{O}_{Communicate} = \bigcup_{\forall X_k \in \mathcal{E}} DOM(X_k) \cup \{okMsg, nokMsg\} \tag{6}$$

Maintaining a belief state The agent doesn't know the exact current state of the system. It only has observations about it. Therefore it should maintain some beliefs concerning this current state. In the context of multiagent information gathering, an agent should not only have beliefs about the state of the environment but also about the other agents. To prevent agents from exploring the same areas and to enable them to choose the most relevant observation to communicate, they should model the knowledge of other agents in their own belief state. Thus we defined an *extended belief state* as following :

Definition 4. *Let a extended belief state of an agent a_i at time t be defined as following :*

$$\mathcal{B}_{i,t} = \mathcal{B}^{\mathcal{E}}_{i,t} \cup \{\mathcal{B}^{j,\mathcal{E}}_{i,t}, \forall j \in \{\mathcal{AG} \backslash i\}\} \tag{7}$$

with $\mathcal{B}^{\mathcal{E}}_{i,t} = \{b^{i,k}_{i,t}, \forall X_k \in \mathcal{E}\}$ being the beliefs of agent a_i concerning the environment $\mathcal{E}$ and $\mathcal{B}^{j,\mathcal{E}}_{i,t} = \{b^{j,k}_{i,t}, \forall X_k \in \mathcal{E}\}$ being the beliefs of agent a_i concerning the beliefs of agent a_j concerning the environment.

Let's note that $\mathcal{B}^{j,\mathcal{E}}_{i,t}$ is an approximation of $\mathcal{B}^{\mathcal{E}}_{j,t}$.

To keep an accurate representation of the current state of the system an agent has to update its beliefs regularly. An update will occur in three cases :

1. the agent receives a new observation from its sensors after an $Explore$ action. It updates its own beliefs concerning the environment : $\mathcal{B}^{\mathcal{E}}_{i,t+1}$.
2. the agent receives a new observation from agent a_j. It updates its own beliefs $\mathcal{B}^{\mathcal{E}}_{i,t+1}$ as well as its beliefs concerning agent a_j : $\mathcal{B}^{j,\mathcal{E}}_{i,t+1}$.
3. the agent sends an observation to agent a_j. It updates its beliefs concerning agent a_j : $\mathcal{B}^{j,\mathcal{E}}_{i,t+1}$.

In all cases the update $\mathcal{B}^{x,\mathcal{E}}_{i,t+1} = update(\mathcal{B}^{x,\mathcal{E}}_{i,t}, o_k)$,$o_k$ being the observation received, is made as usual in Partially Observable Markov Decision Processes :

$$\mathcal{B}_{i,t+1}(s') = \frac{\omega(o_k, s', a) \sum_{s \in \mathcal{S}} p(s'|s, a)\mathcal{B}_{i,t}(s)}{\sum_{s \in \mathcal{S}} \sum_{s'' \in \mathcal{S}} \omega(o_k, s'', a)p(s''|s, a)b^{i,k}_{i,t}} \tag{8}$$

Reward function The best action to perform at a given time is set by a policy. The optimal policy is computed thanks to the reward function. The reward function defines the reward an agent may receive by performing action a in state s. However in an information gathering context we are not interested in reaching some special state of the system but gathering and communicating relevant observations. Therefore the reward function is defined on the belief states of the agent and not on the real states of the system. An agent is rewarded if it collects observations that are relevant for itself and if it communicates observations that are relevant for other agents. As mentioned in section 3.1, an observation must be true to be relevant. Since agents only have beliefs concerning the world, they cannot ensure that an observation is true. However they should not exchange observations that reinforce their existing beliefs, regardless of their veracity. Therefore we need to find a compromise between the agent's belief concerning

the observation and the relevance of this observation. To do so we weight the relevance of a given observation by the agent's belief concerning its truth. This belief is given by the probability of receiving the observation in the state s considered multiplied by the agent's belief that the state s is the current state. The reward function is thus defined as follows:

$$R(\mathcal{B}_{i,t}, Exp(X_k)) = \sum_{s \in \mathcal{S}} \sum_{o_k \in \mathcal{O}} \mathcal{B}_{i,t}(s)\omega(o_k, s, a)rel_i(o_k) - C_{Exp(X_k)}$$

$$R(\mathcal{B}_{i,t}, Comm(o_k, a_j)) = \sum_{s \in \mathcal{S}} \mathcal{B}_{i,t}(s)\omega(o_k, s, a)rel_j(o_k) - C_{Comm(o_k,a_j)}$$

with $C_{Explore(X_k)}$ and $C_{Communicate(o_k,a_j)}$ being the costs of taking the $Explore$ or $Communicate$ action and $\mathcal{B}_{i,t}(s)$ being the belief of agent a_i that state s is the current state.

Resolution In this POMDP actions are epistemic : they don't modify the real state of the system. Therefore it is possible to transform our POMDP into a Belief MDP defined as a tuple $< \Delta, \mathcal{A}, \tau >$ where :

- Δ is the new state space. It corresponds directly to the belief state space in the initial POMDP. $\Delta = \mathcal{B}_i$
- $\mathcal{A}$ is the same state of actions as previously defined
- τ is the new transition function

Theorem 1. *The transition function τ of the Belief MDP is defined as follows :*

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

$$\begin{cases} \sum_{s \in \mathcal{S}} \sum_{o_k \in U_t} \omega(o_k, s, a)\mathcal{B}_{i,t}(s) \; if \; U_t \neq \emptyset \\ 0 \qquad\qquad\qquad\qquad\qquad\qquad otherwise \end{cases}$$

where $U_t = \{o_k \in \mathcal{O} \mid \mathcal{B}_{i,t+1} = update(\mathcal{B}_{i,t}, o_k)\}$ is the set of all observations enabling the transition from state $\mathcal{B}_{i,t}$ to state $\mathcal{B}_{i,t+1}$, $\omega(o_k, s_t, a)$ is the observation function of the POMDP and $\mathcal{B}_{i,t}(s_t)$ is the belief of agent a_i that the current state is s_t.

Proof. If there is no observation such as

$\mathcal{B}_{i,t+1} = update(\mathcal{B}_{i,t}, o_k)\}$, it is not possible to transfer from belief state $\mathcal{B}_{i,t}$ to belief state $\mathcal{B}_{i,t+1}$. Therefore, $\tau(\mathcal{B}_{i,t}, a, \mathcal{B}_{i,t+1}) = 0$. If there exists at least one observation o_k such as $\mathcal{B}_{i,t+1} = update(\mathcal{B}_{i,t}, o_k)\}$ we have the following equations :

$$\begin{aligned} \tau(\mathcal{B}_{i,t}, a, \mathcal{B}_{i,t+1}) &= P(\mathcal{B}_{i,t+1}|\mathcal{B}_{i,t}, a) \\ &= \sum_{o_k \in U_t} P(o_k|\mathcal{B}_{i,t}, a) \\ &= \sum_{s \in \mathcal{S}} \sum_{o_k \in U_t} P(o_k|s, a)\mathcal{B}_{i,t}(s) \\ &= \sum_{s \in \mathcal{S}} \sum_{o_k \in U_t} \omega(o_k, s, a)\mathcal{B}_{i,t}(s) \end{aligned}$$

The value function corresponding to this Belief MDP is defined as following:

$$V(\mathcal{B}_{i,t}) = \mathcal{R}(\mathcal{B}_{i,t}) + \max_{a \in \mathcal{A}} \int_{\mathcal{B}'_{i,t}} \tau(\mathcal{B}_{i,t}, a, \mathcal{B}'_{i,t}) V(\mathcal{B}'_{i,t}) \tag{9}$$

Using discretization techniques on the probability distributions, we may transform equation 9 :

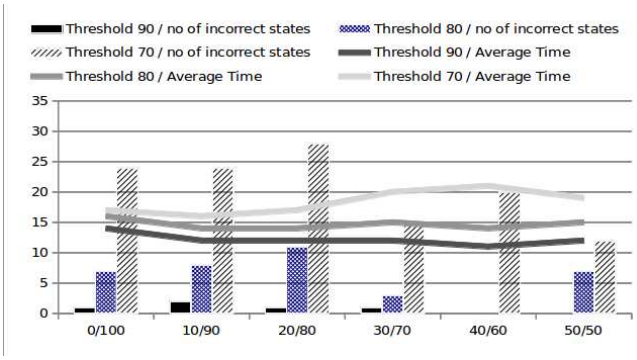
$$V(\mathcal{B}_{i,t}) = \mathcal{R}(\mathcal{B}_{i,t}) + \max_{a \in \mathcal{A}} \sum_{\mathcal{B}'_{i,t} \in Samples} \tau(\mathcal{B}_{i,t}, a, \mathcal{B}'_{i,t}) V(\mathcal{B}'_{i,t}) \tag{10}$$

Then, any technique from the literature may be used to solve this belief-MDP [28].

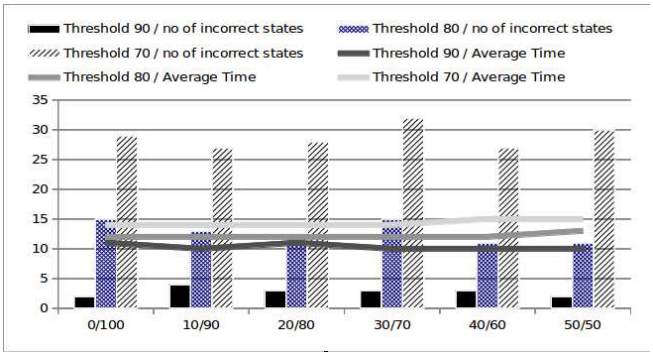
4 Experiments

Simulated Robots The suggested model was implemented on a simple scenario. Two robots have to explore an environment made of 4 different zones connected to each other. For each zone, two observations are possible : *emptyRoom* and *notEmptyRoom*. The optimal policy was computed using different ratios to make the compromise between the Kullback-Leibler ratio and the entropy measure, as well as different probabilities of obtaining a false observation. The system was run 50 times with each set of parameters. To evaluate the policy, we measured the average number of messages sent by the robots, the average time needed to get a stable belief state and the number of false belief states at the end of exploration. We compared those measures with a multirobot system without communication and with a system in which agents communicate each observation they received immediately. The results are presented in figure 1.

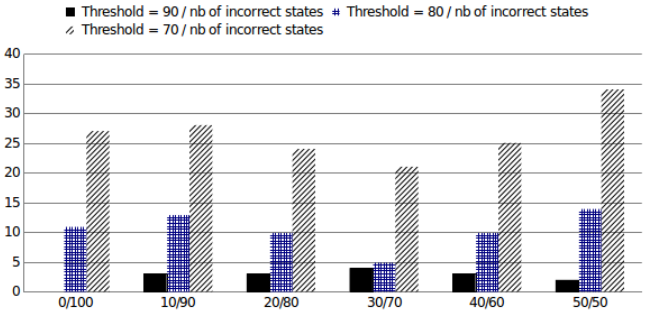
The linear part of the first two graphs represents the average time to reach a stable belief state, measured in number of iterations, an iteration being made up of the execution of an action, the reception of the associated observation and the reception of a communication message from another agent, if any. The bar part represents the average number of final states that were partially or totally incorrect. First of all, we notice that the communicating system (Figure 1a) takes around 18% longer than the explore-only system (Figure 1b) to reach a stable belief state. We did not measured the time needed for the fully communicative (Figure 1c) system because it is not meaningful. Indeed agents are communicating each observation they receive and we consider communication as a separate action, so this system is much longer that the others. However, we notice that the number of false end states is reduced with partial and complete communication. In the worst case, that is to say a probability of 70% of receiving a correct observation after doing an explore action, the system with partial communication reduces the average number of false end states by 28%, and the system with complete communication reduces it by 8%. However it is possible to configure the Kullback-Leibler / Entropy ratio so that this amount increases to 50% for the system with partial communication (by considering Kullback-Leibler and Entropy with the same weights). In the average case, that is to say a probability of 80% of receiving a correct observation, the average number of false end states is reduced by 72% with our system and by 19% with a complete communication. On top of that, Figure 2a shows that the average



(a) Evaluation of the computed policy with relevant communication

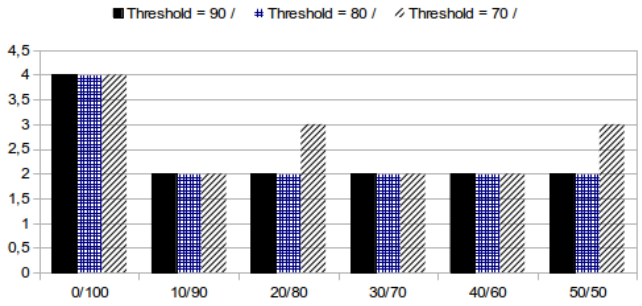


(b) Evaluation of the policy without communication

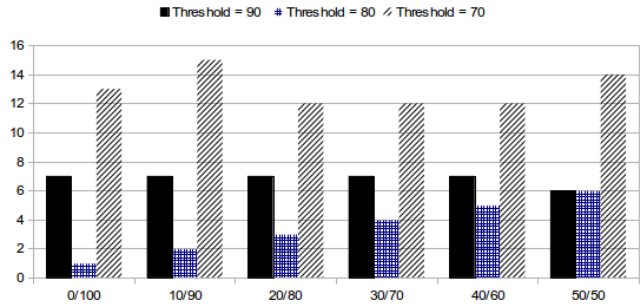


(c) Evaluation of the fully-communicative policy

Fig. 1: Evalutation of the three policies. The X-Axis represents the different ratios Kullback-Leibler / Entropy. The three different thresholds correspond to the probability to obtain a correct observation while doing an explore action



(a) System with partial communication



(b) System with complete communication

Fig. 2: Exploration with μ-troopers

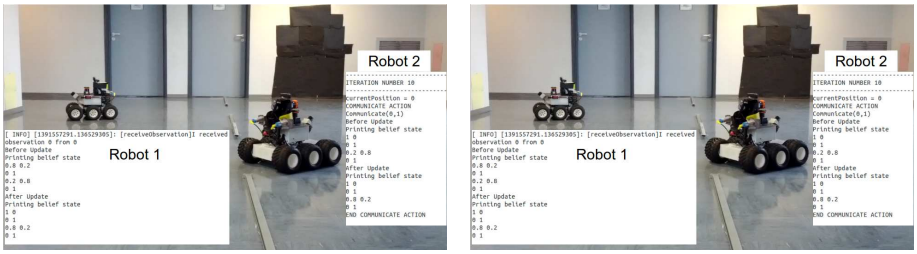
number of messages sent remains almost constant and much lower than the system with complete communication (Figure 2b).

Those experiments seem to validate the hypothesis that choosing relevant information to communicate may improve system performances while reducing the number of communications.

Real Robots We implemented the model on two μ-troopers in a simple scenario where two rooms must be explored. In the figure 3a, robot 1 decided to explore room 1 and robot 2 decided to explore room 2. Since beliefs of robot 2 about the environment are very accurate and it believes that robot 1 has incorrect beliefs, robot 2 decides to communicate the observation 0 to robot 1. Robot 1 receives this observation and updates its beliefs accordingly. This figure presents a case where an agent has approximated beliefs concerning the beliefs of the other agents. However this approximation does not prevent the robots from completing the mission and reaching a stable belief state, as presented on Figure 3b.

5 Conclusion and Prospects

We have introduced a new model of agent-based relevance as well as a decision process to perform active information gathering with a multiagent system. Each agent computes



(a) Sharing the exploration (b) End of exploration

Fig. 3: Exploration with μ-troopers

the relevance of an observation regarding itself or another agent to decide whether it should explore a particular zone or communicate this observation. The relevance of an observation is a compromise between the novelty, modeled by Kullback-Leibler ratio, and the certainty of an observation, modeled by Entropy measure. Therefore it may be tuned depending on the environment considered. In a static environment, as presented in the experiments, the certainty of an observation is more important than its novelty. However, in a highly dynamic environment, the novelty of an information may be the most important. The system has been implemented and tested on real robots. Results show that this approach is more efficient than a fully-communicating system.

The decision process we described focuses on relevance and reasoning on belief states to perform active information gathering. In the system presented in this paper, an agent is able to communicate any observation from the observation set if it is relevant. Therefore, an agent may communicate an observation it has never directly received. Future works would maintain a history of observations received and allow an agent to communicate only observations it has previously received. Moreover, the beliefs about the beliefs of other agents are updated only when there is an explicit communication. We plan to work on a less naive method : since the same policy is used by all agents, we may update those beliefs more often by assuming the action taken by other agents. Finally, future works would consider the integration of the system presented in non-epistemic POMDPs.

References

1. Borlund, P. : The concept of relevance in IR. Journal of the American Society for information Science and Technology. 54(10), 913-925 (2003)
2. Salton, G., Buckley, C. : Improving retrieval performance by relevance feedback. Readings in information retrieval. 24. 5 (1997)
3. Baeza-Yates, R., Ribeiro-Neto, B. et al : Modern information retrieval. ACM press New York. 463 (1999)
4. Floridi, L. : Understanding epistemic relevance. Erkenntnis. 69(1), 69-92 (2008)
5. Roussel, S., Cholvy, L. : Cooperative interpersonal communication and Relevant information. ESSLLI Workshop on Logical Methods for Social Concepts. (2009)
6. Ponzoni Carvalho Chanel, C., Teichteil-Königsbuch, F., Lesire, C. : POMDP-based online target detection and recognition for autonomous UAVs. ECAI. 955-960 (2012)

7. Araya, M., Buffet, O., Thomas, V., Charpillet, F. : A POMDP extension with belief-dependent rewards. Advances in Neural Information Processing System. 64-72 (2010)
8. Agmon, N., Kaminka, G., Kraus, S. : Multi-Robot Adversarial Patrolling Facing a Full-Knowledge Opponent. Journal of Artificial Intelligence. (2011)
9. Basilico, N., Gatti, N., Amigoni, F.: Developing a Deterministic Patrolling Strategy for Security Agents. Web Intelligence and Intelligent Agent Technologies. 2. 565-572 (2009)
10. Paruchuri, P., Tambe, M., Ordóñez, F., Kraus, S. : Security in multiagent systems by policy randomization. Proceedings of the fifth international joint conference on Autonomous agents and multiagent systems. 273-280 (2006)
11. Melo, F., Veloso, M. : Decentralized MDPs with sparse interactions. Artificial Intelligence. 175(11). 1757-1789 (2011)
12. Ferranti, E., Trigoni, N., Levene, M. : Brick Mortar: an on-line multi-agent exploration algorithm. International Conference on Robotics and Automation. 761-767 (2007)
13. Matignon, L., Jeapierre, L., Mouaddib, A. : Coordinated Multi-Robot Exploration Under Communication Constraints Using Decentralized Markov Decision Processes. AAAI. (2012)
14. Seuken, S., Zilberstein, S. : Formal models and algorithms for decentralized decision making under uncertainty. Autonomous Agents and Multi-Agent Systems. 17(2). 190-250 (2008)
15. Bernstein, D., Givan, R., Immerman, N., Zilberstein, S. : The complexity of decentralized control of Markov decision processes. Mathematics of operations research. 27(4). 819-840 (2002)
16. Goldman, C., Zilberstein, S. : Decentralized control of cooperative systems: Categorization and complexity analysis. J. Artif. Intell. Res.(JAIR). 22. 143-174 (2004)
17. Seuken, S., Zilberstein, S. : Memory-Bounded Dynamic Programming for DEC-POMDPs. IJCAI. 2009-2015 (2007)
18. Spaan, M., Veiga, T., Lima, P. : Active cooperative perception in network robot systems using POMDPs. International Conference on Intelligent Robots and Systems. 4800-4805 (2010)
19. Pynadath, D. : The Communicative Multiagent Team Decision Problem: Analyzing Teamwork Theories and Models. Journal of Artificial Intelligence Research. 16. 389-423 (2002)
20. Peshkin, L., Kim, K., Meuleau, N., Kaelbling, L. : Learning to cooperate via policy search. Proceedings of the Sixteenth conference on Uncertainty in artificial intelligence. 489-496 (2000)
21. Roth, M., Simmons, R., Veloso, M. : Reasoning about joint beliefs for execution-time communication decisions. Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems. 786-793 (2005)
22. Cassandra, A., Kaelbling, L., Littman, M. : Acting optimally in partially observable stochastic domains. AAAI. 94. 1023-1028 (1994)
23. Smallwood, R., Sondik, E. : he optimal control of partially observable Markov processes over a finite horizon. Operations Research. 21(5). 1071-1088 (1973)
24. Hansen, E., Feng, Z. : Dynamic Programming for POMDPs Using a Factored State Representation. (2000)
25. Sabbadin, R., Lang, J., Ravoanjanahry, N. : Purely epistemic markov decision processes. Proceedings of the national conference on artificial intelligence. 22(2). 1057 (2007)
26. Sigaud, O., Buffet, O. et al : Markov Decision Processes in Artificial Intelligence. ISTE-Jonh Wiley & Sons (2010)
27. Poupart, P. : Exploiting structure to efficiently solve large scale partially observable Markov decision processes. Citeseer. (2005)
28. Porta, J., Vlassis, N., Spaan, M., Poupart, P. : Point-based value iteration for continuous POMDPs. The Journal of Machine Learning Research. 7. 2329-2367 (2006)
29. Hoey, J., Poupart, P. : Solving POMDPs with continuous or large discrete observation spaces. International Joint Conference on Artificial Intelligence. 19. 1332 (2005)