Distributed decision-theoretic active perception for multi-robot active information gathering

Jennifer Renoux and **Abdel-Illah Mouaddib**[1] and **Simon LeGloannec**[2]

Abstract. Multirobot systems have made tremendous progress in exploration and surveillance. However, information gathering tasks remain passive: the agents receive information from their sensors without acting in order to gather it. In this paper, we present a model and an algorithm for active information gathering using the information relevance concept. 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 Hellinger distance in a reward function to access the relevance of their beliefs and their divergence with the other robots. This model allows to derive a policy for gathering information to make the entropy low and a communication policy to reduce the divergence between robot's beliefs. An experimental scenario has been developed for an indoor information gathering mission. Our model has been compared to two different systems : one without communication and one communicating each received observation. The results show that our approach is more efficient than both systems.

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 to gather information : the system needs to create a complete and accurate view of the situation. Therefore it 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.

Classical framework and designs used to make exploration and surveillance - such as Game Theory, Partially Observable Markov Decision Processes - are usually not designed to have information gathering as a goal. Information gathering is used as a mean to reach another goal and is usually passive : the agent receives data about its environment but does not act to collect specifically those data. Some systems have been developed for mono-agent active information gathering but the plunge to multi-agent systems has not been taken to our knowledge. We suggest in this paper a formal definition of the relevance of a piece of information as well as a new model dedicated to information gathering that is able to explore actively its environment and communicate relevant information. The result is a distributed decision model where each robot uses the information relevance to derive an information communication policy. This approach has been implemented on a multirobot system. The system has been compared to a system without communication and to a system in which robots communicate each observation they receive at each time step. The results show that our approach is more efficient than both other approaches.

Section 2 presents some background. Section 3 presents the proposed model to do active sensing with a multi-robot system. It defines an agent-oriented relevance degree and describes an augmented 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

Robots situated in an environment perceive a huge amount of data through their sensors. To reason on these data they need to process them to extract higher-level features. However the interest of a feature for a given robot depends on several parameters such as the situation, the problem to be dealt etc. Since it is counterproductive to communicate or 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 [7] 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 [10] argued that there were no agent-oriented relevance theory and suggested a base of epistemic relevance. In this study information is considered as an answer to a query that an agent may ask. The degree of relevance of information is defined proportionally to how it answers a question weighted by the probability that the agent receiving information asks this question. Roussel and Cholvy [21] deepened Floridi's study in the context

[1] University of Caen, France, email:firstName.lastName@unicaen.fr
[2] Airbus Defence and Space, France, email: simon.legloannec@cassidian.com

of BDI agents and multimodal logic. They created a new operator to model the relevance and defined some properties of this operator, setting a theoretical base of relevance for logical models. This study is based on propositional logic and does not enable to reason on imprecise knowledge. To our knowledge, a theoretical base of agent-oriented relevance is still missing to allow reasoning on imprecise knowledge.

Active information gathering

Using relevance, a robot 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. [17] 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. [2] 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

Multiagent information gathering has gained interest in recent years, particularly in the patrolling and exploration domains. However, most of the patrolling system suggested in the literature are modeled using Game Theory and research studies focused on the best strategy for some patrolling agents to adopt [1] . These strategies are usually deterministic and do not depend on the situation [4]. Even if some strategies adopt some kind of random movements [16] they can not adapt to a new situation. Multiagent exploration problems are also well studied [15]. Ferranti et al. [9] suggested an algorithm to make rapid exploration in small closed environment. They assume that robots communicate using tags in the environment. Matignon et al. [14] suggested a decision process based on Markov Decision Processes (MDP) to make exploration, based on the fact that agents are not always able to interact with each others. Both works address the issue of coordination between agents. Robots sense their environment passively and their actions are not driven by information gathering. Partially Observable Markov Decision Processes (POMDP), as well as their multiagent extensions [23] are also often used to make information gathering. However, solving a multiagent POMDP is a problem NEXP-complete [5] and usually assume that communication is free and instantaneous. Therefore it is usually not possible to apply those frameworks in real problems. Roth et al. [20] 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 $DOM(X_k)$.

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

Let $DOM(\mathcal{E})$ be the set of all possible values for all $X_k \in \mathcal{E}$.

$$DOM(\mathcal{E}) = \{DOM(X_k) \forall X_k \in \mathcal{E}\}$$

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 $b_{i,t}^k$ 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. When receiving a new observation, agent a_i updates its beliefs as follows : $\mathcal{B}_{i,t+1}^{\mathcal{E}} = update(\mathcal{B}_{i,t}^{\mathcal{E}}, o_k)$ [8, 25]

First of all we considered that observations received are true, that is to say reflect the real state of the system and are not due to observational error. As a matter of fact, an observation cannot be relevant if it is a false observation [10]. We discuss about 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
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 $\mathcal{B}_{i,t+1}^{\mathcal{E}}$ and $\mathcal{B}_{i,t}^{\mathcal{E}}$ are distant from each other. In information gathering, the Kullback-Leibler divergence [13] is often used to measure the dissimilarity between two probability distributions. However, this measure is not symmetric and really difficult to bound. Therefore, we decided to use instead the Hellinger distance[6], which quantifies the similarity between two probability distributions. In addition of being symmetric, the Hellinger distance is bounded between 0 and 1 which simplifies its use in this problem.

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

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

ϵ is a fixed problem-dependent threshold and $D_H(\mathcal{B}_{i,t}^{\mathcal{E}} || \mathcal{B}_{i,t+1}^{\mathcal{E}})$ is the Hellinger distance between the two belief states and defined by

$$D_H(\mathcal{B}_{i,t}^{\mathcal{E}} || \mathcal{B}_{i,t+1}^{\mathcal{E}}) = \sum_{X_k \in \mathcal{E}} \frac{1}{\sqrt{2}} \sqrt{\sum_{p=1}^{n} (\sqrt{b_{i,t}^k(x_p)} - \sqrt{b_{i,t+1}^k(x_p)})^2} \tag{2}$$

where $b^k_{i,t}(x_p)$ is the belief of agent a_i that the random variable X_k takes value x_p.

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

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*

$$rel_i(o_k) = (1-\delta) D_H(\mathcal{B}^{\mathcal{E}}_{i,t} || \mathcal{B}^{\mathcal{E}}_{i,t+1}) + \delta \frac{H(\mathcal{B}^{\mathcal{E}}_{i,t}) - H(\mathcal{B}^{\mathcal{E}}_{i,t+1})}{H_{max}} \tag{4}$$

with $\mathcal{B}^{\mathcal{E}}_{i,t+1} = update(\mathcal{B}^{\mathcal{E}}_{i,t}, o_k)$, H_{max} is the maximum entropy, and δ is a weight associated to the dynamism of the system.

Because the Hellinger distance is bounded, it is easy to show that the above relevance degree is also bounded.

3.2 Decision Process for multiagent active information gathering

Let a multiagent system be defined as a tuple $\langle \mathcal{E}, \mathcal{AG}, \mathcal{B}, \mathcal{D} \rangle$ with $\mathcal{E}$ being the environment as defined previously, $\mathcal{AG}$ being the set of agents, $\mathcal{B}$ being the set of all agent's beliefs on the environment and $\mathcal{D}$ being the set of all agent's decision functions $\mathcal{D} = \{\mathcal{D}_i, \forall i \in \mathcal{AG}\}$. Each $\mathcal{D}_i$ is represented as a Factored Partially Observable Markov Decision Process (FPOMDP)[11].

A POMDP is a tuple $\langle \mathcal{S}, \mathcal{A}, \mathcal{O}, \mathcal{T}, \omega, \mathcal{R}, b_0 \rangle$ with

- $\mathcal{S}$ is a set of states, corresponding to the joint instantiations of the random variables $X_k \in \mathcal{E}$.
- $\mathcal{A}$ is a set of epistemic actions [22]
- $\mathcal{O}$ is the set of observations
- $\mathcal{T}$ is the transition function, defined as in [11]. For more information on how to represent it, see [24] and [19].
- ω is the observation function, defined as in [19].
- $\mathcal{R}$ is the reward function
- b_0 is the initial belief state

Set of actions

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

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

The size of the action set is :

$$|\mathcal{A}| = |\mathcal{A}_{Explore}| + |\mathcal{A}_{Communicate}| \tag{5}$$

$$|\mathcal{A}| = |\mathcal{E}| + (2^{|\mathcal{O}|} - 1) \times (2^{|\mathcal{AG}|} - 1) \tag{6}$$

It is obvious that this amount will create combinatorial explosion even with a small number of observations and agents. It is however possible to reduce the size of the action set with a reasonable hypothesis. We may assume the simple case where an agent communicates only one observation to one other agent. In that case equation 5 becomes

$$|\mathcal{A}| = |\mathcal{E}| + |\mathcal{O}| \times |\mathcal{AG}| \tag{7}$$

In the remainder of this paper we will only consider this simplified $Communication$-type action. However one should note that the model is still valid if one needs to relax the hypothesis.

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. This $\mathcal{O}_{Explore}$ set of observations depends on the problem considered. When performing a $Communicate$-type action the agent receives an observation stating that the message has been properly sent or not. Therefore two observations are possible for any $Communicate$-type action:

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

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. As a matter of fact, to avoid agents 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 an extended belief state of an agent a_i at time t be defined as following :*

$$\mathcal{B}_{i,t} = \mathcal{B}^{\mathcal{E}}_{i,t} \cup \mathcal{B}^{j,\mathcal{E}}_{i,t} \tag{8}$$

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 $B^{j,\mathcal{E}}_{i,t}$ is an approximation of $B^{\mathcal{E}}_{j,t}$. We may use a matrix representation. Rows represent the different random variables describing the environment and columns represent agent's i beliefs on each agent's beliefs, including itself :

$$\mathcal{B}_{i,t} = \begin{pmatrix} b^{1,1}_{i,t} & \dots & b^{j,1}_{i,t} & \dots & b^{i,1}_{i,t} & \dots & b^{|\mathcal{AG}|,1}_{i,t} \\ \vdots & & \vdots & & \vdots & & \vdots \\ b^{1,k}_{i,t} & \dots & b^{j,k}_{i,t} & \dots & b^{i,k}_{i,t} & \dots & b^{|\mathcal{AG}|,k}_{i,t} \\ \vdots & & \vdots & & \vdots & & \vdots \\ b^{1,|\mathcal{E}|}_{i,t} & \dots & b^{j,|\mathcal{E}|}_{i,t} & \dots & b^{i,|\mathcal{E}|}_{i,t} & \dots & b^{|\mathcal{AG}|,|\mathcal{E}|}_{i,t} \end{pmatrix} \tag{9}$$

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 :

$$\mathcal{B}_{i,t+1}(s') = \frac{\omega(o_k, s', a)\sum_{s\in\mathcal{S}} p(s'|s,a)\mathcal{B}_{i,t}(s)}{\sum_{s\in\mathcal{S}}\sum_{s''\in\mathcal{S}}\omega(o_k, s'', a)p(s''|s,a)b^{i,k}_{i,t}} \quad (10)$$

In the matrix representation, the columns that may be updated are:

$$\mathcal{B}_{i,t+1} = \begin{pmatrix} b^{1,1}_{i,t} & \dots & \mathbf{b^{j,1}_{i,t+1}} & \dots & \mathbf{b^{i,1}_{i,t+1}} & \dots & b^{|\mathcal{AG}|,1}_{i,t} \\ \vdots & & \vdots & & \vdots & & \vdots \\ b^{1,k}_{i,t} & \dots & \mathbf{b^{j,k}_{i,t+1}} & \dots & \mathbf{b^{i,k}_{i,t+1}} & \dots & b^{|\mathcal{AG}|,k}_{i,t} \\ \vdots & & \vdots & & \vdots & & \vdots \\ b^{1,|\mathcal{E}|}_{i,t} & \dots & \mathbf{b^{j,|E|}_{i,t+1}} & \dots & \mathbf{b^{i,|E|}_{i,t+1}} & \dots & b^{|\mathcal{AG}|,|\mathcal{E}|}_{i,t} \end{pmatrix}$$

Reward function

The best action to perform at a given time is set by a policy, computed considering a relevance based reward function. This 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. In order to prevent agent to communicate false information, we weight the relevance of an observation by the probability that this observation may be true. 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. On top of that, agents should ensure that they have homogeneous beliefs. As a matter of fact, agents having different beliefs means that one agent at least is wrong. Therefore the reward function for communication includes a term to ensure that the different between agent's own beliefs and its beliefs on beliefs of other agents are low. The reward function is thus defined as follows:

$$R(\mathcal{B}_{i,t}, Exp(X_k)) = \sum_{s\in\mathcal{S}}\sum_{o_k\in\mathcal{O}}\mathcal{B}_{i,t}(s)\omega(o_k, s, a)rel_i(o_k) - C_{Exp(X_k)} \quad (11)$$

$$\begin{aligned} &R(\mathcal{B}_{i,t}, Comm(o_k, a_j)) = \\ &\sum_{s\in\mathcal{S}}\mathcal{B}_{i,t}(s)\omega(o_k, s, a)rel_j(o_k) \\ &+\gamma(D_H(\mathcal{B}^{\mathcal{E}}_{i,t}||\mathcal{B}^{j,\mathcal{E}}_{i,t}) - D_H(\mathcal{B}^{\mathcal{E}}_{i,t+1}||\mathcal{B}^{j,\mathcal{E}}_{i,t+1})) \\ &-C_{Comm(o_k,a_j)} \end{aligned} \quad (12)$$

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. This cost may represent battery loss, bandwith used etc.

Resolution

To solve the POMDP presented above, we can rely on classic MDP algorithms. As a matter of fact, actions are epistemic and so 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 correspond 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 defined as following

The transition function τ of the Belief MDP is defined as following :

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

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

where $U_t = \{o_k \in \mathcal{O}$, such as, $\mathcal{B}_{i,t+1} = update(\mathcal{B}_{i,t}, o_k)\}$ 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.

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

$$V(\mathcal{B}_{i,t}) = R(\mathcal{B}_{i,t}) + \max_{a\in\mathcal{A}}\int_{\mathcal{B}'_{i,t}}\tau(\mathcal{B}_{i,t}, a, \mathcal{B}'_{i,t})V(\mathcal{B}'_{i,t}) \quad (13)$$

Using discretization techniques (by discretizing the probability distributions) we may transform equation 13 in :

$$V(\mathcal{B}_{i,t}) = R(\mathcal{B}_{i,t}) + \max_{a\in\mathcal{A}}\sum_{\mathcal{B}'_{i,t}\in Samples}\tau(\mathcal{B}_{i,t}, a, \mathcal{B}'_{i,t})V(\mathcal{B}'_{i,t}) \quad (14)$$

Then, any technique from the literature may be used to solve this belief-MDP [18] [12]. However, we face a problem of scalability. Indeed, the size of the state space grows exponentially with the number of variables. To overcome this issue, we can exploit the structure of the problem considered. In a lot of surveillance and exploration applications, some variables are independent from the others. Therefore the set of variables can be divided into independent subsets of variables.

$$\mathcal{S} = \{S_i\}$$

such as $\forall j, k, \forall X_n \in S_j, \forall X_m \in S_k$, X_n and X_m are independent. In a similar way, the probability to receive an observation often depends on a given subset of variables. The observation function can also be divided. Using these assumptions, it is possible to build several independent sub-MDPs, to compute an optimal policy on these sub-MDPs and then to build the global policy again.

4 Experiments

The model has been implemented on a exploration application : two robots have to explore an environment made of 4 zones connected to each others. For each zone, two observations are possible : *emptyRoom* and *notEmptyRoom*. The optimal policy has been computed using different probabilities to obtain a false observation (referred as "threshold"). The system has been simulated 50 times with each

set of parameters. To compute the solution, we used custom implementation of a value-iteration algorithm. To evaluate the policy, we measured the evolution of the belief's entropy during the exploration. We compared the results with those obtained using a system without communication (referred as "no communication system") and using a system where agents communicate each observation they receive immediately (referred as "dummy communication system"). Results are presented on figures 1c, 1b and 1a. We observe that our system performs better in the three cases. Indeed, it enables to obtain a lower belief's entropy, that is to say a better belief state with regards to information gathering. It is also faster to obtain a good belief state. These results seem to prove that a relevant communication is better than no communication, which was expected, but also than a dummy communication. This latest result is more surprising but can finally be explained. Indeed, in a system with dummy communication, the agents will communicate all the observations they receive including observations created after an observational error. Therefore, false observations will be spread among the system. In our approach, the agent communicates an observation only if its beliefs support this observation enough. Therefore, our system is more resistant to observational errors.

5 Conclusion and Prospects

We have introduced a new model of agent-based relevance as well as a decision process to make active information gathering with a multiagent system. Each agent computes 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 Hellinger distance, and the certainty of an observation, modeled by Entropy measure. Therefore it may be tuned depending on the environment considered. In 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 make active information gathering. In the system presented in this paper, an agent is able to communicate any observation from the observations 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, this work focuses on epistemic-POMDPs, that is to say POMDPs where the agents' actions don't modify the real state of the system. Future works would consider the integration of this model in non-epistemic POMDPs and take into account agents' variables such as its position.

REFERENCES

[1] Noa Agmon, Gal A. Kaminka, and Sarit Kraus, 'Multi-robot adversarial patrolling facing a full-knowledge opponent', *Journal of Artificial Intelligence*, (2011).

[2] Mauricio Araya, Olivier Buffet, Vincent Thomas, and Françcois Charpillet, 'A pomdp extension with belief-dependent rewards', in *Advances in Neural Information Processing Systems*, pp. 64–72, (2010).

[3] Ricardo Baeza-Yates, Berthier Ribeiro-Neto, et al., *Modern information retrieval*, volume 463, ACM press New York, 1999.

[4] N. Basilico, N. Gatti, and F. Amigoni, 'Developing a deterministic patrolling strategy for security agents', in *Web Intelligence and Intelligent Agent Technologies, 2009. WI-IAT '09. IEEE/WIC/ACM International Joint Conferences on*, volume 2, pp. 565–572, (2009).

[5] Daniel S Bernstein, Robert Givan, Neil Immerman, and Shlomo Zilberstein, 'The complexity of decentralized control of markov decision processes', *Mathematics of operations research*, **27**(4), 819–840, (2002).

[6] Anil Bhattacharyya, 'On a measure of divergence between two multinomial populations', *Sankhyā: The Indian Journal of Statistics*, 401–406, (1946).

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

[8] Anthony R Cassandra, Leslie Pack Kaelbling, and Michael L Littman, 'Acting optimally in partially observable stochastic domains', in *AAAI*, volume 94, pp. 1023–1028, (1994).

[9] E. Ferranti, N. Trigoni, and M. Levene, 'Brick mortar: an on-line multi-agent exploration algorithm', in *Robotics and Automation, 2007 IEEE International Conference on*, pp. 761–767, (2007).

[10] Luciano Floridi, 'Understanding epistemic relevance', *Erkenntnis*, **69**(1), 69–92, (2008).

[11] Eric A Hansen and Zhengzhu Feng, 'Dynamic programming for pomdps using a factored state representation.', (2000).

[12] Jesse Hoey and Pascal Poupart, 'Solving pomdps with continuous or large discrete observation spaces', in *International Joint Conference on Artificial Intelligence*, volume 19, p. 1332, (2005).

[13] Solomon Kullback and Richard A Leibler, 'On information and sufficiency', *The Annals of Mathematical Statistics*, **22**(1), 79–86, (1951).

[14] Laëtitia Matignon, Laurent Jeanpierre, and Abdel-Illah Mouaddib, 'Coordinated multi-robot exploration under communication constraints using decentralized markov decision processes.', in *AAAI*, (2012).

[15] Francisco S Melo and Manuela Veloso, 'Decentralized mdps with sparse interactions', *Artificial Intelligence*, **175**(11), 1757–1789, (2011).

[16] Praveen Paruchuri, Milind Tambe, Fernando Ordóñez, and Sarit Kraus, 'Security in multiagent systems by policy randomization', in *Proceedings of the fifth international joint conference on Autonomous agents and multiagent systems*, pp. 273–280. ACM, (2006).

[17] Caroline Ponzoni Carvalho Chanel, Florent Teichteil-Königsbuch, and Charles Lesire, 'Pomdp-based online target detection and recognition for autonomous uavs.', in *ECAI*, pp. 955–960, (2012).

[18] Josep M Porta, Nikos Vlassis, Matthijs TJ Spaan, and Pascal Poupart, 'Point-based value iteration for continuous pomdps', *The Journal of Machine Learning Research*, **7**, 2329–2367, (2006).

[19] Pascal Poupart, *Exploiting structure to efficiently solve large scale partially observable Markov decision processes*, Ph.D. dissertation, Citeseer, 2005.

[20] Maayan Roth, Reid Simmons, and Manuela Veloso, 'Reasoning about joint beliefs for execution-time communication decisions', in *Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*, p. 786793, (2005).

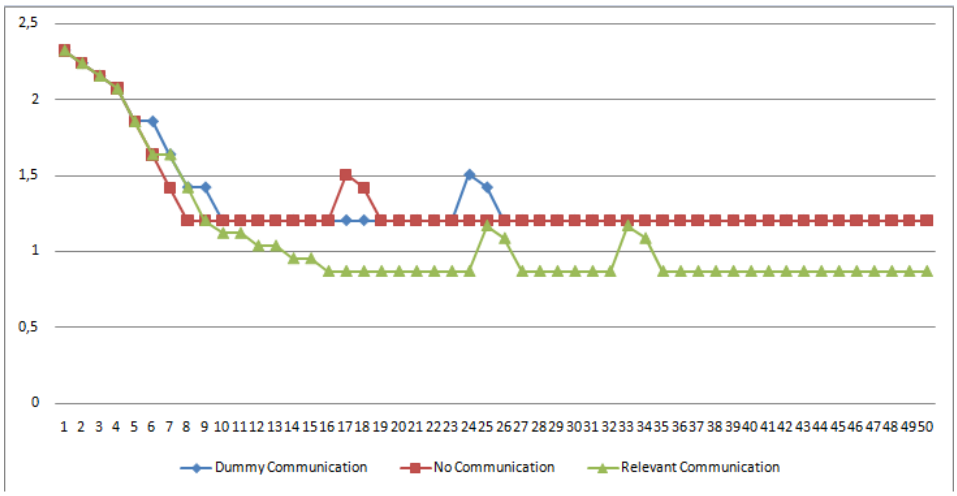
[21] Stéphanie Roussel and Laurence Cholvy, 'Cooperative interpersonal communication and relevant information', in *ESSLLI Workshop on Logical Methods for Social Concepts, Bordeaux*. Citeseer, (2009).

[22] Régis Sabbadin, Jérôme Lang, and Nasolo Ravoanjanahry, 'Purely epistemic markov decision processes', in *Proceedings of the national conference on artificial intelligence*, volume 22, p. 1057. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, (2007).

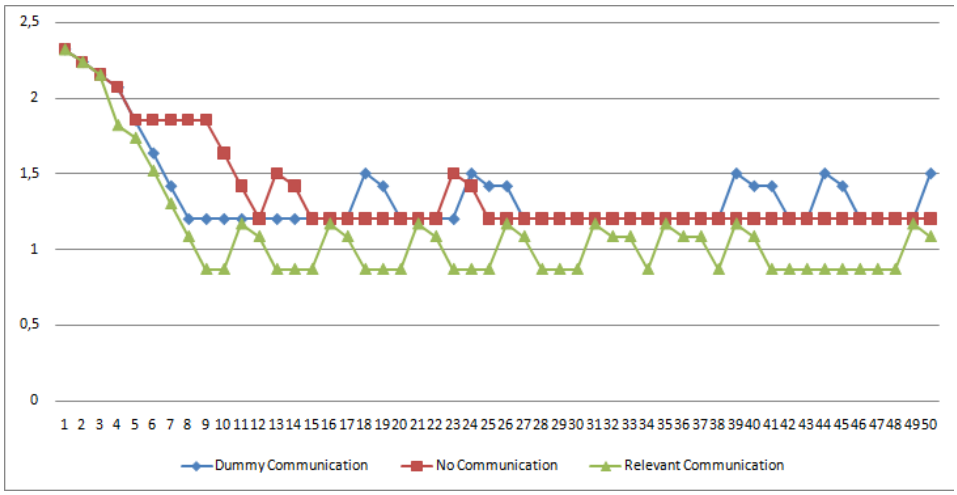
[23] Sven Seuken and Shlomo Zilberstein, 'Formal models and algorithms for decentralized decision making under uncertainty', *Autonomous Agents and Multi-Agent Systems*, **17**(2), 190–250, (2008).

[24] Olivier Sigaud, Olivier Buffet, et al., 'Markov decision processes in artificial intelligence', *ISTE-Jonh Wiley & Sons*, (2010).

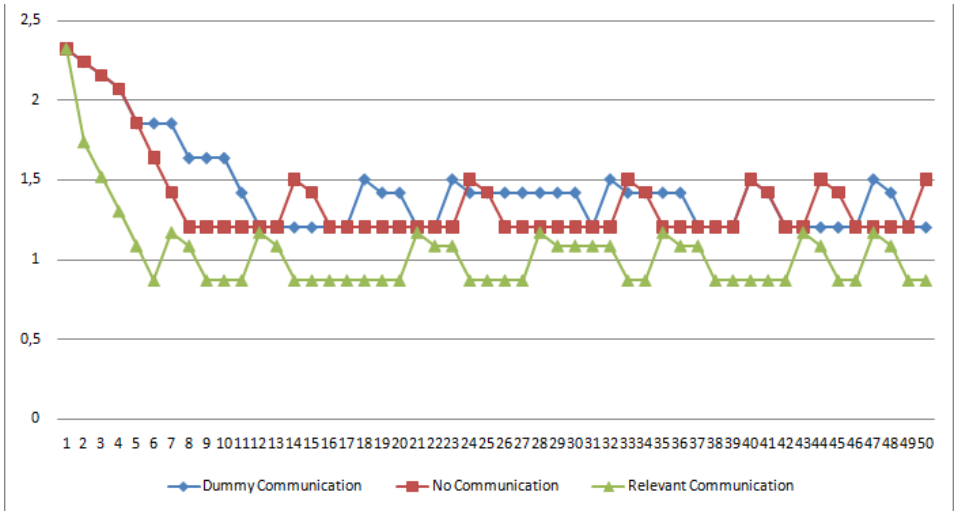
[25] Richard D Smallwood and Edward J Sondik, 'The optimal control of partially observable markov processes over a finite horizon', *Operations Research*, **21**(5), 1071–1088, (1973).



(a) Threshold = 90



(b) Threshold = 80



(c) Threshold = 70

Figure 1: Evalutation of the three policies at different thresholds. The X-Axis represents the iterations. The Y-Axis is the entropy.