OPTIMAL PLANNING FOR AUTONOMOUS AGENTS UNDER TIME AND RESOURCE UNCERTAINTY

Aurélie Beynier, Laurent Jeanpierre, Abdel-Illah Mouaddib
Bd Marechal Juin, Campus II
BP5186, 14032 Caen Cedex, France

Keywords: Decision Theory, Mobile robots, Autonomous Agents, Planning under Uncertainty.

Abstract: In this paper we develop an approach for planning under time and resource uncertainty with task dependencies. We follow the line of research described by (Bresina et al., 2002), overcoming limitations of existing approaches: they only handle simple time constraints and assume a simple model of durations and resources' uncertainty.In many domains, such as space applications (rovers, satellites), these assumptions are invalid. Our approach considers temporal and resource constraints and fully deals with their uncertainty. From a mission graph, we build a MDP that can be optimally solved by dynamic programming, even for large mission graphs.

1 INTRODUCTION

In this paper we develop an approach of planning under time and resource uncertainty along the research lines described in a challenge paper of (Bresina et al., 2002). In that paper, authors claim that existing methods fail because they suffer from many limitations where some of them consist of: (1) they only handle simple time constraints, (2) they assume a simple model of uncertainty concerning action durations and resource consumption. In many domains such as space applications (rovers, satellites), these assumptions are not valid. We present an approach that relaxes those assumptions by using a model of tasks with complex dependencies, uncertain execution time and probabilistic resource consumptions. In this approach, we consider the following: (1) a mission is an acyclic graph of tasks that expresses complex dependencies between tasks, (2) a task has a temporal interval during which it can be executed, (3) durations and resource consumptions of tasks are uncertain. This class of problems is found in some rover scenarios where the objective of the rover is to maximize the overall value of the mission: given the mission graph, we want the mission to be executed optimally by an autonomous agent.

The major objectives of future space missions (Estlin et al., 1999; Bresina et al., 2002) will consist of maximizing science return and enabling certain types of activities by using a robust approach. To show the significance of problems we deal with, let us give some examples of robotic vehicles where we express the different constraints we consider in this paper:

- *Time windows*: a number of tasks need to be done at particular but approximate times: for example "about noon", "at sunrise". There is no explicit time window but we can represent these using time windows or soft constraints on a precise time. Examples include atmospheric measurements (look at the sun through the atmosphere must be done at sunrise, sunset, and noon). Since the rover will be power-constrained, most of the rover's operations will occur between 10:00 and 15:00 to make sure that there is enough sunlight to operate. Driving will certainly happen during that part of the day. Communication also has to start at a particular time since this requires synchronization with Earth. There are operations that cannot be done outside of a time window – night-time operations of cameras or daytime operations of the spectrometer. Other constraints can be considered such as illumination constraints (which involves time of day and position), setup times (warm-up or cool-down delays for instruments), time separation between images to infer 3D shape from shadows.
- *Precedence dependencies*: a tasks usually has preconditions that must hold before the agent can execute it. For instance, instruments must be turned on and calibrated, in order the agent to perform the measurements. Such dependencies lead to precedence constraints between the tasks.

Beynier A., Jeanpierre L. and Mouaddib A. (2006).
OPTIMAL PLANNING FOR AUTONOMOUS AGENTS UNDER TIME AND RESOURCE UNCERTAINTY.
In *Proceedings of the Third International Conference on Informatics in Control, Automation and Robotics*, pages 182-187
DOI: 10.5220/0001203901820187

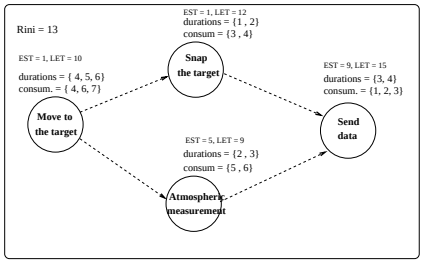

Figure 1: An acyclic graph of tasks.

• *Bounded resources*: many activities require power and data storage.

In summary, these scenarios share common characteristics such as: 1)Activities have an associated temporal window. 2)Task durations and resource consumptions are uncertain. 3)The temporal constraints are, in general, soft. 4)Precedence dependencies between tasks exist. 5)The mission can be represented by an acyclic graph. These scenarios are well suited to space-rovers because of the communication delay which prevents the agent from any interaction that could help it to achieve its tasks. For example, communication from Earth to Mars suffers from several minutes delays, and requires synchronization. Therefore missions generally include a whole workday, that is around one hundred tasks to accomplish. Mission plans are generally computed on Earth, transmitted to the distant agent once a day, and then executed autonomously during the rest of the day. However, even if we focus mainly on rovers, the approach we present in this paper can be applied to any problem formalized as a dependency graph. We could apply it to cooking a dinner; in this case, we would have a recipe to follow, ingredients to buy and mix in the proper order, and choices on which meal to prepare depending on the remaining time before the guests arrival...

Existing work on planning under uncertainty for rovers encounter difficulties in their application to real problems because most of them handle only simple time constraints (deadline) and instantaneous actions. More sophisticated techniques have been developed to deal with uncertain duration (Tsamardinos et al., 2003; Vidal and Ghallab, 1996; Morris et al., 2001) but they fail to optimally control the execution.

The requirements for a richer model of time and actions are more problematic for those planning techniques (Bresina et al., 2002). The techniques based on MDP and POMDP representations can be used to represent actions with uncertainty on their outcomes (Mouaddib and Zilberstein, 1998; Cardon et al., 2001) but they have difficulties when those actions involve temporal dependencies and uncertain durations. Given the temporal constraints of the rover problem, which are not purely Markovian, the approach should handle irregular multi-step transitions.

In this paper, we proposed a formalism that can deal with temporal and resource constraints to represent realistic problems. Our approach deals with uncertainty on task durations and resource consumptions. Like Semi-Markov Decision Processes (SMDPs) (Sutton et al., 1999), it considers different possible durations for each task. We propose to formalize the problem using an MDP that allows for temporal and resource constraints. This MDP is solved optimally by dynamic programming. We demonstrate that problems of a hundred tasks can be considered.

2 PRINCIPLES OF THE APPROACH

Let's consider a planetary rover that have a mission to complete. First, it must move to a target. Then, depending on the time and on available resources, it can snap the target or complete atmospheric measurements. To end its mission, the rover must send the data it has collected. As shown in Figure 1, this small mission can be represented by an acyclic graph. Each node stands for a task t_i and edges represent temporal constraints: the rover must move to the target before it can snap it. "Rini" is the initial resource rate. We consider only "or" nodes : the agent must snap the target or complete atmospheric measurements. In order to send the data, one of these predecessor tasks must be completed. Temporal constraints are indicated by the Earliest Start Time (EST) and the Latest End Time (LET) of each task.

In the rest of the paper, we assume that the mission graph is given. We will use the terms task, activity, and node interchangeably. A task will be denoted by t_i. Nonetheless, a task t_i differs from an action a: an action consists of a task and a start time.

Each task t_i is characterized by its time window and by the uncertainty about its duration $\delta_i(t_i)$ and about its resource consumption $\delta_r(t_i)$. When it is unambiguous, we will suppress the activity argument t_i for conciseness.

Temporal window of Tasks Each task is assigned a temporal window [EST, LET] during which it must be executed. The execution of the activity (both start and end times) must be included in this interval for the task to succeed.

Uncertain execution time The uncertainty on durations has been considered in several approaches developed in (Mouaddib and Zilberstein, 1998; Zilberstein and Mouaddib, 1999; Cardon et al., 2001). All

of them ignore the uncertainty on the start time. We show in this paper how extensions can be considered, taking different temporal constraints into account.

Uncertain resource consumption The consumption of resources (energy, memory, etc ...) is uncertain. As the agent has limited initial resources, it must consider resource constraints.

We assume that resource consumption and execution time are dependent. The representation adopted for these distributions is discrete.

Definition 1 *Let $\mathcal{SP}$ be a set of pairs $(\Delta t, \Delta r)$. We denote $P((\Delta t, \Delta r))$ the probability the task consumes Δr resources and Δt units of time.*

If resource consumption and execution time are independent, we simply have :

$$P((\Delta t, \Delta r)) = P(\Delta t).P(\Delta r).$$

Overview In order to formalize our problem as a MDP, we need further information about the tasks. In fact, we must know the possible execution intervals of each task and their probabilities. Once these information are known, the MDP can be defined and solved.

Our approach can be divided into several steps. It first propagates the temporal constraints and computes the set of possible time intervals for each task, computing theire probabilities. The MDP is then constructed using these information and solved by the Policy Iteration algorithm (Howard, 1960).

3 TEMPORAL INTERVAL AND PROBABILITY PROPAGATION

The decision process bases its decision on the remaining resources, the latest executed task and the execution interval of this task.

Many possible execution intervals exist for each task. In order to explore the whole state space, we need to know the set of possible execution intervals. Therefore, we developed an algorithm that computes all the possible time intervals from the mission graph, weighted by a probability. This probabilistic weight allows us to know the probability that an activity will be executed during a given interval of time.

The set of possible start times of a task is a subset of $\{EST, EST+1, \ldots, LET-\min\delta_i\}$ (EST: Earliest Start Time, LET: Latest End Time). We denote the Latest Start Time as $LST = LET - \min\delta_i$ where $\min\delta_i$ is the minimum duration of the task.

To start the execution of a task t_i, the agent must have executed one of the predecessors of t_i. The start time of t_i therefore depends on the end time of its predecessor. We can compute off-line all the possible end times of an activity's predecessors and consequently compute the possible start times of the activity. The algorithm is similar to the one described in (Bresina and Washington, 2000). The possible execution intervals I are determined by a simple forward propagation of temporal constraints in the graph. This propagation organizes the graph into levels such that: l_0 is the root of the graph, l_1 contains all nodes that are constrained only by the root node, ..., l_i contains all the successor nodes of the nodes in level l_{i-1}. For each node in a given level l_i, we compute all its possible execution intervals from its predecessors.

- *level* l_0: the start time and the end times of the root node (the first task of the mission) are computed as follows:
  - *start time*: $st(root) = EST(root)$
  - *end times*: $ET(root) = \{st(root) + \delta_i(root), \forall\delta_i(root)\}$

  Possible execution intervals of the root are given by $I = [st(root), et(root)]$ with $et(root) \in ET(root)$. Note that there is potentially a non-zero probability that some end times violate the deadline LET.
- *level* l_i: for each node in level l_i, the possible start times are the times at which the predecessor activities can finish. We first compute all the possible start times, and then we eliminate the start times that do not respect the constraints of earliest start time: $st < EST$, and latest start time: $st > LST$.

  For each possible start time, we compute the possible end times, accounting for the possible durations of the node. Note that potentially $st(node) + \delta_i(node) > LET(node)$

Many classical algorithms exist in literature, like PERT, but most of them don't deal with uncertainties on execution time.

As explained in the next section, we must compute the possible execution intervals to define the state space. Probabilities on these intervals must also be known to compute the transition probabilities.

The probability of an execution interval P_w depends on its start time and on the probability of its duration P_c.

A task cannot start before its predecessor finishes. Therefore, the probability on the start time depends on the predecessor' end time. We will consider a task t_i. $P_w(I|et(I'))$ denotes the probability that t_i is executed in the interval I when its predecessor finishes at $et(I')$. This value measures the probability that an activity starts at $st(I)$ and ends at $et(I)$.

In order to respect temporal constraints, an agent executes the next task as soon as possible: if it delays the execution of the next task, it will increase the probability of violating the deadline with no expected

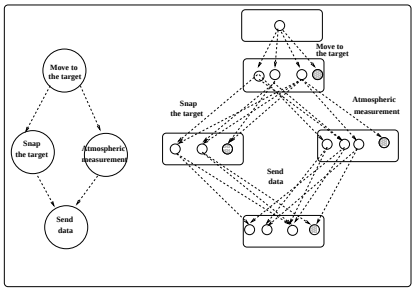

Figure 2: Relationship between the original graph structure and the MDP state space.

gain. If $st(I) \geq et(I')$ and there is no possible start time for t_i in $]et(I'), st(I)[$, the agent will start the execution at $st(I)$.

The probability $P_w(I|et(I'))$ of an execution interval I is defined as follows:

- *If $st(I)$ is the first possible start time such that $st(I) \geq et(I')$:* $P_w(I|et(I')) = P_c(et(I) - st(I))$
- *Otherwise:* $P_w(I|et(I')) = 0$

These probabilities are computed off-line during the forward propagation of temporal constraints. By propagating temporal constraints and their probabilities through the mission graph, the algorithm allows for developing the entire state space.

4 A DECISION MODEL: MDP

As mentioned above, we model this problem with a Markov Decision Process.

Definition 2 *A Markov Decision Process (MDP) is defined by a tuple $< \mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R} >$:*

- *$\mathcal{S}$ is a finite set of states s*
- *$\mathcal{A}$ is a finite set of stochastic actions a*
- *$\mathcal{P} : \mathcal{S} \times \mathcal{A} \times \mathcal{S}$ is a Markovian transition function that gives the probability the agent moves to a state s when it executes action a from s'.*
- *$\mathcal{R}$ is a reward function*

To solve our problem, we need to define the state space, the action set, the transition function and the reward function of our model. These components can be deduced from the initial mission graph. Figure 2 gives a representation of the relationship between the original graph structure, the state space and the transitions of the MDP. The left part of the figure stands for the mission graph. The right part represents the state space of the MDP and its transitions. Each box groups the states associated with a task t_i. Each node stands for a state and edges represent the transitions between the states.Some nodes have no successor: they are terminal states.

State Space At each decision step, the agent must decide for the execution of a task. The success of this execution depends on the temporal constraints and the available resources. If temporal constraints are not respected or if the agent lacks resources, the execution of the task fails and the agent moves to a failure state. Otherwise, the execution succeeds and the agent moves to a success state. The decision relies on three parameters: the latest executed task t_i, the available resources r, and the interval of time I during which task t_i has been executed. A state is therefore defined by a triplet $[t_i, r, I]$. States are fully observable since the temporal intervals are determined off-line and the resource consumption is observed by the agent as an environment feedback.

For each task, we develop a set of states by combining the execution intervals and the remaining resources. Possible resource rates are computed by propagating resource consumptions through the mission graph. The consequence of representing all the intervals and resource levels is that the state space becomes fully observable. The decision process can perform its action selection using the Bellman equation defined below.

Action Space From each non terminal state, the agent should choose which task to execute and when to start its execution. Note that the decision depends only on the current state and thus this process has the Markov property. The set of actions $a = E_i(st)$ to perform consists of *Executing task* t_i at time st, where t_i is a successor of the latest executed task. This action a is probabilistic since the duration and resource consumption of the task are uncertain. Standard MDPs used to represent one time unit actions. Our approach extends this model and allows for representing various possible durations for each action.

Transition model When the agent starts executing an action from a state $[t_i, r, I]$, it can move to various states. If temporal and resource constraints are respected, the execution succeeds. If the agent lacks resources, starts too late or violates the deadline, it fails. Failure states are considered as terminal states. It is straightforward to adapt our system to non terminal failure states. States associated with the last task of the mission are also terminal states.

Four kinds of transitions have to be considered:

- *Successful Transition:* The action allows the agent to move to $[t_{i+1}, r', I']$ where task t_{i+1} has been achieved during the interval I', respecting the EST and LET of this task, and r' are the remaining resources. The probability P_{SUC} of moving to the state $[t_{i+1}, r', I']$ is :

$$P_{SUC} = \sum_{(\Delta t, \Delta r) \in \mathcal{SP} \ | \Delta r \leq r \text{ and } st(I') + \Delta t = et(I')} P((\Delta r, \Delta t))$$

- *Too late start time (TLST) Transition:* The agent starts too late (after LST). In such a case, the agent moves to a failure state $[failure, r, [st, +\infty]]$ (in fact, the resource and interval arguments are unimportant since the state is terminal). The transition probability P_{TLST} is defined as: $P_{TLST} = Pr(st > LST)$.
- *Deadline met Transition:* The agent starts to execute the task at time st but the duration δ_i is so long that the deadline is met. This transition moves to the state $[failure, r, [st, +\infty]]$. The probability P_{DMT} of moving to this state is: $P_{DMT} = Pr(st \leq LST) \cdot \sum_{(\Delta t, \Delta r) \in \mathcal{SP} \ | \Delta r \leq r \text{ and } LET - \Delta t < st \leq LST} P((\Delta t, \Delta r))$
- *Insufficient resource Transition:* The execution requires more resources than available. The agent moves to a state $[failure, 0, [st, +\infty]]$. The probability P_{IRT} of moving to this state is: $P_{IRT} = \sum_{(\Delta t, \Delta r) \in \mathcal{SP} | \Delta r > r} P((\Delta t, \Delta r))$

Reward function Each time the agent finishes to execute a task within temporal, precedence and resource constraints, it obtains a reward which depends on the executed task. R_i is the reward the agent obtains when it has accomplished the task t_i.
As the number of tasks and start times are finite, the horizon of the MDP is finite. Since time is considered in each state, the MDP have no loop.

MDP resolution Once the MDP is defined, it can be solved optimally. To compute an optimal policy for the agent, we use dynamic programing and Bellman principle of optimality. A policy is a mapping from states to actions: for each state it dictates which action to execute. It allows the agent to execute its tasks autonomously. The value of each state relies on the immediate gain (reward) and the expected value that takes into account the various possible transitions. Each state is valued as follows:

$$V([t_i, r, I]) = R_i + \max_{\mathsf{E_k(st \geq et(I)), k=successor(t_i)}} (V')$$

where V' is the expected value of future tasks. We decompose V' as $V' = V_{SUC} + V_{FAIL}$ such that:

• *Expected Value of successful transition:*

$$V_{SUC} = \sum_{(\Delta t, \Delta r) \in \mathcal{SP} \ | \Delta r \leq r \text{ and } st(I') + \Delta t = et(I')} P((\Delta r, \Delta t).V([t_{i+1}, r - \Delta_r, I'])$$

This value is used when the task starts and finishes with success: there is enough resources and I' is the execution interval of the next activity.

• *Expected Value of failure transition:* Otherwise, the execution fails because the agent lacks resources or temporal constraints are violated.

$$V_{FAIL} = P_{FAIL} \cdot V([\textit{failure}, *, *])$$

with $P_{FAIL} = P_{TLST} + P_{DMT} + P_{IRT} = 1 - P_{SUC}$. Note that resources and intervals are unimportant.

The policy π of a state $[t_i, r, I]$ is given by:

$$\pi([t_i, r, I]) = R_i + \arg \max_{\mathsf{E_k(st \geq et(I)), k=successor(a_i)}} (V')$$

Optimality The MDP is easily solved using the standard dynamic programming algorithm *policy iteration*. Therefore, the obtained policy is optimal. Precedence constraints allows for ordering the policy revision: the tasks are considered from the leaves to the node. Thus, there is no need to iterate the algorithm since there is no loop.

5 RELATED WORK

There has been considerable work in planning under uncertainty that leads to two categories of planners: conformant planners and contingent planners. These planners are characterized by 2 important criteria: *representation of uncertainty* and *observability*. The first criterion, the *uncertainty representation*, has been addressed in two ways in many planners using *disjunction* or *probability* while the second criterion is composed of *Non-observability* (NO), *partial observability* (PO) or *a full observability* (FO) of planners. A survey on all classes of planners can be found in Blythe (Blythe, 1999) and Boutilier (Boutilier et al., 1999) where details are given on NO, PO, or FO disjunctive planners and on NO, PO or FO probabilistic planners. Let us just recall some of those planners: C-PLAN NO-disjunctive planner, Puccini PO-disjunctive planner, Warplan FO-disjunctive planner, Buridan NO probabilistic planner, POMDP, C-MAXPLAN PO probabilistic planners and JIC, MDP FO probabilistic planners. In this section we focus on why those planners are unsuitable for our concern and why our work is a contribution to overcome those limits. These planners encounter some difficulties in our domain of interest:

• *Model of time*: the existing planners do not support explicit time constraints nor complex temporal dependencies.

• *Model of actions*: the existing planners assume that actions are instantaneous.

• *Scalability*: the existing planners do not scale to large problems. For rover operations, a daily plan can involve on the order of a hundred operations, many of which having uncertain outcomes.

The approach we present in this paper meets the requirements for a rich model of time and actions and for scalability. It complements the work initiated in (Bresina and Washington, 2000) by using a similar model of time and utility distribution and by using

a decision-theoretic approach. The advantage of using such an approach is to achieve optimality. Another contribution consists of handling uncertainty on resource consumption combined with uncertainty on execution time.

In the MDP we present, actions are not instantaneous as in the previous planners and can deal with complex time constraints such as a temporal window of execution and temporal precedence constraints. We also show that our approach can solve large problems with a hundred operations, many of which are uncertain. Another requirement mentioned for the rover applications is concurrent actions. This problem is under development, taking advantage of some multiagent concepts : as soon as the rover needs to execute two actions, we consider those actions are concurrent agents. This new line of research allows for bridging the gap between the multiagent systems and the distributed MDP.

6 CONCLUSION AND FUTURE WORK

In this paper we presented an MDP planning technique that allows for a plan where operations have complex dependencies, time and resource constraints. The operations are organized in an acyclic graph where each operation has a temporal window during which it can be executed and an uncertain resource consumption and execution time. This approach is based on an MDP using a rich model of time and resources with dependencies between tasks. This technique allows us to deal with the variable duration of actions. We presented experimental results showing that our approach can scale to large robotic problems (a hundred of operations). Our approach also overcomes some of the limitations described in (Bresina et al., 2002). Indeed, our model is able to handle more complex time constraints and uncertainty on durations and resource consumptions. Moreover, as required in (Bresina et al., 2002), our system can consider plans of more than a hundred tasks.

In this paper, we have focused on planetary rover applications. Nonetheless, our approach is not restricted to space applications. It can be applied to any scenario formalized by an acyclic graph with temporal constraints. However, this approach needs to be extended to other requirements such as continuous variables. In our current version of the approach we use a discrete representation of time and resources. We are specially interested in finding tradeoffs between the scalability, execution errors and discretization. The other extension we are developing consists of dealing with multiple agents using a distributed MDP. Future work may concern the construction of the graphs that we consider as given in this paper.

REFERENCES

Blythe, J. (1999). *Planning under uncertainty in Dynamic domains*. PhD, Carnegie Mellon University.

Boutilier, C., Dean, T., and Hanks, S. (1999). Decision-theoretic planning: Structural asumptions and computational leverage. *Journal of Articicial Intelligence Research*, 1:1–93.

Bresina, J., Dearden, R., Meuleau, N., Ramkrishnan, S., Smith, D., and Washington, R. (2002). Planning under continuous time and resource uncertainty: A challenge for ai. In *Proceedings of the 18th Annual Conference on Uncertainty in Artificial Intelligence (UAI-02)*, pages 77–84, San Francisco, CA. Morgan Kaufmann Publishers.

Bresina, J. and Washington, R. (2000). Expected utility distributions for flexible contingent execution. In *AAAI Workshop on Representation issues for Real World Planning system*.

Cardon, S., Mouaddib, A., Zilberstein, S., and Washington, R. (2001). Adaptive control of acyclic progressive processing task structures. In *IJCAI*, pages 701–706.

Estlin, T. A., Gray, A., Mann, T., Rabideau, G., Castano, R., Chien, S., and Mjolsness, E. (1999). An integrated system for multi-rover scientific exploration. In *AAAI/IAAI*, pages 613–620.

Howard, R. A. (1960). *Dynamic Programming and Markov Processes*. MIT Press.

Morris, P., Muscettola, N., and Vidal, T. (2001). Dynamic control of plans with temporal uncertainty. In *IJCAI*, pages 494–502.

Mouaddib, A.-I. and Zilberstein, S. (1998). Optimal scheduling for dynamic progressive processing. In *ECAI-98*, pages 499–503.

Sutton, R. S., Precup, D., and Singh, S. P. (1999). Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning. volume 112, pages 181–211.

Tsamardinos, I., Pollack, M., and Ramakrishnan, S. (2003). Assessing the probability of legal execution of plans with temporal uncertainty. In *ICAPS Workshop on Planning under uncertainty and Incomplete information*.

Vidal, T. and Ghallab, M. (1996). Dealing with uncertain durations in temporal constraints networks dedicated to planning. In 12^{th} *ECAI*, pages 48–54.

Zilberstein, S. and Mouaddib, A.-I. (1999). Reactive control for dynamic progressive processing. In *IJCAI-99*, pages 1269–1273.