One-Shot Learning of Temporal Actions Models via Constraint Programming

Antonio Garrido and **Sergio Jiménez**

Abstract. We present a *Constraint Programming* (CP) formulation for learning temporal planning action models from the observation of a single plan execution (*one shot*). Inspired by the CSP approach to *temporal planning*, our CP formulation models *time-stamps* for states and actions, *causal-link* relationships, *threats* and effect *interferences* and evidences the connection between the tasks of *plan synthesis*, *plan validation* and *action model learning* in the temporal planning setting. The CP formulation is solver-independent so off-the-shelf CSP solvers can be used for the resolution of any of these three tasks. The performance of the CP formulation is assessed when learning and validating action models of several temporal planning domains specified in PDDL2.1.

1 INTRODUCTION

Temporal planning is an expressive planning model that relaxes the assumption of instantaneous actions of *classical planning* [9]. Actions in temporal planning are called *durative*, because each action has an associated duration and hence, the conditions/effects of durative actions may hold/happen at different times [6]. This means that *durative actions* can be executed in parallel and overlap in several different ways [3], and that valid solutions for temporal planning instances specify the precise time-stamp when durative actions start and end [14].

Despite the potential of state-of-the-art planners, their application to real world problems is still somewhat limited mainly because of the difficulty of specifying correct and complete planning models [16]. The more expressive the planning model, the more evident becomes this knowledge acquisition bottleneck that jeopardizes the usability of AI planning technology. With the aim of relieving this limitation there are growing efforts in the planning community for the machine learning of action models [17, 19, 21, 22]. There is a wide range of different approaches for learning classical action models from sequential plans: since pioneering learning systems like ARMS [21], we have seen systems able to learn action models with *quantifiers* [1, 25], from *noisy* actions or states [19, 22], from *null state information* [2], or from *incomplete* domain models [23, 24].

As far as we know this paper presents the first approach for learning action models for the *temporal planning* setting. While learning an action model for classical planning means computing the actions' conditions and effects that are consistent with the input observations, learning temporal action models requires additionally: i) identifying how conditions and effects are temporally distributed within the actions, and ii) estimate the action duration.

Most of the cited approaches for model learning are purely inductive and require large input datasets, e.g. hundreds of plan observations, to compute statistically significant models and focus on learning models from sequential plans for classical planning. With the aim of understanding better the connection between the learning of durative action models, *temporal planning* and the validation of temporal plans, this paper follows a radically different approach and studies the singular learning scenario where just the observation of a single plan execution (*one-shot*) is available.

The contributions of this work are two-fold:

1. The learning of action models from observations of plans with overlapping actions. This feature makes our approach appealing for learning action models from observations of multi-agent environments [7].
2. A solver-independent CP formulation that connnects the *learning* of planning action models with the *synthesis* and the *validation* of temporal plans. Off-the-shelf CSP solvers can be used for any of these tasks.

2 BACKGROUND

This section formalizes the *temporal planning* and *Constraint Satisfaction* models that we follow in this work.

2.1 Temporal Planning

We assume that *states* are factored into a set F of Boolean variables. A state s is a time-stamped assignment of values to all the variables in F. A *temporal planning problem* is a tuple $P = \langle F, I, G, A \rangle$ where the *initial state* I is a fully observed state (i.e. a total assigment of the state variables $|I| = |F|$) and that is time-stamped with $t = 0$; $G \subseteq F$ is a conjunction of *goal conditions* over the variables in F that defines the set of goal states; and A represents the set of *durative actions*.

A *durative action* has an associated duration and may have conditions/effects on F at different times [8, 20]. Like in PDDL and to compactly represent temporal planning problems, we assume that the state variables in F are intantiations of a given set of predicates Ψ and that durative actions in A are fully grounded from *action schemes* (also known as *operators*) . PDDL2.1 is the input representation language for the temporal track of the International Planning Competition (IPC) [6, 11]. According to PDDL2.1, a durative action $a \in A$ is defined with the following elements:

1. $\mathsf{dur}(a)$, a positive value indicating the *duration* of the action.
2. $\mathsf{cond}_s(a), \mathsf{cond}_o(a), \mathsf{cond}_e(a)$ representing the three types of action *conditions*. Unlike the *pre*conditions of classical actions, action conditions in PDDL2.1 must hold: before a is executed (*at start*), during the entire execution of a (*over all*) or when a finishes (*at end*), respectively.

3. $\mathsf{eff}_s(a)$ and $\mathsf{eff}_e(a)$ represent the two types of action effects. In PDDL2.1, effects can happen *at start* or *at end* of action a respectively, and can be either positive or negative (i.e. asserting or retracting variables).

PDDL2.1 is a restricted temporal planning model that defines the semantics of a *durative action* a as two discrete events, $\mathsf{start}(a)$ and $\mathsf{end}(a) = \mathsf{start}(a) + \mathsf{dur}(a)$. This means that if a starts on state s with time-stamp $\mathsf{start}(a)$, then $\mathsf{cond}_s(a)$ must hold in s. Ending action a in state s', with time-stamp $\mathsf{end}(a)$, means $\mathsf{cond}_e(a)$ must hold in s'. *Over all* conditions must hold at any state between s and s' or, in other words, throughout the closed interval $[\mathsf{start}(a)..\mathsf{end}(a)]$. Likewise, *at start* and *at end* effects are instantaneously applied at states s and s', respectively (continuous effects are not considered in this work). Figure 1 shows an example of two schemes for PDDL2.1 durative actions taken from the *driverlog* domain. The schema `board-truck` has a fixed duration while the duration of `drive-truck` depends on the driving time associated to the two given locations.

```
(:durative-action board-truck
  :parameters (?d - driver ?t - truck ?l - location)
  :duration (= ?duration 2)
  :condition (and (at start (at ?d ?l))
                  (at start (empty ?t))
                  (over all (at ?t ?l)))
  :effect (and (at start (not (at ?d ?l)))
               (at start (not (empty ?t)))
               (at end (driving ?d ?t))))


(:durative-action drive-truck
  :parameters (?t - truck ?l1 - location ?l2 - location
               ?d - driver)
  :duration (= ?duration (driving-time ?l1 ?l2))
  :condition (and (at start (at ?t ?l1))
                  (at start (link ?l1 ?l2))
                  (over all (driving ?d ?t)))
  :effect (and (at start (not (at ?t ?l1)))
               (at end (at ?t ?l2))))
```

Figure 1. Two action schemes of durative actions represented in PDDL2.1.

PDDL2.2 is an extension of the PDDL2.1 language that includes the notion of *Timed Initial Literal* [13], denoted as $\mathsf{til}(f, t)$, and representing that variable $f \in F$ becomes true at a certain time $t > 0$, independently of the actions in the plan [4]. TILs are useful to model *exogenous happenings*; for instance, a time window when a warehouse is open in a logistics scenario can be modeled with two timed initial literals as follows, $\mathsf{til}(\mathit{open}, 8)$ and $\mathsf{til}(\mathit{not\text{-}open}, 20)$.

A *temporal plan* is a set of pairs $\pi = \{(a_1, t_1), (a_2, t_2) \dots (a_n, t_n)\}$. Each pair (a_i, t_i) contains a durative action a_i and the $t_i = \mathsf{start}(a_i)$ time-stamp. The execution of a temporal plan starting from a given initial state I induces a state sequence formed by the union of all states $\{s_{t_i}, s_{t_i + \mathsf{dur}(a_i)}\}$, where there exists an initial state $s_0 = I$, and a state s_{end} that is the last state induced by the execution of the plan. Note then that sequential plans can be expressed as temporal plans but not the opposite. A *solution* to a given temporal planning problem P is a temporal plan π such that its execution, starting from the corresponding initial state, eventually reaches a state that meets the goal conditions, $G \subseteq s_{end}$. A solution is *optimal* iff it minimizes the plan *makespan* (i.e., the maximum $\mathsf{end}(a) = \mathsf{start}(a) + \mathsf{dur}(a)$ of an action in the plan).

2.2 Constraint Satisfaction

A *Constraint Satisfaction Problem* (CSP) is a tuple $\langle X, D, C \rangle$, where X is a set of finite-domain *variables*, D represents the *domain* for each of these variables and C is a set of *constraints* among the variables in X that bound their possible values in D.

A *solution* to a CSP as an assignment of values to all the variables in X that is *consistent* with all the input constraints. Given a CSP there may be many different solutions to that problem, i.e., different variable assignments that are *consistent* with the input constraints.

A *cost-function* can be defined to specify user preferences about the space of possible solutions. Given a CSP and cost-function, then an *optimal solution* is a variable assignment that is consistent with the constraints of the CSP and minimizes the value of the defined cost-function.

3 One-shot learning of temporal actions models

We formalize the task of the *one-shot learning of temporal action models* as a tuple $\mathcal{L} = \langle F, I, G, A?, O, C \rangle$ where:

```
(at driver1 loc1) (at truck1 loc1) (driving driver1 truck1)
(empty truck1) (path loc1 loc1) (link loc1 loc1)
```

Figure 2. Set of six *candidates* to appear in the conditions/effects of the ground action `board-truck(driver1,truck1,loc1)`.

- $\langle F, I, G, A? \rangle$ is a *temporal planning problem* such that the actions in $A?$ are *partially specified*. This means that the exact conditions/effects, their temporal annotation, and the duration of actions are unknown while the actions *header* (i.e., the *name* and *parameters* of each action) is known. With this regard, we say that a fluent $f \in F$ is a *candidate* to appear in the condition/effects of an action $aA?$ iff f appears in the set of FOL interpretations of the predicates over the action parameters $pars(a)$. For instance Figure 2 shows the set of six *candidates* to appear in the conditions/effect of the ground action `board-truck(driver1,truck1,loc1)`.

```
(:objects driver1 driver2 - driver
          truck1 truck2 - truck
          package1 package2 - obj
          s0 s1 s2 p1-0 p1-2 - location)

(:init (at driver1 s2) (at driver2 s2) (at truck1 s0)
       (empty truck1) (at truck2 s0) (empty truck2)
       (at package1 s0) (at package2 s0)
       (path s1 p1-0) (path p1-0 s1) (path s0 p1-0)
       (path p1-0 s0) (path s1 p1-2) (path p1-2 s1)
       (path s2 p1-2) (path p1-2 s2)
       (link s0 s1) (link s1 s0) (link s0 s2) (link s2 s0)
       (link s2 s1) (link s1 s2))

(:observation :time-stamp 56
              (at driver1 s1) (at truck1 s1))

(:observation :time-stamp 78
              (at package1 s0) (at package2 s0))
```

Figure 3. Example of a set of three observations (containing the fully observed initial state and two time-stamped partial states) extracted from the execution of a plan from the *driverlog* domain.

- O is the set of *observations* over a plan execution. Al least this set contains a full observation of the initial state (time-stamped with $t = 0$) and a final state observation that equals the goals G of the temporal planning problem (time-stamped with t_{end}, the makespan of the observed plan). Additionally, it can contain time-stamped observations of traversed intermediate *partial*

states[1] (e.g., $\mathsf{obs}(f,t)$ denotes that $f \in F$ was observed at time t) and the times when actions start and/or end their execution. For instance, $\mathsf{obs}(is\_start(a), t_i)$ represents that action a starts at t_i while $\mathsf{obs}(is\_end(a), t_j)$ represents that action a ends at t_j. Figure 3 shows an example of the observation of a plan execution taken from the *driverlog* domain.

$\forall truck, driver: \quad \neg empty(truck) \vee \neg driving(driver, truck).$
$\forall driver, loc_1, loc_2: \quad \neg at(driver, loc_1) \vee \neg at(driver, loc_2), \neq (loc_1, loc_2).$
$\forall driver, truck1, truck2: \neg driving(driver, truck_1) \vee \neg driving(driver, truck_2), \neq (truck_1, truck_2).$
$\forall driver1, driver2, truck: \neg driving(driver_1, truck) \vee \neg driving(driver_2, truck), \neq (driver_1, driver_2).$
$\forall driver, location, truck: \neg at(driver, location) \vee \neg driving(driver, truck).$

Figure 4. Examples of five mutex constraints for the *driverlog* domain.

- C is a set of *constraints* that captures domain-specific expert knowledge. In this work these constraints are mutually-exclusive (*mutex*) constraints that allow us to (1), deduce new observations and (2), prune action models inconsistent with these constraints. Figure 4 shows an example of a set of five mutex constraints for the *driverlog* domain.

A *solution* for the learning task $\mathcal{L}$ is a fully specified model of durative actions $\mathcal{A}$ such that the *conditions*, *effects*, their temporal annotations and the *duration* of any action in $\mathcal{A}$ are: i) completely specified; and ii) *consistent* with $\mathcal{L} = \langle F, I, G, A?, O, C\rangle$. By *consistent* we mean that there exists a valid plan that exclusively contains actions in $\mathcal{A}$ and whose execution starting in I, produces all the observations in O at the associated time-stamps, while it satisfies all constraints in C, and reaches a final state that satisfies G.

4 One-Shot learning of action models with CSPs

Given a one-shot learning task $\mathcal{L}$, as defined in Section 3, we automatically create a CSP, whose solution induces an action model that solves $\mathcal{L}$. This method is solver-independent and integrates well with previous work on *temporal planning* as CP [8, 20].

Table 1. The CSP variables, their domains and semantics.

ID	Variable	Domain	Description
X1	$\mathsf{start}(a)$	$[0..t_{end}]$	*Start time* of action a
X2	$\mathsf{end}(a)$	$[0..t_{end}]$	*End time* of action a
X3	$\mathsf{dur}(a)$	$[0..t_{end}]$	*Duration* of action a
X4	$\mathsf{is\_cond}(f,a)$	$\{0,1\}$	1 if f is a *condition* of a; 0 otherwise
X5	$\mathsf{is\_eff}(f,a)$	$\{0,1\}$	1 if f is an *effect* of a; 0 otherwise
X6.1	$\mathsf{req\_start}(f,a)$,		
X6.2	$\mathsf{req\_end}(f,a)$	$[0..t_{end}]$	Interval when action a requires f
X7	$\mathsf{sup}(f,a)$	$\{b\}_{b\in A?} \cup \emptyset$	Supporters for causal link $\langle b, f, a\rangle$
X8	$\mathsf{time}(f,a)$	$[0..t_{end}]$	Time when the effect f of a happens

4.1 The CSP variables

For each action $a \in A?$ and *candidate* f to appear in the conditions/effects of a, we create the following eight CSP variables that are shown in Table 1.

Variables X1 and X2 represent the times when a given action *starts* and *ends* while variable X3 represents the duration of that action. For simplicity, we model time in $\mathbb{Z}^+$ and bound all maximum times to the *makespan* of the observed plan (t_{end} observed in O). If the observation of t_{end} is unavailable, we consider a large enough domain for time. The value of these three variables (X1,X2 and X3) can be either observed in O or derived from the expression $\mathsf{end}(a) = \mathsf{start}(a) + \mathsf{dur}(a)$. Variables X4 and X5 are Boolean variables modeling whether f is actually a condition of a and whether f is an effect of a. Variables X6.1 and X6.2 model the closed interval throughout condition f must hold, provided that $\mathsf{is\_cond}(f,a)$=*true*. Variable X7 models *causal links* representing that actions b support f that is required by a. If f is not a condition of a ($\mathsf{is\_cond}(f,a)$=*false*) then $\mathsf{sup}(f,a)$=$\emptyset$, thus representing an empty supporter. Last but not least variable X8 models the time-stamp when effect f happens in a, provided $\mathsf{is\_eff}(f,a)$=*true*.

This formulation is able to model til and *observations*. The intuition is that modeling a til is analogous to modeling the full observation of the initial state (both represent information that is given at a particular time but externally to the execution of the plan). Likewise modeling an observation is analogous to the modeling of the goal of the planning task, as they both represent conditions that must be satisfied in the execution of the plan at a particular time. On the one hand, $\mathsf{til}(f,t)$ is modeled as a *dummy* action that starts at time t and has instantaneous duration ($\mathsf{start}(\mathsf{til}(f,t)) = t$ and $\mathsf{dur}(\mathsf{til}(f,t)) = 0$) with no conditions and the single effect f that happens at time t ($\mathsf{is\_eff}(f, \mathsf{til}(f,t))$=*true* and $\mathsf{time}(f, \mathsf{til}(f,t)) = t$). On the other hand, $\mathsf{obs}(f,t)$ is modeled as another *dummy* action that also starts at time t and has instantaneous duration ($\mathsf{start}(\mathsf{obs}(f,t)) = t$ and $\mathsf{dur}(\mathsf{obs}(f,t)) = 0$) but with only one condition f, which is the value observed for fact f ($\mathsf{is\_cond}(f, \mathsf{obs}(f,t))$=*true*, $\mathsf{sup}(f, \mathsf{obs}(f,t)) \neq \emptyset$ and $\mathsf{req\_start}(f, \mathsf{obs}(f,t)) = \mathsf{req\_end}(f, \mathsf{obs}(f,t)) = t$), and no effects at all.

Our variable formulation is accommodating a level of expressiveness beyond PDDL2.1 since it allows conditions/effects to be at any time, even outside the execution of the action. For example, we allow a condition f to hold in $\mathsf{start}(a)\pm 2$: $\mathsf{req\_start}(f,a) = \mathsf{start}(a) - 2$ and $\mathsf{req\_end}(f,a) = \mathsf{start}(a) + 2$. An effect f might also happen after the action ends e.g., $\mathsf{time}(f,a) = \mathsf{end}(a) + 2$.

4.2 The CSP constraints

Table 2 shows the constraints defined among the CSP variables of Table 1.

Constraint C1 represents the duration of an action. Constraint C2 indicates that actions must end before t_{end}. C3 is a double implication meaning that supporters are only required by action conditions. Constraint C4 forces to have valid values for the $[\mathsf{req\_start}, \mathsf{req\_end}]$ interval that defines when conditions are required. Constraint C5 models that the time when b supports f must be before a requires because of the causal link $\langle b, f, a\rangle$[2]. Given a causal link $\langle b, f, a\rangle$, constraint C6 avoids threats of actions c deleting f. The thread is solved via *promotion* or *demotion* [11], which means bringing $\mathsf{time}(not\text{-}f, c)$ backward or forward, respectively, in time. Constraint C7 avoids action a from being a supporter of f when $\mathsf{is\_eff}(f,a)$=*false*. Constraint C8 models the fact that the same action requires and deletes f; then the effect cannot happen before the condition. Note

[1] In this work, not all variables can be observed at any time; that is, we deal with *partial observations* (e.g. just a subset of variables is observable by associated sensors). Observations are noiseless, which means that if a value is observed, that is the actual value of that variable.

[2] $\mathsf{time}(f,b) < \mathsf{req\_start}(f,a)$ and not $\leq$ because, like in PDDL2.1 [6], our temporal planning model assumes $\epsilon > 0$ (ϵ denotes a small tolerance that implies no collision between the time when effect f is supported and when it is required). When time is modeled in $\mathbb{Z}^+$, $\epsilon = 1$ so $\leq$ becomes $<$.

Table 2. The CSP constraints and a brief description.

ID	Constraint	Description
C1	$\mathsf{end}(a) = \mathsf{start}(a) + \mathsf{dur}(a)$	Relationship among start, end and duration of a
C2	$\mathsf{end}(a) \leq \mathsf{start}(\mathsf{goal})$	Always goal is the last action of the plan
C3	**iff** ($\mathsf{is\_cond}(f, a)$=*false*) **then** $\mathsf{sup}(f, a) = \emptyset$	f is not a condition of $a \iff$ the supporter of f in a is $\emptyset$
C4	**if** ($\mathsf{is\_cond}(f, a)$=*true*) **then** $\mathsf{req\_start}(f, a) \leq \mathsf{req\_end}(f, a)$	$[\mathsf{req\_start}(f, a)..\mathsf{req\_end}(f, a)]$ is a valid interval
C5	**if** ($\mathsf{is\_eff}(f, b)$=*true*) **AND** ($\mathsf{is\_cond}(f, a)$=*true*) **AND** ($\mathsf{sup}(f, a) = b$)) **then** $\mathsf{time}(f, b) < \mathsf{req\_start}(f, a)$	Modeling the causal link $\langle b, f, a\rangle$: supporting f before it is required (obviously $b \neq \emptyset$)
C6	**if** ($\mathsf{is\_eff}(f, b)$=*true*) **AND** ($\mathsf{is\_cond}(f, a)$=*true*) **AND** ($\mathsf{is\_eff}(\textit{not-}f, c)$=*true*) **AND** ($\mathsf{sup}(f, a) = b$) **AND** ($c \neq a$) **then** ($\mathsf{time}(\textit{not-}f, c) < \mathsf{time}(f, b)$) **OR** ($\mathsf{time}(\textit{not-}f, c) > \mathsf{req\_end}(f, a)$)	Solving threat of c to causal link $\langle b, f, a\rangle$ by promotion or demotion (obviously $b \neq \emptyset$)
C7	**if** ($\mathsf{is\_eff}(f, a)$=*false*) **then forall** b that requires f: $\mathsf{sup}(f, b) \neq a$	a cannot be a supporter of f for any other action b
C8	**if** ($\mathsf{is\_cond}(f, a)$=*true*) **AND** ($\mathsf{is\_eff}(\textit{not-}f, a)$=*true*) **then** $\mathsf{time}(\textit{not-}f, a) \geq \mathsf{req\_end}(f, a)$	a requires and deletes f: the condition holds before the effect
C9	**if** ($\mathsf{is\_eff}(f, b)$=*true*) **AND** ($\mathsf{is\_eff}(\textit{not-}f, c)$=*true*) **then** $\mathsf{time}(f, b) \neq \mathsf{time}(\textit{not-}f, c)$	Solving effect interference at the same time (f and *not-*f)
C10	**forall** condition f_i and effect f_j of a: $\sum \mathsf{is\_cond}(f_i, a) \geq 1$ **AND** $\sum \mathsf{is\_eff}(f_j, a) \geq 1$	Every non-dummy action has at least one condition/effect

the $\geq$ inequality here: if one condition and one effect of the same action happen at the same time, the underlying semantics in planning considers the condition is checked instantly before the effect [6]. Constraint C9 prevents two actions have contradictory effects. Constraint C10 forces actions to have at least one condition and one effect (actions without effects are unnecessary for any plan). Constraint C9 applies to any type of action, including the dummy actions (init, goal, til and obs), while constraint C10 only applies to *non-dummy* actions.

Some conditions of Table 2 are redundant. For instance C5 and C6, $\mathsf{sup}(f, a) = b$ means obligatorily $\mathsf{is\_eff}(f, b) = \textit{true}$. We include them here to define an homogeneous formulation but they are not included in our implementation. For simplicity, the value of some unnecessary variables is not bounded in the table. For instance, if $\mathsf{is\_cond}(f, a)$=*false*, variables $\mathsf{req\_start}(f, a)$ and $\mathsf{req\_end}(f, a)$ become useless.

Table 3. Constraints to learn PDDL2.1-compliant action models.

ID	Constraint
C11.1	$(\mathsf{req\_start}(f, a) = \mathsf{start}(a))$ **OR** $(\mathsf{req\_start}(f, a) = \mathsf{end}(a))$
C11.2	$(\mathsf{req\_end}(f, a) = \mathsf{start}(a))$ **OR** $(\mathsf{req\_end}(f, a) = \mathsf{end}(a))$
C12	$(\mathsf{time}(f, a) = \mathsf{start}(a))$ **OR** $(\mathsf{time}(f, a) = \mathsf{end}(a))$
C13.1	$\forall f_i : (\forall a_j : \mathsf{req\_start}(f_i, a_j) = \mathsf{start}(a_j))$ **OR** $(\forall a_j : \mathsf{req\_start}(f_i, a_j) = \mathsf{end}(a_j))$
C13.2	$\forall f_i : (\forall a_j : \mathsf{req\_end}(f_i, a_j) = \mathsf{start}(a_j))$ **OR** $(\forall a_j : \mathsf{req\_end}(f_i, a_j) = \mathsf{end}(a_j))$
C14	$\forall f_i : (\forall a_j : \mathsf{time}(f_i, a_j) = \mathsf{start}(a_j))$ **OR** $(\forall a_j : \mathsf{time}(f_i, a_j) = \mathsf{end}(a_j))$
C15	$\forall a_i, a_j$ occurrences of the same action: $\mathsf{dur}(a_i) = \mathsf{dur}(a_j)$
C16	$\sum_{i=1}^{n} \mathsf{time}(f_i, a) > n \times \mathsf{start}(a)$

4.2.1 Constraints for the PDDL2.1 model

Here we show that making the presented CP formulation PDDL2.1-compliant is straightforward, by adding the constraints of Table 3 for all *non-dummy* actions.

Constraints C11 limit the *conditions* of an action to be only at *at start*, *over all* or *at end*. Likewise, constraint C12 limits the *effects* of an action to only happen *at start* or *at end*. In PDDL2.1 the structure of conditions/effects of all actions $\{a_j\}$ grounded from a particular operator are fixed. With this regard, constraints C13 makes the conditions of all $\{a_j\}$ equal and constraint C14 makes the effects of all $\{a_j\}$ equal. Constraint C15 makes the duration of all occurrences of the same action equals. Last but not least, constraint C16 forces all actions to have at least one of its *n*-effects *at end*. Actions with only *at start* effects would turn the value of the duration irrelevant and they could exceed the plan makespan. Although this constraint is not specific of PDDL2.1, we include it to learn more rational *durative actions*.

4.2.2 Mutex constraints

The set of mutexes that is given to a learning task $\mathcal{L}$ allows to infer new information in form of *dynamic observations*. In more detail, if two Boolean variables $\langle f_i, f_j\rangle$ are mutex they cannot hold simultaneously. This means that if we observe f_i, then we can infer $\neg f_j$ (despite $\neg f_j$ was not observed). This source of knowledge is specially relevant for the learning of *negative effects* when thre is absence of many observations. Mutex information helps to fill this void by inferring the observation of negated variables, which forces later to satisfy the *causal links* of negative variables.

Given a $\langle f_i, f_j\rangle$ mutex, in our *temporal planning* model, $\neg f_i$ does not necessarily implies f_j. To illustrate this see action `drive-truck` of Figure 1, where `(at ?t ?l1)` and `(at ?t ?l2)` are mutex (as defined in Figure 4). Effects `(not (at ?t ?l1))` and `(at ?t ?l2)` happen *at start* and *at end*, respectively. This means that the same truck cannot be in two locations simultaneously, but it is valid that the truck is, for some time, at no location (i.e., the truck is at some location `l2` some time after *not being* at location `l1`). These situations do not happen in STRIPS, where actions have instantaneous effects, so if $\langle f_i, f_j\rangle$ are mutex f_i implies *not-*f_j and vice versa.

Mutex-constraints can be exploited in a pre-proces step for completing the input observations given in a one-shot learning task $\mathcal{L}$. Furthermore, *dynamic observations* can be created to exploit the mutex constraints for any generated intermediate state. This include states that where not observed but that are inferred by the CSP solutions. Given a mutex $\langle f_i, f_j\rangle$ it means that, immediately after a asserts f_i, we need to ensure the observation *not-*f_j. This is done while performing the CSP search, and if $\mathsf{is\_eff}(f_i, a)$ takes the value *true*, then the next observation is added: $\mathsf{obs}(\textit{not-}f_j, \mathsf{time}(f_i, a)+\epsilon)$. The time of the observation cannot be just $\mathsf{time}(f_i, a)$, as we first need to assert f_i and one ϵ later observe *not-*f_j. Adding the variables and constraints for this new observation is trivial for *Dynamic CSPs* (DCSPs), in which the original formulation can be altered. Otherwise, we need to statically define a new type of observation $\mathsf{obs}(f_i, a, \textit{not-}f_j)$, where a supports f_i which is mutex with f_j and, consequently, we will need to observe *not-*f_j. The difference *w.r.t.* an original obs is twofold: i) the observation time is now initially unknown, and ii) the observation will be activated or not according to the following constraints:

if ($\mathsf{is\_eff}(f_i, a)$=*true*) **then** ($\mathsf{start}(\mathsf{obs}(f_i, a, \textit{not-}f_j)) = \mathsf{time}(f_i, a) + \epsilon$) **AND** ($\mathsf{is\_cond}(\textit{not-}f_j, \mathsf{obs}(f_i, a, \textit{not-}f_j))$=*true*)
else $\mathsf{is\_cond}(\textit{not-}f_j, \mathsf{obs}(f_i, a, \textit{not-}f_j))$=*false*

4.3 The CSP cost functions

The *conditions* of actions that are not deleted by any action are specialy difficult to be learned with a CSP that exclusively implements a

pure satisfiabiliy aproach. This is an issue when attempting to learn action models in which *static predicates* appear in the action *conditions* [12]. For instance the `(link ?l1 ?l2)` condition of the `drive-truck` action showed in the Figure 1.

This issue can be addressed extending the CP formulation to not only deal with the satisfaction of hard constraints but also to optimize a given cost function that defines the user preferences among different possible solutions. In more detail, the cost function is used to produce solutions that are *consistent* with the given input knowledge of the one-shot learning task $\mathcal{L}$ but to prefer solutions that support the input observations in a way that is as *tight* as possible.

To prefer this kind of *tight* support of the input observations we define the following two positive functions:

ϕ_1 *Causal-links*. This function counts the number of causal links that are created to support the provided observations.

ϕ_2 *Side-effects*. This function counts the number of possitive effects that are added by the actions in a plan but that do not build any causal link.

Our aim is to compute solutions to the CSP that minimize function ϕ_1 while function ϕ_2. is maximized. To achieve this we ask the CSP solve to *pareto optimize* functions ϕ_2 and $-\phi_1$(i.e. the negation of function 1).

5 A UNIFIED CP FORMULATION FOR PLANNING, VALIDATION AND LEARNING

Our CP formulation is connected to the tasks of plan *synthesis* and plan *validation* in the *temporal planning* settting. This connection lies on the fact that we can constrain the domain of the variables of our CP formulation to given known values. This feature is useful to leverage a priori knowledge of a given planning domain. For instance, because we have some available *prior knowledge* about the possible durations of a given action or because we already know that a given action produces for sure certain effects or requires some conditions. In this case the value of the corresponding variables is a priori specified while the remaining variables are then regular variables whose value will be determined solving the CSP.

If all the variables that represent the conditions, effects and duration of the actions are are a priori constrained to a single value (variables X3, X4 and X5) then solving the CSP is equivalent to solving a temporal planning task (that is synthesizing a plan that reaches a set of goals from certain initial stat and with a given action model). Likewise, if all the variables that represent when the different actions appear in a solution plan (when the start times of actions happen, variables X1,X2 and X3) then solving the CSP is equivalent to validating a plan in a given temporal planning problem.

What is more, we can either synthesize (or validate) a plan despite some of the variables that representing the conditions, effects or duration of an action do not have a fixed value (its value is initially unknown). That is planning (and validating plans) when the action model is partially specified. Therefore, that the plan validation ability of our CP formulation is beyond the functionality of VAL (the standard plan validation tool [14]) since it can address plan validation of partial, or even empty, action models and with partially observed plan traces (VAL requires both a full plan and a full action model for plan validation).

To wrap up, when addressing either *learning*, *planning* or *validating* tasks, our formulation is flexible to accept different levels of specification of the input knowledge:

- Partial knowledge of the conditions/effects of actions.
- Partial knowledge of actions durations (i.e. a set of possible durations).
- Partial knowledge of the plan to validate or synthesize.

To illustrate this, let us assume that the distribution of all (or just a few) conditions and/or effects is known and, in consequence, represented in the model $A?$ of $\mathcal{L}$. If a solution to the CSP is found, then that structure of conditions/effects is consistent for the learned model. On the contrary, if no solution is found that structure is inconsistent and cannot be explained. We can also represent known values for the durations by bounding the value of $\mathsf{dur}(a)$ variables to a given value. We can also introduce a priori knowledge about plans by bounding the value of the $\mathsf{start}(a)$ variables.

Last but not least this connection applies not only to temporal planning but also to the classical planning model, the vanilla model of AI planning where actions are instantaneous [9].

6 EVALUATION

[DE MOMENTO ESTO ESTA EN EL AIRE PORQUE NO SABEMOS COMO LO VAMOS A ABORDAR??]

The CP formulation has been implemented in Choco[3], an open-source Java library for constraint programming that provides an object-oriented API to state the constraints to be satisfied. Choco uses a static model of variables and constraints, i.e. it is not a DCSP.

The empirical evaluation of a learning task can be addressed from two perspectives. From a pure syntactic perspective, learning can be considered as an automated design task to create a new model that is similar to a reference (or *ground truth*) model. Consequently, the success of learning is an accuracy measure of how similar these two models are, which usually counts the number of differences (in terms of incorrect durations or distribution of conditions/effects). Unfortunately, there is not a unique reference model when learning temporal models at real-world problems. Also, a pure syntax-based measure usually returns misleading and pessimistic results, as it may count as incorrect a different duration or a change in the distribution of conditions/effects that really represent equivalent reformulations of the reference model. For instance, given the example of Figure 1, the condition learned `(over all (link ?from ?to))` would be counted as a difference in action `drive-truck`, as it is `at start` in the reference model; but it is, semantically speaking, even more correct. Analogously, some durations may differ from the reference model but they should not be counted as incorrect. As seen in section **??**, some learned durations cannot be granted, but the underlying model is still consistent. Therefore, performing a syntactic evaluation in learning is not always a good idea.

From a semantic perspective, learning can be considered as a classification task where we first learn a model from a training dataset, then tune the model on a validation test and, finally, asses the model on a test dataset. Our approach represents a one-shot learning task because we only use one plan sample to learn the model and no validation step is required. Therefore, the success of the learned model can be assessed by analyzing the success ratio of the learned model *vs.* all the unseen samples of a test dataset. In other words, we are interested in learning a model that fits as many samples of the test dataset as possible. This is the evaluation that we consider most valuable for learning, and define the success ratio as the percentage of samples of the test dataset that are consistent with the learned model. A higher

[3] `http://www.choco-solver.org`

ratio means that the learned model explains, or adequately fits, the observed constraints the test dataset imposes.

6.1 Learning from partially specified action models

We have run experiments on nine IPC planning domains. It is important to highlight that these domains are encoded in PDDL2.1, with the number of operators shown in Table 4, so we have included the constraints given in section 4.2.1. We first get the plans for these domains by using five planners (*LPG-Quality* [10], *LPG-Speed* [10], *TP* [15], *TFD* [5] and *TFLAP* [18]), where the planning time is limited to 100s. The actions and observations on each plan are automatically compiled into a CSP learning instance. Then, we run the one-shot learning task to get a temporal action model for each instance, where the learning time is limited to 100s on an Intel i5-6400 @ 2.70GHz with 8GB of RAM. In order to assess the quality of the learned model, we validate each model *vs.* the other models *w.r.t.* the *struct*ure, the *dur*ation and the *struct*ure+*dur*ation, as discussed in section 5. For instance, the *zenotravel* domain contains 78 instances, which means learning 78 models. Each model is validated by using the 77 remaining models, thus producing 78×77=6006 validations per struct, dur and struct+dur each. The value for each cell is the average success ratio. In *zenotravel*, the struct value means that the distribution of conditions/effects learned by using only one plan sample is consistent with all the samples used as dataset (100% of the 6006 validations), which is the perfect result, as also happens in *floortile* and *sokoban* domains. The dur value means the durations learned explain 68.83% of the dataset. This value is usually lower because any learned duration that leads to inconsistency in a sample counts as a failure. The struct+dur value means that the learned model explains entirely 35.76% of the samples. This value is always the lowest because a subtle structure or duration that leads to inconsistency in a sample counts as a failure. As seen in Table 4, the results are specially good, taking into consideration that we use only one sample to learn the temporal action model. These results depend on the domain size (number of operators, which need to be grounded), the relationships (causal links, threats and interferences) among the actions, and the size and quality of the plans.

Table 4. Number of operators to learn. Instances used for validation. Average success ratio of the one-shot learned model *vs.* the test dataset in different IPC planning domains.

	ops	ins	struct	dur	struct+dur
zenotravel	5	78	100%	68.83%	35.76%
driverlog	6	73	97.60%	44.86%	21.04%
depots	5	64	55.41%	76.22%	23.19%
rovers	9	84	78.84%	5.35%	0.17%
satellite	5	84	80.74%	57.13%	40.53%
storage	5	69	58.08%	70.10%	38.36%
floortile	7	17	100%	80.88%	48.90%
parking	4	49	86.69%	81.38%	54.89%
sokoban	3	51	100%	87.25%	79.96%

We have observed that some planners return plans with unnecessary actions, which has a negative impact for learning precise durations. The worst result is returned in the *rovers* domain, which models a group of planetary rovers to explore the planet they are on. Since there are many parallel actions for taking pictures/samples and navigation of multiple rovers, learning the duration and the structure+duration is particularly complex in this domain.

6.2 Learning from scratch

7 CONCLUSIONS

We have presented a purely declarative CP formulation, which is independent of any CSP solver, to address the learning of temporal action models. Learning in planning is specially interesting to recognize past behavior in order to predict and anticipate actions to improve decisions. The main contribution is a simple formulation that is automatically derived from the actions and observations on each plan execution, without the necessity of specific hand-coded domain knowledge. It is also flexible to support a very expressive temporal planning model, though it can be easily modified to be PDDL2.1-compliant. Formal properties are inherited from the formulation itself and the CSP solver. The formulation is correct because the definition of constraints to solve causal links, threats and effect interferences are supported, which avoids contradictions. It is also complete because the solution needs to be consistent with all the imposed constraints, while a complete exploration of the domain of each variable returns all the possible learned models in the form of alternative consistent solutions.

Unlike other approaches that need to learn from datasets with many samples, we perform a one-shot learning. This reduces both the size of the required datasets and the computation time. The one-shot learned models are very good and explain a high number of samples in the datasets used for testing. Moreover, the same CP formulation is valid for learning and for validation, by simply adding constraints to the variables. This is an advantage, as the same formulation allows us to carry out different tasks: from entirely learning, partial learning/validation (structure and/or duration) to entirely plan validation. According to our experiments, learning the structure of the actions in a one-shot way leads to representative enough models, but learning the precise durations is more difficult, and even impossible, when many actions are executed in parallel.

Our CP formulation can be adapted straightforward to address learning, planning or validation tasks within the classical planning model. In this case actions cannot have conditions *overall* or *at end* as well as they cannot have *at start* effects. Therefore the variables representing this kind of information can be removed from the CSP model (or be set to `false`). Further the duration of any action is fixed to one unit [15]. Finally, our CP formulation can be represented and solved by Satisfiability Modulo Theories, which is part of our current work.

REFERENCES

[1] Eyal Amir and Allen Chang, 'Learning partially observable deterministic action models', *Journal of Artificial Intelligence Research*, **33**, 349–402, (2008).

[2] S. N. Cresswell, T.L. McCluskey, and M.M West, 'Acquiring planning domain models using LOCM', *The Knowledge Engineering Review*, **28(2)**, 195–213, (2013).

[3] William Cushing, Subbarao Kambhampati, Daniel S Weld, et al., 'When is temporal planning really temporal?', in *Proceedings of the 20th international joint conference on Artificial intelligence*, pp. 1852–1859. Morgan Kaufmann Publishers Inc., (2007).

[4] S. Edelkamp and J. Hoffmann, 'PDDL2.2: the language for the classical part of IPC–4', in *Proc. of the International Conference on Automated Planning and Scheduling (ICAPS-04) – International Planning Competition*, pp. 2–6, (2004).

[5] Patrick Eyerich, Robert Mattmüller, and Gabriele Röger, 'Using the context-enhanced additive heuristic for temporal and numeric planning', in *Nineteenth International Conference on Automated Planning and Scheduling*, (2009).

[6] Maria Fox and Derek Long, 'PDDL2.1: An extension to PDDL for expressing temporal planning domains', *Journal of artificial intelligence research*, **20**, 61–124, (2003).

[7] Daniel Furelos Blanco, Antonio Bucchiarone, and Anders Jonsson, 'Carpool: Collective adaptation using concurrent planning', in *AAMAS 2018. 17th International Conference on Autonomous Agents and Multiagent Systems; 2018 Jul 10-15; Stockholm, Sweden.[Richland]: IFAAMAS; 2018.* International Foundation for Autonomous Agents and Multiagent Systems (IFAAMAS), (2018).

[8] Antonio Garrido, Marlene Arangu, and Eva Onaindia, 'A constraint programming formulation for planning: from plan scheduling to plan generation', *Journal of Scheduling*, **12**(3), 227–256, (2009).

[9] Hector Geffner and Blai Bonet, 'A concise introduction to models and methods for automated planning', *Synthesis Lectures on Artificial Intelligence and Machine Learning*, **8**(1), 1–141, (2013).

[10] Alfonso Gerevini, Alessandro Saetti, and Ivan Serina, 'Planning through stochastic local search and temporal action graphs in LPG', *Journal of Artificial Intelligence Research*, **20**, 239–290, (2003).

[11] Malik Ghallab, Dana Nau, and Paolo Traverso, *Automated Planning: theory and practice*, Elsevier, 2004.

[12] Peter Gregory and Stephen Cresswell, 'Domain model acquisition in the presence of static relations in the lop system', in *Twenty-Fifth International Conference on Automated Planning and Scheduling*, (2015).

[13] J. Hoffmann and S. Edelkamp, 'The deterministic part of IPC-4: an overview', *Journal of Artificial Intelligence Research*, **24**, 519–579, (2005).

[14] Richard Howey, Derek Long, and Maria Fox, 'VAL: Automatic plan validation, continuous effects and mixed initiative planning using PDDL', in *16th IEEE International Conference on Tools with Artificial Intelligence. ICTAI 2004*, pp. 294–301. IEEE, (2004).

[15] Sergio Jiménez, Anders Jonsson, and Héctor Palacios, 'Temporal planning with required concurrency using classical planning', in *Proceedings of the 25th International Conference on Automated Planning and Scheduling (ICAPS)*, (2015).

[16] Subbarao Kambhampati, 'Model-lite planning for the web age masses: The challenges of planning with incomplete and evolving domain models', in *Proceedings of the National Conference on Artificial Intelligence (AAAI-07)*, volume 22(2), pp. 1601–1604, (2007).

[17] Jirí Kucera and Roman Barták, 'LOUGA: learning planning operators using genetic algorithms', in *Pacific Rim Knowledge Acquisition Workshop, PKAW-18*, pp. 124–138, (2018).

[18] Eliseo Marzal, Laura Sebastia, and Eva Onaindia, 'Temporal landmark graphs for solving overconstrained planning problems', *Knowledge-Based Systems*, **106**, 14–25, (2016).

[19] Kira Mourão, Luke S. Zettlemoyer, Ronald P. A. Petrick, and Mark Steedman, 'Learning STRIPS operators from noisy and incomplete observations', in *Conference on Uncertainty in Artificial Intelligence, UAI-12*, pp. 614–623, (2012).

[20] Vincent Vidal and Héctor Geffner, 'Branching and pruning: An optimal temporal pocl planner based on constraint programming', *Artificial Intelligence*, **170**(3), 298–335, (2006).

[21] Qiang Yang, Kangheng Wu, and Yunfei Jiang, 'Learning action models from plan examples using weighted MAX-SAT', *Artificial Intelligence*, **171**(2-3), 107–143, (2007).

[22] Hankz Hankui Zhuo and Subbarao Kambhampati, 'Action-model acquisition from noisy plan traces', in *International Joint Conference on Artificial Intelligence, IJCAI-13*, pp. 2444–2450, (2013).

[23] Hankz Hankui Zhuo and Subbarao Kambhampati, 'Model-lite planning: Case-based vs. model-based approaches', *Artificial Intelligence*, **246**, 1–21, (2017).

[24] Hankz Hankui Zhuo, Tuan Anh Nguyen, and Subbarao Kambhampati, 'Refining incomplete planning domain models through plan traces', in *International Joint Conference on Artificial Intelligence, IJCAI-13*, pp. 2451–2458, (2013).

[25] Hankz Hankui Zhuo, Qiang Yang, Derek Hao Hu, and Lei Li, 'Learning complex action models with quantifiers and logical implications', *Artificial Intelligence*, **174**(18), 1540–1569, (2010).