

Learning and Evaluation of STRIPS Action Models with Classical Planning

Diego Aineto^a, Sergio Jiménez Celorrio^a, Eva Onaindia^a

^a*Department of Computer Systems and Computation, Universitat Politècnica de València, Spain*

Abstract

This paper presents a novel approach for learning STRIPS action models from observations of plan executions that compiles this learning task into classical planning. The compilation approach is flexible to various amount and kind of available input knowledge; the learning examples can range from plans (with their corresponding initial and final states), sequences of state observations or just a set of initial and final states (where no intermediate action or state is given). The compilation accepts also partially specified action models and can be used to validate whether the observation of a plan execution follows a given STRIPS action model, even if this model is not fully specified. What is more, the compilation is extensible to assess how well a STRIPS action model matches observations of plan executions. The performance of our compilation approach is evaluated learning action models for a wide range of classical planning domains from the International Planning Competition (IPC) and assessing the learned models with respect to (1), test sets of observations of plan executions and (2), the true models.

Keywords: Classical planning, Planning and learning, Learning action models, Generalized planning

1. Introduction

Besides *plan synthesis* [1], planning action models are also useful for *plan/goal recognition* [2]. At both planning tasks, an automated planner is required to reason about action models that correctly and completely capture the possible world transitions [3]. Unfortunately, building planning action models is complex, even for planning experts, and this knowledge acquisition task is a bottleneck that limits the potential of AI planning [4].

Machine Learning (ML) has shown to be able to compute a wide range of different kinds of models from examples [5]. The application of inductive ML to the learning of STRIPS action models, the vanilla action model for planning [6], is not straightforward though:

- The *input* to ML algorithms (the learning/training data) usually are finite vectors encoding the value of fixed features in a given set of objects. The input for learning planning action models are observations of plan executions (where each plan possibly has a different length and involve a different number of objects).
- The *output* of ML algorithms usually is a scalar value (an integer, in the case of *classification* tasks, or a real value, in the case of *regression* tasks). When learning STRIPS action models the output is, for each action, the sets of preconditions, negative and positive effects, that define the possible state transitions.

Motivated by recent advances on the synthesis of different kinds of generative models with classical planning [7, 8, 9, 10], this paper introduces an innovative approach for learning STRIPS action models that can be defined as a classical planning compilation. A solution to the classical planning task that results from our compilation is a

sequence of actions that determines the learned action model, i.e. the preconditions and effects of the target STRIPS operator schemas.

The compilation approach is appealing by itself, because it leverages off-the-shelf planners and because it opens up a way towards *bootstrapping* planning action models, enabling a planner to gradually learn/update its action model. Moreover, the practicality of the compilation allow us to report learning results over a wide range of IPC planning domains. Apart from these, the compilation approach presents the following contributions:

1. *Flexibility in the input knowledge.* The compilation is flexible to various amounts and kind of available input knowledge. Learning examples can range from a set of plans (with their corresponding initial and final states) or state observations, to just a set of initial and final states where no intermediate action or state is observed. The compilation also accepts previous knowledge about the structure of the actions in the form of partially specified action models, where some preconditions and effects are a priori known.
2. *Model Evaluation.* The compilation can assess how well a STRIPS action model matches a given set of observations of plan executions. Our compilation is extensible to accept a learned model as input besides the observations of plan executions. This extension allows us to transform the input model into a new model that induces the observations whilst assessing the amount of edition required by the input model to induce the given observations.

A first description of the compilation previously appeared in several conference papers [11]. Compared to the conference papers, the present paper includes the following novel material:

- Unifies the formulation of our approach for learning STRIPS action models from state observations and from observed plans.
- Shows that our compilation is also valid for the particular scenario where the observed states are not full states but *partial states*, in the sense that some of the observed fluents (either with positive or negative value) are missing.
- Proposes a distance metric to assess how well a STRIPS action model matches a given set of plans. Defines the compilation to compute this distance as well as the edit distance between two given STRIPS action models.

Section 2 reviews related work on learning planning action models. Section 3 formalizes the classical planning model with *conditional effects* (a requirement of the proposed compilation) and the STRIPS action model (the output of the addressed learning task). Section 4 formalizes the learning of STRIPS action models with regard to different amounts of available input knowledge. Sections 5 and 6 describe our compilation approach for addressing the formalized learning tasks and how the compilation is also extensible to assess learned action models. Section 7 reports the data collected in a two-fold empirical evaluation of our learning approach: First, the learned STRIPS action models are tested with observations of plan executions and second, the learned models are compared to the actual models. Finally, Section 8 discusses the strengths and weaknesses of the compilation approach and proposes several opportunities for future research.

2. Related work

Back in the 90's various systems aimed learning operators mostly via interaction with the environment. LIVE captured and formulated observable features of objects and used them to acquire and refine operators [12]. OBSERVER updated preconditions and effects by removing and adding facts, respectively, accordingly to observations [13]. These early works were based on lifting the observed states supported by exploratory plans or external teachers, but none provided a theoretical justification for this second source of knowledge.

More recent work on learning planning action models [14] shows that although learning STRIPS operators from pure interaction with the environment requires an exponential number of samples, access to an external teacher can provide solution traces on demand.

Whilst the aforementioned works deal with full state observability, action model learning has also been studied in domains where there is partial or missing state observability. ARMS works when no partial intermediate state is given. It defines a set of weighted constraints that must hold for the plans to be correct, and solves the weighted

propositional satisfiability problem with a MAX-SAT solver [15]. In order to efficiently solve the large MAX-SAT representations, ARMS implements a hill-climbing method that models the actions approximately. SLAF also deals with partial observability [16]. Given a formula representing the initial belief state, a sequence of executed actions and the corresponding partially observed states, it builds a complete explanation of observations by models of actions through a CNF formula. The learning algorithm updates the formula of the belief state with every action and observation in the sequence. This update makes sure that the new formula represents all the transition relations consistent with the actions and observations. The formula returned at the end includes all consistent models, which can then be retrieved with additional processing.

Unlike the previous approaches, the one described in [17] deals with both missing and noisy predicates in the observations. An action model is first learnt by constructing a set of kernel classifiers which tolerate noise and partial observability and then STRIPSrules are derived from the classifiers' parameters.

LOCM only requires the example plans as input without need for providing information about predicates or states [18]. This makes LOCM be most likely the learning approach that works with the least information possible. The lack of available information is addressed by LOCM by exploiting assumptions about the kind of domain model it has to generate. Particularly, it assumes a domain consists of a collection of objects (sorts) whose defined set of states can be captured by a parameterized Finite State Machine (FSM). The intuitive assumptions of LOCM, like the continuity of object transitions or the association of parameters between consecutive actions in the training sequence, yield a learning model heavily reliant on the kind of domain structure. The inability of LOCM to properly derive domain theories where the state of a sort is subject to different FSMs is later overcome by LOCM2 by forming separate FSMs, each containing a subset of the full transition set for the sort [19]. LOP (LOCM with Optimized Plans [20]), the last contribution of the LOCM family, addresses the problem of inducing static predicates. Because LOCM approaches induce similar models for domains with similar structures, they face problems at generating models for domains that are only distinguished by whether or not they contain static relations (e.g. *blocksworld* and *freecell*). In order to mitigate this drawback, LOP applies a post-processing step after the LOCM analysis which requires additional information about the plans, namely a set of optimal plans to be used in the learning phase.

3. Background

This section defines the planning model used on this work, Classical planning with conditional effects, and the output of the learning tasks addressed in the paper, a STRIPS action model.

3.1. Classical planning with conditional effects

We use F to denote the set of *fluents* (propositional variables) describing a state. A *literal* l is a valuation of a fluent $f \in F$, i.e. either $l = f$ or $l = \neg f$. A set of literals L represents a partial assignment of values to fluents (without loss of generality, we will assume that L does not assign conflicting values to any fluent). We use $\mathcal{L}(F)$ to denote the set of all literal sets on F , i.e. all partial assignments of values to fluents.

A *state* s is a full assignment of values to fluents, i.e. $|s| = |F|$, so the size of the state space is $2^{|F|}$. Explicitly including negative literals $\neg f$ in states simplifies subsequent definitions but often, we will abuse notation by defining a state s only in terms of the fluents that are true in s , as is common in STRIPS planning.

A *classical planning frame* is a tuple $\Phi = \langle F, A \rangle$, where F is a set of fluents and A is a set of actions. An action $a \in A$ is defined with *preconditions*, $\text{pre}(a) \subseteq \mathcal{L}(F)$, *positive effects*, $\text{eff}^+(a) \subseteq \mathcal{L}(F)$, and *negative effects* $\text{eff}^-(a) \subseteq \mathcal{L}(F)$. We say that an action $a \in A$ is *applicable* in a state s iff $\text{pre}(a) \subseteq s$. The result of applying a in s is the *successor state* denoted by $\theta(s, a) = \{s \setminus \text{eff}^-(a)\} \cup \text{eff}^+(a)$.

An action $a \in A$ with conditional effects is defined as a set of *preconditions* $\text{pre}(a) \in \mathcal{L}(F)$ and a set of *conditional effects* $\text{cond}(a)$. Each conditional effect $C \triangleright E \in \text{cond}(a)$ is composed of two sets of literals $C \in \mathcal{L}(F)$, the *condition*, and $E \in \mathcal{L}(F)$, the *effect*. An action $a \in A$ is *applicable* in a state s if and only if $\text{pre}(a) \subseteq s$, and the *triggered effects* resulting from the action application are the effects whose conditions hold in s :

$$\text{triggered}(s, a) = \bigcup_{C \triangleright E \in \text{cond}(a), C \subseteq s} E,$$

The result of applying action a in state s is the *successor* state $\theta(s, a) = \{s \setminus \text{eff}_c^-(s, a) \cup \text{eff}_c^+(s, a)\}$ where $\text{eff}_c^-(s, a) \subseteq \text{triggered}(s, a)$ and $\text{eff}_c^+(s, a) \subseteq \text{triggered}(s, a)$ are, respectively, the triggered *negative* and *positive* effects.

A *classical planning problem* is a tuple $P = \langle F, A, I, G \rangle$, where I is an initial state and $G \subseteq \mathcal{L}(F)$ is a goal condition. A *plan* for P is an action sequence $\pi = \langle a_1, \dots, a_n \rangle$ that induces the *state trajectory* $\langle s_0, s_1, \dots, s_n \rangle$ such that $s_0 = I$ and, for each $1 \leq i \leq n$, a_i is applicable in s_{i-1} and generates the successor state $s_i = \theta(s_{i-1}, a_i)$. The *plan length* is denoted with $|\pi| = n$. A plan π *solves* P iff $G \subseteq s_n$, i.e. if the goal condition is satisfied at the last state reached after following the application of the plan π in the initial state I .

3.2. STRIPS action schemas

This work addresses the learning of PDDL action schemas that follow the STRIPS requirement [21, 22]. Figure 1 shows the *stack* action schema, coded in PDDL, from a four-operator *blocksworld* [23].

```
(:action stack
:parameters (?v1 ?v2 - object)
:precondition (and (holding ?v1) (clear ?v2))
:effect (and (not (holding ?v1)) (not (clear ?v2)) (handempty) (clear ?v1) (on ?v1 ?v2))
)
```

Figure 1: STRIPS operator schema coding, in PDDL, the *stack* action from a four-operator *blocksworld*.

To formalize the output of the learning task, we assume that fluents F are instantiated from a set of *predicates* Ψ , as in PDDL. Each predicate $p \in \Psi$ has an argument list of arity $\text{ar}(p)$. Given a set of *objects* Ω , the set of fluents F is induced by assigning objects in Ω to the arguments of predicates in Ψ , i.e. $F = \{p(\omega) : p \in \Psi, \omega \in \Omega^{\text{ar}(p)}\}$ s.t. Ω^k is the k -th Cartesian power of Ω .

Let $\Omega_v = \{v_i\}_{i=1}^{\max_{a \in A} \text{ar}(a)}$ be a new set of objects ($\Omega \cap \Omega_v = \emptyset$), denoted as *variable names*, and that is bound by the maximum arity of an action in a given planning frame. For instance, in a three-block *blocksworld* $\Omega = \{\text{block}_1, \text{block}_2, \text{block}_3\}$ while $\Omega_v = \{v_1, v_2\}$ because the operators with the maximum arity, *stack* and *unstack*, have arity two. We define F_v , a new set of fluents s.t. $F \cap F_v = \emptyset$, that results from instantiating Ψ using only the objects in Ω_v , i.e. the variable names, and that defines the elements that can appear in an action schema. For the *blocksworld*, $F_v = \{\text{handempty}, \text{holding}(v_1), \text{holding}(v_2), \text{clear}(v_1), \text{clear}(v_2), \text{ontable}(v_1), \text{ontable}(v_2), \text{on}(v_1, v_1), \text{on}(v_1, v_2), \text{on}(v_2, v_1), \text{on}(v_2, v_2)\}$.

We assume also that actions $a \in A$ are instantiated from STRIPS operator schemas $\xi = \langle \text{head}(\xi), \text{pre}(\xi), \text{add}(\xi), \text{del}(\xi) \rangle$ where:

- $\text{head}(\xi) = \langle \text{name}(\xi), \text{pars}(\xi) \rangle$, is the operator *header* defined by its name and the corresponding *variable names*, $\text{pars}(\xi) = \{v_i\}_{i=1}^{\text{ar}(\xi)}$. The headers of a four-operator *blocksworld* are $\text{pickup}(v_1)$, $\text{putdown}(v_1)$, $\text{stack}(v_1, v_2)$ and $\text{unstack}(v_1, v_2)$.
- The preconditions $\text{pre}(\xi) \subseteq F_v$, the negative effects $\text{del}(\xi) \subseteq F_v$, and the positive effects $\text{add}(\xi) \subseteq F_v$ such that, $\text{del}(\xi) \subseteq \text{pre}(\xi)$, $\text{del}(\xi) \cap \text{add}(\xi) = \emptyset$ and $\text{pre}(\xi) \cap \text{add}(\xi) = \emptyset$.

Finally, given the set of predicates Ψ and the header of a STRIPS operator schema ξ , we define $F_v(\xi) \subseteq F_v$ as the subset of elements that can appear in the action schema ξ and that confine its space of possible action models. For instance, for the *stack* action schema $F_v(\text{stack}) = F_v$ while $F_v(\text{pickup}) = \{\text{handempty}, \text{holding}(v_1), \text{clear}(v_1), \text{ontable}(v_1), \text{on}(v_1, v_1)\}$ excludes the fluents from F_v that involve v_2 because the action header $\text{pickup}(v_1)$ contains the single parameter v_1 .

Therefore, given the set of predicates Ψ and the header of the operator schema ξ , then $2^{|F_v(\xi)|}$ is an upper-bound on the number of possible STRIPS action model for ξ . In practice this number is lower because the previous STRIPS constraints require that a negative effect must appear as a precondition and that cannot be a positive effect and also, that a positive effect cannot appear as a precondition.

4. Learning STRIPS action models

Learning STRIPS action models from fully available input knowledge, i.e. from plans where the *pre-* and *post-* states of every action in a plan are available, is straightforward. When any intermediate state is available, STRIPS operator schemes are derived lifting the literals that change between the pre and post-state of the corresponding action executions. Preconditions are derived lifting the minimal set of literals that appears in all the pre-states of the corresponding actions [24]. This section formalizes a set of more challenging action model learning tasks, where less input knowledge is available.

4.1. Learning from observations of plan executions

The first learning task corresponds to observing an agent acting in the world but watching only the states that result of its plan executions, the actual executed actions are unobserved. This learning task is formalized as $\Lambda = \langle \mathcal{M}, \mathcal{O}, \Psi, \rangle$:

- \mathcal{M} is the set of *empty* operator schemas, wherein each $\xi \in \mathcal{M}$ is only composed of $head(\xi)$. In some cases, we may not require to start learning from scratch that is, the operator schemas in \mathcal{M} may be not *empty* but partially specified operator schemas where some preconditions and some positive or negative effects are a priori known.
- $\mathcal{O} = \langle s_0, s_1, \dots, s_n \rangle$ is a sequence of *state observations* obtained observing the execution of an *unobserved* plan $\pi = \langle a_1, \dots, a_n \rangle$ such that, for each $1 \leq i \leq n$, a_i is applicable in s_{i-1} and generates the successor state $s_i = \theta(s_{i-1}, a_i)$. The observed states are not required to be full states, they can be *partial states*, in the sense that some of the observed fluents (either with positive or negative value) are missing (they are not known to be either positive or negative).
- Ψ is the set of predicates that define the abstract state space of a given planning domain. This set can be given as input or inferred from the state observations given in \mathcal{O} when these observations are full states.

A solution to the $\Lambda = \langle \mathcal{M}, \mathcal{O}, \Psi \rangle$ learning task is a set of operator schema \mathcal{M}' compliant with the headers in \mathcal{M} , the state observation sequence \mathcal{O} and the predicates Ψ . In this learning scenario, a solution must not only determine a possible STRIPS action model but also the plan π , that explains the given observations using the learned STRIPS model. Figure 2 shows a $\Lambda = \langle \mathcal{M}, \mathcal{O}, \Psi \rangle$ task for learning the *blocksworld* STRIPS action model from the five-state observations sequence that corresponds to inverting a 2-block tower.

Here we redefine the task of learning STRIPS action models from observations of plan executions to cover the learning scenario where the actions executed by the observed agent are known. There may be missing state observations in \mathcal{O} , because otherwise the learning task is trivial as previously explained. In this case, the learning task is formalized as $\Lambda = \langle \mathcal{M}, \mathcal{O}, \Psi, \pi \rangle$, where:

- The plan $\pi = \langle a_1, \dots, a_n \rangle$, is an action sequence that produces the sequence of state observations given in \mathcal{O} . Note that if the plan is diverse enough it can be used to infer the set of *empty* operator schemas \mathcal{M} .

Figure 3 shows an example of a *blocksworld* learning task $\Lambda = \langle \mathcal{M}, \mathcal{O}, \Psi, \pi \rangle$, that corresponds to observing the execution of an eight-action plan for inverting a four-block tower. Note that \mathcal{M}, Ψ are skipped since they are the same as for Figure 2.

The previous definitions formalize the learning STRIPS action models from observations of a single plan execution. These definitions are extensible to the more general case where learning from the execution of multiple plans:

- When learning from states observations, $\Lambda = \langle \mathcal{M}, \mathcal{O}, \Psi \rangle$ is still a valid formalization by simply considering that now $\mathcal{O} = \langle s_0^1, s_1^1, \dots, s_n^1, \dots, s_0^t, s_1^t, \dots, s_n^t, \dots, s_0^\tau, s_1^\tau, \dots, s_n^\tau \rangle$ are sequences of *state observations* obtained observing the execution of multiple *unobserved* plans $\pi^t = \langle a_1, \dots, a_n \rangle$, $1 \leq t \leq \tau$ such that, for each $1 \leq i \leq n$, a_i^t is applicable in s_{i-1}^t and generates the successor state $s_i^t = \theta(s_{i-1}^t, a_i^t)$.
- When learning from a set of observed plans, the task is defined as $\Lambda = \langle \mathcal{M}, \Psi, \mathcal{O}, \Pi \rangle$, where $\Pi = \{\pi_1, \dots, \pi_\tau\}$ is the given set of example plans producing the corresponding given set of state observations.

```

;;;;; Headers in  $\mathcal{M}$ 

(pickup v1) (putdown v1)
(stack v1 v2) (unstack v1 v2)

;;;;; Predicates  $\Psi$ 

(handempty) (holding ?o - object)
(clear ?o - object) (ontable ?o - object)
(on ?o1 - object ?o2 - object)

;;;;; Observations  $\mathcal{O}$ 

;;; observation #0
(clear B) (on B A) (ontable A) (handempty)

;;; observation #1
(holding B) (clear A) (ontable A)

;;; observation #2
(clear A) (ontable A) (clear B) (ontable B) (handempty)

;;; observation #3
(holding A) (clear B) (ontable B)

;;; observation #4
(clear A) (on blockA B) (ontable B) (handempty)

```

Figure 2: Example of a $\Lambda = \langle \mathcal{M}, \mathcal{O}, \Psi \rangle$ task for learning a STRIPS action model in the *blocksworld* from a sequence of five state observations.

```

;;;;; Observations  $\mathcal{O}$ 

;;; observation #0
(clear B) (on B A) (ontable A) (handempty)

;;; observation #4
(clear A) (on blockA B) (ontable B) (handempty)

;;; Plan  $\pi$ 

0: (unstack B A)
1: (putdown B)
2: (pickup A)
3: (stack A B)

```

Figure 3: Example of a $\Lambda = \langle \mathcal{M}, \mathcal{O}, \Psi, \pi \rangle$ task for learning a STRIPS action model in the *blocksworld* from a four-action plan and two state observations.

5. Learning STRIPS action models with classical planning

Our approach for addressing a Λ learning task is compiling it into a classical planning task P_Λ with conditional effects. A planning compilation is a suitable approach for addressing Λ because a solution must not only determine the STRIPS action model \mathcal{M}' but also, the *unobserved* plans that explain the inputs to the learning task. The intuition behind the compilation is that a solution to the resulting classical planning task is a sequence of actions that:

1. **Programs the action model \mathcal{M}' .** A solution plan starts with a *prefix* that, for each $\xi \in \mathcal{M}$, determines which fluents $f \in F_v(\xi)$ belong to its *pre*(ξ), *del*(ξ) and *add*(ξ) sets.
2. **Validates the action model \mathcal{M}' .** The solution plan continues with a *postfix* that reproduces the given input knowledge (the available observations of the plan executions) using the programmed action model \mathcal{M}' .

5.1. Learning from state observations

Given a learning task $\Lambda = \langle \mathcal{M}, \mathcal{O}, \Psi \rangle$ the compilation outputs a classical planning task $P_\Lambda = \langle F_\Lambda, A_\Lambda, I_\Lambda, G_\Lambda \rangle$:

- F_Λ contains:
 - The set of fluents F built instantiating the predicates Ψ with the objects Ω appearing in the input observations \mathcal{O} , i.e. `block1` and `block2` in Figure 2. Formally, $\Omega = \bigcup_{s \in \mathcal{O}} \text{obj}(s)$, where obj is a function that returns the objects that appear in a given state.
 - Fluents $\text{pre}_f(\xi)$, $\text{del}_f(\xi)$ and $\text{add}_f(\xi)$, for every $f \in F_v(\xi)$, that represent the programmed action model. If a fluent $\text{pre}_f(\xi)/\text{del}_f(\xi)/\text{add}_f(\xi)$ holds, it means that f is a precondition/negative/positive effect in the schema $\xi \in \mathcal{M}'$. For instance, the preconditions of the *stack* schema (Figure 1) are represented by the pair of fluents `pre_holding_stack_v1` and `pre_clear_stack_v2` set to *True*.
 - The fluents $\text{mode}_{\text{prog}}$ and mode_{val} indicating whether the operator schemas are programmed or validated and the fluents $\{\text{test}_i\}_{1 \leq i \leq |\mathcal{O}|}$, indicating the observation where the action model is validated.
- I_Λ contains the fluents from F that encode s_0 (the first observation) and $\text{mode}_{\text{prog}}$ set to true. Our compilation assumes that initially, operator schemas are programmed with every possible precondition (the most specific learning hypothesis), no negative effect and no positive effect. Therefore fluents $\text{pre}_f(\xi)$, for every $f \in F_v(\xi)$, hold also at the initial state.
- $G_\Lambda = \bigcup_{1 \leq i \leq |\mathcal{O}|} \{\text{test}_i\}$, requires that the programmed action model is validated in all the input observations.
- A_Λ comprises three kinds of actions:

1. Actions for *programming* operator schema $\xi \in \mathcal{M}$:

- Actions for **removing** a *precondition* $f \in F_v(\xi)$ from the action schema $\xi \in \mathcal{M}$.

$$\begin{aligned} \text{pre}(\text{programPre}_{t,\xi}) &= \{\neg \text{del}_f(\xi), \neg \text{add}_f(\xi), \\ &\quad \text{mode}_{\text{prog}}, \text{pre}_f(\xi)\}, \\ \text{cond}(\text{programPre}_{t,\xi}) &= \{\emptyset\} \triangleright \{\neg \text{pre}_f(\xi)\}. \end{aligned}$$

- Actions for **adding** a *negative* or *positive* effect $f \in F_v(\xi)$ to the action schema $\xi \in \mathcal{M}$.

$$\begin{aligned} \text{pre}(\text{programEff}_{t,\xi}) &= \{\neg \text{del}_f(\xi), \neg \text{add}_f(\xi), \\ &\quad \text{mode}_{\text{prog}}\}, \\ \text{cond}(\text{programEff}_{t,\xi}) &= \{\text{pre}_f(\xi)\} \triangleright \{\text{del}_f(\xi)\}, \\ &\quad \{\neg \text{pre}_f(\xi)\} \triangleright \{\text{add}_f(\xi)\}. \end{aligned}$$

2. Actions for *applying* a programmed operator schema $\xi \in \mathcal{M}$ bound with objects $\omega \subseteq \Omega^{ar(\xi)}$. Given that the operators headers are known, the variables $\text{pars}(\xi)$ are bound to the objects in ω that appear at the same position. Figure 4 shows the PDDL encoding of the action for applying a programmed operator *stack*.

$$\begin{aligned} \text{pre}(\text{apply}_{\xi,\omega}) &= \{\text{pre}_f(\xi) \implies p(\omega)\}_{\forall p \in \Psi, f = p(\text{pars}(\xi))} \\ &\quad \cup \{\neg \text{mode}_{\text{val}}\}, \\ \text{cond}(\text{apply}_{\xi,\omega}) &= \{\text{del}_f(\xi)\} \triangleright \{\neg p(\omega)\}_{\forall p \in \Psi, f = p(\text{pars}(\xi))}, \\ &\quad \{\text{add}_f(\xi)\} \triangleright \{p(\omega)\}_{\forall p \in \Psi, f = p(\text{pars}(\xi))}, \\ &\quad \{\text{mode}_{\text{prog}}\} \triangleright \{\neg \text{mode}_{\text{prog}}\}, \\ &\quad \{\emptyset\} \triangleright \{\text{mode}_{\text{val}}\}. \end{aligned}$$

```

(:action apply_stack
:parameters (?o1 - object ?o2 - object)
:precondition
  (and (or (not (pre_on_stack_v1_v1)) (on ?o1 ?o1))
        (or (not (pre_on_stack_v1_v2)) (on ?o1 ?o2))
        (or (not (pre_on_stack_v2_v1)) (on ?o2 ?o1))
        (or (not (pre_on_stack_v2_v2)) (on ?o2 ?o2))
        (or (not (pre_ontable_stack_v1)) (ontable ?o1))
        (or (not (pre_ontable_stack_v2)) (ontable ?o2))
        (or (not (pre_clear_stack_v1)) (clear ?o1))
        (or (not (pre_clear_stack_v2)) (clear ?o2))
        (or (not (pre_holding_stack_v1)) (holding ?o1))
        (or (not (pre_holding_stack_v2)) (holding ?o2))
        (or (not (pre_handempty_stack)) (handempty))))
:effect
  (and (when (del_on_stack_v1_v1) (not (on ?o1 ?o1)))
        (when (del_on_stack_v1_v2) (not (on ?o1 ?o2)))
        (when (del_on_stack_v2_v1) (not (on ?o2 ?o1)))
        (when (del_on_stack_v2_v2) (not (on ?o2 ?o2)))
        (when (del_ontable_stack_v1) (not (ontable ?o1)))
        (when (del_ontable_stack_v2) (not (ontable ?o2)))
        (when (del_clear_stack_v1) (not (clear ?o1)))
        (when (del_clear_stack_v2) (not (clear ?o2)))
        (when (del_holding_stack_v1) (not (holding ?o1)))
        (when (del_holding_stack_v2) (not (holding ?o2)))
        (when (del_handempty_stack) (not (handempty)))
        (when (add_on_stack_v1_v1) (on ?o1 ?o1))
        (when (add_on_stack_v1_v2) (on ?o1 ?o2))
        (when (add_on_stack_v2_v1) (on ?o2 ?o1))
        (when (add_on_stack_v2_v2) (on ?o2 ?o2))
        (when (add_ontable_stack_v1) (ontable ?o1))
        (when (add_ontable_stack_v2) (ontable ?o2))
        (when (add_clear_stack_v1) (clear ?o1))
        (when (add_clear_stack_v2) (clear ?o2))
        (when (add_holding_stack_v1) (holding ?o1))
        (when (add_holding_stack_v2) (holding ?o2))
        (when (add_handempty_stack) (handempty))
        (when (modeProg) (not (modeProg)))))

```

Figure 4: PDDL action for applying an already programmed schema *stack* (implications coded as disjunctions).

3. Actions for *validating* an observation $1 \leq i \leq |O|$.

$$\begin{aligned}
 \text{pre}(\text{validate}_i) &= s_i \cup \{test_j\}_{j \in 1 \leq j < i} \\
 &\quad \cup \{\neg test_j\}_{j \in i \leq j \leq |O|} \cup \{mode_{val}\}, \\
 \text{cond}(\text{validate}_i) &= \{\emptyset\} \triangleright \{test_i, \neg mode_{val}\}.
 \end{aligned}$$

Known preconditions and effects (that is, a partially specified STRIPS action model) can be encoded as fluents $pre_f(\xi)$, $del_f(\xi)$ and $add_f(\xi)$ set to true at the initial state I_Λ . In this case, the corresponding programming actions, $\text{programPre}_{f,\xi}$ and $\text{programEff}_{f,\xi}$, become unnecessary and are removed from A_Λ making the classical planning task P_Λ easier to be solved. In the extreme, when a fully specified STRIPS action model Ξ is given in Ξ_0 , the compilation validates whether an observed plan follows the given model:

- If a solution plan is found to P_Λ , it means that the given STRIPS action model is *valid* for the given examples.
- If P_Λ is unsolvable it means that the given STRIPS action model is invalid since it is not compliant with all the given examples.

To illustrate how our compilation works, Figure 5 shows a plan that solves a classical planning task resulting from the compilation. The plan programs and validates the *stack* schema using the five state observations shown in Figure 2 as well as previously specified operator schemes for *pickup*, *putdown* and *unstack*. Note that this compilation is valid for the particular scenario where the observed states are not full states but *partial states*, in the sense that some of the observed fluents (either with positive or negative value) are missing.

Even more, if any reference to the $mode_{val}$ fluent is removed from the compilation then, it is valid for learning when there are missing state observations in O . In that case, the compilation is more flexible but the resulting classical planning task P_Λ becomes harder to be solved since the classical planner must determine how many *apply* actions are necessary between two *validate* actions.

```

00 : (program_pre_clear_stack_v1)
01 : (program_pre_handempty_stack)
02 : (program_pre_holding_stack_v2)
03 : (program_pre_on_stack_v1_v1)
04 : (program_pre_on_stack_v1_v2)
05 : (program_pre_on_stack_v2_v1)
06 : (program_pre_on_stack_v2_v2)
07 : (program_pre_ontable_stack_v1)
08 : (program_pre_ontable_stack_v2)
09 : (program_eff_clear_stack_v1)
10 : (program_eff_clear_stack_v2)
11 : (program_eff_handempty_stack)
12 : (program_eff_holding_stack_v1)
13 : (program_eff_on_stack_v1_v2)
14 : (apply_unstack block2 block1)
15 : (validate.1)
16 : (apply_putdown block2)
17 : (validate.2)
18 : (apply_pickup block1)
19 : (validate.3)
20 : (apply_stack block1 block2)
21 : (validate.4)

```

Figure 5: Plan for programming and validating the *stack* schema using the five state observations shown in Figure 2 as well as previously specified operator schemes for *pickup*, *putdown* and *unstack*.

5.2. Learning from plans

We extend the compilation to consider observed actions. Given a learning task $\Lambda = \langle \mathcal{M}, O, \Psi, \pi \rangle$, the compilation outputs a classical planning task $P_{\Lambda'} = \langle F_{\Lambda'}, A_{\Lambda'}, I_{\Lambda'}, G_{\Lambda'} \rangle$ such that:

- $F_{\Lambda'}$ extends F_Λ with $F_\pi = \{plan(name(\xi), \Omega^{ar(\xi)}, j)\}_{1 \leq j \leq |\pi|}$, the fluents to code the j^{th} step of the observed plan $\pi = \langle a_1, \dots, a_n \rangle$ that contains the action $name(\xi), \Omega^{ar(\xi)}$. Fluents at_j and $next_{j,j+1}$, $1 \leq j < |\pi|$, are also added to represent the current plan step and iterate through the steps of the plan.
- $I_{\Lambda'}$ extends I_Λ with fluents F_π plus fluents at_1 and $\{next_{j,j+1}\}$, $1 \leq j < |\pi|$, for indicating the plan step where the action model is validated. Goals are as in the original compilation. In other words, observed states are used for validation adding a $test_t$, $1 \leq t \leq |O|$ fluent for each observation in O .
- With respect to $A_{\Lambda'}$.
 1. The actions for *programming* the preconditions/effects of a given operator schema $\xi \in \Xi$ and the actions for *validating* the programmed action model in the given state observations are the same.
 2. The actions for *applying* an already programmed operator have an extra precondition $f \in F_\pi$ that encodes the current plan step, and extra conditional effects $\{at_j\} \triangleright \{\neg at_j, at_{j+1}\}_{j \in [1,n]}$ for advancing to the next plan step. With this mechanism we ensure that these actions are applied in the same order as in the example plan π .

To illustrate this, the classical plan of Figure 6 is a solution to a learning task $\Lambda = \langle \mathcal{M}, \mathcal{O}, \Psi, \pi \rangle$ for getting the *blocksworld* action model where operator schemes for *pickup*, *putdown* and *unstack* are specified in \mathcal{M} . This plan programs and validates the operator schema *stack* from *blocksworld*, using the plan π and the two state observations \mathcal{O} shown in Figure 3. Plan steps [0, 8] program the preconditions of the *stack* operator, steps [9, 13] program the operator effects and steps [14, 18] validate the programmed operators following the plan π shown in the Figure 3.

```

00 : (program_pre_clear_stack_v1)
01 : (program_pre_handempty_stack)
02 : (program_pre_holding_stack_v2)
03 : (program_pre_on_stack_v1_v1)
04 : (program_pre_on_stack_v1_v2)
05 : (program_pre_on_stack_v2_v1)
06 : (program_pre_on_stack_v2_v2)
07 : (program_pre_ontable_stack_v1)
08 : (program_pre_ontable_stack_v2)
09 : (program_eff_clear_stack_v1)
10 : (program_eff_clear_stack_v2)
11 : (program_eff_handempty_stack)
12 : (program_eff_holding_stack_v1)
13 : (program_eff_on_stack_v1_v2)
14 : (apply_unstack b a i1 i2)
15 : (apply_putdown b i2 i3)
16 : (apply_pickup a i3 i4)
17 : (apply_stack a b i4 i5)
18 : (validate_1)

```

Figure 6: Plan for programming and validating the *stack* schema using plan π and state observations \mathcal{O} (shown in Figure 3) as well as previously specified operator schemes for *pickup*, *putdown* and *unstack*.

The compilation is extensible for the learning scenario where there are missing actions in π . The extension does not code the corresponding $plan(name(\xi), \Omega^{ar(\xi)}, j$ fluent and does not add the extra precondition $f \in F_\pi$ to the corresponding corresponding *apply* action. In that case the resulting classical planning task P_Λ becomes harder to be solved since the classical planner must determine the *apply* action necessary for the validating the learned action model.

Now we explain how to address the learning from multiple plans $\Pi = \{\pi_1, \dots, \pi_\tau\}$, $1 \leq t \leq \tau$. Let us first define a set of classical planning instances $P_t = \langle F, \emptyset, I_t, G_t \rangle$ that belong to the same planning frame (i.e. same fluents and actions but different initial states and goals). The set of actions, $A = \emptyset$, is empty because the action model is initially unknown. Finally, the initial state I_t is given by the state s_0^t , and the goals G_t are defined by the state s_n^t . Addressing the learning task $\Lambda = \langle \mathcal{M}, \mathcal{O}, \Psi, \Pi \rangle$ with classical planning requires introducing a small modification to our compilation. In particular, the actions in P_Λ for *validating* the plan $\pi_t \in \Pi$, $1 \leq t \leq \tau$, are now defined as:

$$\begin{aligned}
\text{pre}(\text{validate}_t) &= G_t \cup \{\text{test}_j\}_{1 \leq j < t} \cup \{\neg \text{test}_j\}_{t \leq j \leq \tau} \cup \{\neg \text{mode}_{prog}\}, \\
\text{cond}(\text{validate}_t) &= \{\emptyset\} \triangleright \{\text{test}_i\} \cup \{\neg f\}_{\forall f \in G_t, f \notin I_{t+1}} \cup \{f\}_{\forall f \in I_{t+1}, f \notin G_t}.
\end{aligned}$$

5.3. Compilation properties

Lemma 1. Soundness. Any classical plan π that solves P_Λ induces an action model \mathcal{M}' that solves $\Lambda = \langle \mathcal{M}, \mathcal{O}, \Psi, \Pi \rangle$.

Proof sketch. Once operator schemas \mathcal{M}' are programmed, they can only be applied and validated, because of the mode_{prog} fluent. In addition, P_Λ is only solvable if fluents $\{\text{test}_i\}$, $1 \leq i \leq n$ hold at the last reached state. These goals can only be achieved executing an applicable sequence of programmed operator schemas that reaches every state $s_i \in \mathcal{O}$, starting from s_0 and following the sequence $1 \leq i \leq n$. This means that the programmed action model \mathcal{M}' complies with the provided input knowledge and hence, solves Λ . \square

Lemma 2. Completeness. Any STRIPS action model \mathcal{M}' that solves a $\Lambda = \langle \mathcal{M}, \Psi, \mathcal{O}, \Pi \rangle$ learning task, is computable solving the corresponding classical planning task P_Λ .

Proof sketch. By definition, $F_v(\xi) \subseteq F_\Lambda$ fully captures the full set of elements that can appear in a STRIPS action schema $\xi \in \mathcal{M}$ given its header and the set of predicates Ψ . The compilation does not discard any possible STRIPS action schema definable within F_v that satisfies the state trajectory constraint given by \mathcal{O}, Π . \square

The size of the classical planning task P_Λ output by the compilation depends on:

- The arity of the actions headers in \mathcal{M} and the predicates Ψ that are given as input to the Λ learning task. The larger these numbers, the larger the $F_v(\xi)$ sets, that define the $pre_f(\xi)/del_f(\xi)/add_f(\xi)$ fluents set and the corresponding set of *programming* actions.
- The number of given state observations. The larger $|\mathcal{O}|$, the more *test_i* fluents and *validate_i* actions in P_Λ .

5.4. Exploiting static predicates to optimize the compilation

A *static predicate* $p \in \Psi$ is a predicate that does not appear in the effects of any action [25]. Therefore, one can get rid of the mechanism for programming these predicates in the effects of any action schema while keeping the compilation complete. Given a static predicate p :

- Fluents $del_f(\xi)$ and $add_f(\xi)$, such that $f \in F_v$ is an instantiation of the static predicate p in the set of *variable objects* Ω_v , can be discarded for every $\xi \in \Xi$.
- Actions $programEff_{f,\xi}$ (s.t. $f \in F_v$ is an instantiation of p in Ω_v) can also be discarded for every $\xi \in \Xi$.

Static predicates can also constrain the space of possible preconditions by looking at the given set of state observation \mathcal{O} . One can assume that if a precondition $f \in F_v$ (s.t. $f \in F_v$ is an instantiation of a static predicate in Ω_v) is not compliant with the observations in \mathcal{O} then, fluents $pre_f(\xi)$ and actions $programPre_{f,\xi}$ can be discarded for every $\xi \in \mathcal{M}$. For instance in the *zenotravel* domain *pre_next_board_v1_v1*, *pre_next_debark_v1_v1*, *pre_next_fly_v1_v1*, *pre_next_zoom_v1_v1*, *pre_next_refuel_v1_v1* can be discarded (and their corresponding programming actions) because a precondition (*next ?v1 ?v1 - flevel*) will never hold at any state in \mathcal{O} .

Looking at the given example plans, fluents $pre_f(\xi)$ and actions $programPre_{f,\xi}$ are also discardable for every $\xi \in \Xi$ if a precondition $f \in F_v$ (s.t. $f \in F_v$ is an instantiation of a static predicate in Ω_v) is not possible according to Π and \mathcal{O} . Back to the *zenotravel* domain, if an example plan $\pi_i \in \Pi$ contains the action (*fly plane1 city2 city0 f13 f12*) and the corresponding state observations contain the static literal (*next f12 f13*) but does not contain (*next f12 f12*), (*next f13 f13*) or (*next f13 f12*) the only possible precondition including the static predicate is *pre_next_fly_v5_v4*.

6. Evaluating STRIPS action models

We assess how well a STRIPS action model \mathcal{M} explains given observations of plan executions according to the amount of *edition* required by \mathcal{M} to induce that observations.

6.1. The STRIPS edit distance

We first define the allowed *operations* to edit a given STRIPS action model. With the aim of keeping a tractable branching factor of the planning instances that result from our compilations, we only define two *edit operations*:

- *Deletion.* A fluent $pre_f(\xi)/del_f(\xi)/add_f(\xi)$ is removed from the operator schema $\xi \in \mathcal{M}$, such that $f \in F_v(\xi)$.
- *Insertion.* A fluent $pre_f(\xi)/del_f(\xi)/add_f(\xi)$ is added to the operator schema $\xi \in \mathcal{M}$, s.t. $f \in F_v(\xi)$.

We can now formalize an *edit distance* that quantifies how similar two given STRIPS action models are. The distance is symmetric and meets the *metric axioms* provided that the two *edit operations*, deletion and insertion, have the same positive cost.

Definition 3. Let \mathcal{M} and \mathcal{M}' be two STRIPS action models, such that both models are built from the same set of possible elements F_v . The **edit distance**, denoted as $\delta(\mathcal{M}, \mathcal{M}')$, is the minimum number of edit operations that is required to transform \mathcal{M} into \mathcal{M}' .

Since F_v is a bound set, the maximum number of edits that can be introduced to a given action model defined within F_v is bound as well. In more detail, for an operator schema $\xi \in \mathcal{M}$ the maximum number of edits that can be introduced to their precondition set is $|F_v(\xi)|$ while the max number of edits that can be introduced to the effects is twice $|F_v(\xi)|$.

Definition 4. The **maximum edit distance** of an STRIPS action model \mathcal{M} built from the set of possible elements F_v is $\delta(\mathcal{M}, *) = \sum_{\xi \in \mathcal{M}} 3|F_v(\xi)|$.

We define now an edit distance to assess the quality of a learned action model with respect to a sequence of state observations.

Definition 5. Given \mathcal{M} , a STRIPS action models built from F_v , and the observations sequence $O = \langle s_0, s_1, \dots, s_n \rangle$ such that each observation in O is built with fluents in F . The **observation edit distance**, denoted by $\delta(\mathcal{M}, O)$, is the minimal edit distance from \mathcal{M} to any model \mathcal{M}' , defined also within F_v , such that \mathcal{M}' can produce a valid plan $\pi = \langle a_1, \dots, a_n \rangle$ that induces O ;

$$\delta(\mathcal{M}, O) = \min_{\forall \mathcal{M}' \rightarrow O} \delta(\mathcal{M}, \mathcal{M}')$$

Unlike the error function defined by ARMS [15], the *observation edit distance* assess, with a single expression, the flaws in the preconditions and effects of a given learned model. This fact enables the recognition of STRIPS action models. The idea, taken from *plan recognition as planning* [26], is to map distances into likelihoods. These *edit distance* could be mapped into a likelihood with the following expression $P(O|\mathcal{M}) = 1 - \frac{\delta(\mathcal{M}, O)}{\delta(\mathcal{M}, *)}$.

This edit distance could also be defined quantifying the amount of edition necessary to modify the observations of the plan executions to match the given STRIPS action model. This would imply *edit operations* that modify the observations instead of modifying the action schemes. Our definition of the edit distance is however more practical since usually F_v is much smaller than F . Finally, the edit distance can also be defined with respect to a set of plans.

Definition 6. Given an action model \mathcal{M} , built from F_v and a set of valid plans $\Pi = \{\pi_1, \dots, \pi_\tau\}$ that only contain actions built grounding the schemes in \mathcal{M} . The **plans edit distance**, denoted by $\delta(\mathcal{M}, \Pi)$, is the minimal edit distance from \mathcal{M} to any model \mathcal{M}' , defined also within F_v , such that \mathcal{M}' can produce all the plans $\pi_i \in \Pi$, $1 \leq i \leq \tau$;

$$\delta(\mathcal{M}, \Pi) = \min_{\forall \mathcal{M}' \rightarrow \Pi} \delta(\mathcal{M}, \mathcal{M}')$$

An Λ learning task can swap the roles of two operators whose headers match or two action parameters that belong to the same type (e.g. the *blocksworld* operator *stack* can be *learned* with the preconditions and effects of the *unstack* operator and vice versa, or the parameters of the *stack* operator can be swapped). Both the *observation edit distance* and the *plans edit distance* are semantic measures robust to role changes of this particular kind.

6.2. Computing the observations and plans edit distance

Our compilation is extensible to compute the *observation edit distance* by simply considering that the model \mathcal{M} , given in $\Lambda = \langle \mathcal{M}, O, \Psi \rangle$, is *non-empty*. In other words, now \mathcal{M} is a set of given operator schemas, wherein each $\xi \in \mathcal{M}$ initially contains *head*(ξ) but also the *pre*(ξ), *del*(ξ) and *add*(ξ) sets. A solution to the planning task resulting from the extended compilation is a sequence of actions that:

1. **Edits the action model \mathcal{M} .** A solution plan starts with a *prefix* that modifies the preconditions and effects of the action schemas in \mathcal{M} using to the two *edit operations*, deletion and insertion.
2. **Validates the edited model \mathcal{M}' in observations of the plan executions.** The solution plan continues with a postfix that validates the edited model on the given observations O (or in Π), as explained in Section 5.

Now Λ is not a learning task but the task of editing \mathcal{M} to produce the observations O , which results in the edited model \mathcal{M}' . The output of the extended compilation is a classical planning task $P'_\Lambda = \langle F_\Lambda, A'_\Lambda, I'_\Lambda, G_\Lambda \rangle$:

- F_Λ and G_Λ are defined as in the previous compilation.
- I'_Λ contains the fluents from F that encode s_0 and $mode_{prog}$ set to true. In addition, the input action model \mathcal{M} is now encoded in the initial state. This means that the fluents $pre_f(\xi)/del_f(\xi)/add_f(\xi)$, $f \in F_v(\xi)$, hold in the initial state iff they appear in \mathcal{M} .
- A'_Λ , comprises the same three kinds of actions of A_Λ . The actions for *applying* an already programmed operator schema and the actions for *validating* an observation are defined exactly as in the previous compilation. The only difference here is that the actions for *programming* the operator schema now implement the two *edit operations* (i.e. include actions for *inserting* a precondition and for *deleting* a negative/positive effect).

To illustrate this, the plan of Figure 7 solves the classical planning task P'_Λ that corresponds to editing a *blocksworld* action model where the positive effects (`handempty`) and (`clear ?v1`) of the `stack` schema are missing. The plan edits first the `stack` schema, *inserting* these two positive effects, and then validates the edited action model in the five-observation sequence of Figure 2.

```

00 : (insert_add_handempty_stack)
01 : (insert_add_clear_stack_var1)
02 : (apply_unstack block2 block1)
03 : (validate_1)
04 : (apply_putdown block2)
05 : (validate_2)
06 : (apply_pickup block1)
07 : (validate_3)
08 : (apply_stack block1 block2)
09 : (validate_4)

```

Figure 7: Plan for editing a given *blocksworld* schema and validating it at the state observations shown in Figure 2.

Our interest when computing the *observation edit distance* is not in the resulting action model \mathcal{M}' but in the number of required *edit operations* for that \mathcal{M}' is validated in the given observations, e.g. $\delta(\mathcal{M}, \mathcal{O}) = 2$ for the example in Figure 7. In this case $\delta(\mathcal{M}, *) = 4 \times 3 \times 11$ since there are 4 action schemes (`pickup`, `putdown`, `stack` and `unstack`) and the size of F_v is 11 (as shown in Section 3). The *observation edit distance* is exactly computed if P'_Λ is optimally solved (according to the number of edit actions); is approximated if P'_Λ is solved with a satisfying planner (our case); and is a less accurate estimate (but faster to be computed) if the solved task is a relaxation of P'_Λ [27].

If a set of plans Π is available, the compilation can be adapted to compute the *plan edit distance* $\delta(\mathcal{M}, \Pi)$. The modifications to the compilation explained in Section 5 are useful here to redefine the learning task as the task of editing \mathcal{M} to produce the set of plans Π , which results in the edited model \mathcal{M}' . Figure 8 shows the plan for editing a given *blocksworld* action model where again the positive effects (`handempty`) and (`clear ?v1`) of the `stack` schema are missing. In this case the edited action model is however validated it at the plan shown in Figure 3.

```

00 : (insert_add_handempty_stack)
01 : (insert_add_clear_stack_var1)
02 : (apply_unstack b a i1 i2)
03 : (apply_putdown b i2 i3)
04 : (apply_pickup a i3 i4)
05 : (apply_stack a b i4 i5)
06 : (validate_1)

```

Figure 8: Plan for editing a given *blocksworld* schema and validating it at the plan shown in Figure 3.

This compilation is flexible to compute the *edit distance* between two STRIPS action models, \mathcal{M} and \mathcal{M}' , built from the same set of possible elements F_v . A solution to the planning task resulting from this compilation is a sequence

of actions that edits the action model \mathcal{M} to produce \mathcal{M}' using to the two *edit operations*, deletion and insertion. F_Λ, I_Λ are defined like in the previous compilation. With respect to the actions, A_Λ again implement the two *edit operations* (i.e. include actions for *inserting* a precondition and for *deleting* a negative/positive effect) but does not contain apply actions because the STRIPS action model are not validated in any observation of plan executions. Finally, the goals are also different and are now defined by the set of fluents, $pre_f(\xi)/del_f(\xi)/add_f(\xi)$ that represent all the operator schema $\xi \in \mathcal{M}'$, such that $f \in F_v(\xi)$.

To illustrate this, the plan of Figure 9 solves the classical planning task that corresponds to computing the distance between a *blocksworld* action model, where the positive effects (`handempty`) and (`clear ?v1`) of the stack schema are missing, and the actual four-operator *blocksworld* model. The plan edits first the stack schema, *inserting* these two positive effects. Again our interest is in the number of required *edit operations*, e.g. $\delta(\mathcal{M}, \mathcal{M}') = 2$.

```
00 : (insert_add_handempty_stack)
01 : (insert_add_clear_stack_var1)
```

Figure 9: Plan for computing the distance between a *blocksworld* action model, where the positive effects (`handempty`) and (`clear ?v1`) of the stack schema are missing, and the actual four-operator *blocksworld* model.

7. Recognition of STRIPS action models

Given a set of possible STRIPS models and set of observations of plan executions, the *recognition of STRIPS models* is the task of computing the model with the highest probability according to the cited observations. According to the Bayes rule, the probability of an hypothesis \mathcal{H} given the observations \mathcal{O} can be computed with $P(\mathcal{H}|\mathcal{O}) = \frac{P(\mathcal{O}|\mathcal{H})P(\mathcal{H})}{P(\mathcal{O})}$. In our scenario, the hypotheses are about the possible STRIPS action models that can be built within a given set of predicates Ψ and a given a set of operator headers (in other words, given the $F_v(\xi)$ sets). With this regard, $P(\mathcal{M}|\mathcal{O})$, the probability distribution of the possible STRIPS models (within the $F_v(\xi)$ sets) given an observation sequence \mathcal{O} could be computed by:

1. Computing the *observation edit distance* $\delta(\mathcal{M}, \mathcal{O})$ for every possible model \mathcal{M} .
2. Applying the resulting distances to the above $P(\mathcal{O}|\mathcal{M})$ formula to map these distances into likelihoods
3. Applying the Bayes rule to obtain the normalized posterior probabilities, these probabilities must sum 1.

If a set of plans Π is available, this same strategy can be followed using the *plan edit distance* $\delta(\mathcal{M}, \Pi)$.

8. Evaluation

This section evaluates the performance of our approach for learning STRIPS action models starting from different amounts of available input knowledge.

8.1. Setup

The domains used in the evaluation are IPC domains that satisfy the STRIPS requirement [22], taken from the PLANNING.DOMAINS repository [28]. We only use 5 learning examples for each domain and they are fixed for all the experiments so we can evaluate the impact of the input knowledge in the quality of the learned models. All experiments are run on an Intel Core i5 3.10 GHz x 4 with 4 GB of RAM.

- **Reproducibility.** We make fully available the compilation source code, the evaluation scripts and the used benchmarks at this anonymous repository <https://github.com/anonsub/strips-learning> so any experimental data reported in the paper is fully reproducible.
- **Planner.** The classical planner we use to solve the instances that result from our compilations is MADAGASCAR [29]. We use MADAGASCAR because its ability to deal with planning instances populated with dead-ends. In addition, SAT-based planners can apply the actions for programming preconditions in a single planning step (in parallel) because these actions do not interact. Actions for programming action effects can also be applied in a single planning step reducing significantly the planning horizon.

- **Metrics.** The quality of the learned models is quantified with the *precision* and *recall* metrics. These two metrics are frequently used in *pattern recognition*, *information retrieval* and *binary classification* and are more informative than simply counting the number of errors in the learned model or computing the *symmetric difference* between the learned and the reference model [30]. Intuitively, precision gives a notion of *soundness* while recall gives a notion of the *completeness* of the learned models. Formally, $Precision = \frac{tp}{tp+fp}$, where tp is the number of true positives (predicates that correctly appear in the action model) and fp is the number of false positives (predicates appear in the learned action model that should not appear). Recall is formally defined as $Recall = \frac{tp}{tp+fn}$ where fn is the number of false negatives (predicates that should appear in the learned action model but are missing).

8.1.1. Evaluating with a test set

When a reference model is not available, the learned models are tested with an observation set. Table 1 summarizes the results obtained when evaluating the quality of the learned models with respect to a test set of state observations. Each test set comprises between 20 and 50 observations per domain and is generated executing the plans for various instances of the IPC domains and collecting the intermediate states.

The table shows, for each domain, the *observation edit distance* (computed with our extended compilation), the *maximum edit distance*, and their ratio. The reported results show that, despite learning only from 25 state observations, 12 out of 15 learned domains yield ratios of 90% or above. This fact evidences that the learned models require very small amounts of edition to match the observations of the given test set.

	$\delta(\mathcal{M}, \mathcal{O})$	$\delta(\mathcal{M}, *)$	$1 - \frac{\delta(\mathcal{M}, \mathcal{O})}{\delta(\mathcal{M}, *)}$
blocks	0	90	1.0
driverlog	5	144	0.97
ferry	2	69	0.97
floortile	34	342	0.90
grid	42	153	0.73
gripper	2	30	0.93
hanoi	1	63	0.98
hiking	69	174	0.60
miconic	3	72	0.96
npuzzle	2	24	0.92
parking	4	111	0.96
satellite	24	75	0.68
transport	4	78	0.95
visitall	2	24	0.92
zenotravel	3	63	0.95

Table 1: Evaluation of the quality of the learned models with respect to an observations test set.

8.1.2. Evaluating with a reference model

Here we evaluate the learned models with respect to the actual generative model. Opposite to what usually happens in ML, this model is available when learning is applied to IPC domains. The model learned for each domain is compared with its reference model using:

- $Precision = \frac{tp}{tp+fp}$, where tp is the number of *true positives* (predicates that correctly appear in the action model) and fp is the number of *false positives* (predicates of the learned model that should not appear). Precision gives a notion of *soundness*.
- $Recall = \frac{tp}{tp+fn}$, where fn is the number of *false negatives* (predicates that should appear in the learned model but are missing). Recall gives a notion of *completeness*.

Table 2 shows the precision (**P**) and recall (**R**) computed separately for the preconditions (**Pre**), positive effects (**Add**) and negative Effects (**Del**) while the last two columns report averages values. The reason why the scores in Table 2 are lower than in Table 1 is because the syntax-based nature of *precision* and *recall* make these two metrics

report low scores for learned models that are semantically correct but correspond to *reformulations* of the actual model (changes in the roles of actions with matching headers or parameters with matching types).

8.1.3. Precision and recall robust to model reformulations

To give an insight of the actual quality of the learned models, we defined a method for computing *Precision* and *Recall* that is robust to the mentioned model reformulations.

Precision and recall are often combined using the *harmonic mean*. This expression, called the *F-measure* or the balanced *F-score*, is defined as $F = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$. Given the learned action model \mathcal{M} and the reference action model \mathcal{M}^* , the bijective function $f_{P\&R} : \mathcal{M} \mapsto \mathcal{M}^*$ is the mapping between the learned and the reference model that maximizes the accumulated *F-measure* (considering swaps in the actions with matching headers or parameters with matching types). Table 3 shows that significantly higher values of *precision* and *recall* are reported when a learned action schema, $\xi \in \mathcal{M}$, is compared to its corresponding reference schema given by the $f_{P\&R}$ mapping ($f_{P\&R}(\xi) \in \mathcal{M}^*$). These results evidence that in all of the evaluated domains, except for *ferry* and *satellite*, the learning task swaps the roles of some actions (or parameters) with respect to their role in the reference model.

As we can see in Table 3, the *blocksworld* and *grripper* domains are perfectly learned from only 25 state observations. On the other hand, the learning scores of several domains in Table 3 are still lower than in Table 1. The reason lies in the particular observations comprised by the test sets. As an example, in the *driverlog* domain, the action schema `disembark-truck` is missing from the learned model because this action is never induced from the observations in the training set; that is, such action never appears in the corresponding *unobserved* plan. The same happens with the `paint-down` action of the *floortile* domain or `move-curb-to-curb` in the *parking* domain. Interestingly, these actions do not appear either in the test sets and so the learned action models are not penalized in Table 1. Generating *informative* and *representative* observations for learning planning action models is an open issue. Planning actions include preconditions that are only satisfied by specific sequences of actions, often, with a low probability of being chosen by chance [31].

8.2. Learning from plans

We start evaluating our approach with tasks $\Lambda' = \langle \Psi, \Sigma, \Pi \rangle$, where *labeled plans* are available. We then repeat the evaluation but exploiting potential *static predicates* computed from Σ , which are the predicates that appear unaltered in the initial and final states in every $\sigma_t \in \Sigma$. Static predicates are used to constrain the space of possible action models as explained in the previous section.

Table 4 shows the obtained results. Precision (**P**) and recall (**R**) are computed separately for the preconditions (**Pre**), positive effects (**Add**) and negative Effects (**Del**), while the last two columns of each setting and the last row report averages values. We can observe that identifying static predicates leads to models with better precondition *recall*. This fact evidences that many of the missing preconditions corresponded to static predicates because there is no incentive to learn them as they always hold [20].

Table 5 reports the total planning time, the preprocessing time (in seconds) invested by MADAGASCAR to solve the planning instances that result from our compilation as well as the number of actions of the solution plans. All the learning tasks are solved in a few seconds. Interestingly, one can identify the domains with static predicates by just looking at the reported plan length. In these domains some of the preconditions that correspond to static predicates are directly derived from the learning examples and therefore fewer programming actions are required. When static predicates are identified, the resulting compilation is also much more compact and produces smaller planning/instantiation times.

8.3. Learning from partially specified action models

We evaluate now the ability of our approach to support partially specified action models; that is, addressing learning tasks of the kind $\Lambda'' = \langle \Psi, \Sigma, \Pi, \Xi_0 \rangle$. In this experiment, the model of half of the actions is given in Ξ_0 as an extra input of the learning task.

Tables 6 and 7 summarize the obtained results, which include the identification of static predicates. We only report the *precision* and *recall* of the *unknown* actions since the values of the metrics of the *known* action models is 1.0. In this experiment, a low value of *precision* or *recall* has a greater impact than in the corresponding Λ' tasks because the

	Pre		Add		Del		P	R
	P	R	P	R	P	R		
blocks	0.44	0.44	0.44	0.44	0.44	0.44	0.44	0.44
driverlog	0.0	0.0	0.25	0.43	0.0	0.0	0.08	0.14
ferry	1.0	0.71	1.0	1.0	1.0	1.0	1.0	0.9
floortile	0.38	0.55	0.4	0.18	0.56	0.45	0.44	0.39
grid	0.5	0.47	0.33	0.29	0.25	0.29	0.36	0.35
gripper	0.83	0.83	0.75	0.75	0.75	0.75	0.78	0.78
hanoi	0.5	0.25	0.5	0.5	0.0	0.0	0.33	0.25
hiking	0.43	0.43	0.5	0.35	0.44	0.47	0.46	0.42
miconic	0.6	0.33	0.33	0.25	0.33	0.33	0.42	0.31
npuzzle	0.33	0.33	0.0	0.0	0.0	0.0	0.11	0.11
parking	0.25	0.21	0.0	0.0	0.0	0.0	0.08	0.07
satellite	0.6	0.21	0.8	0.8	1.0	0.5	0.8	0.5
transport	1.0	0.3	0.8	0.8	1.0	0.6	0.93	0.57
visitall	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
zenotravel	0.67	0.29	0.33	0.29	0.33	0.14	0.44	0.24

Table 2: Precision and recall values obtained without computing the $f_{P\&R}$ mapping with the reference model.

	Pre		Add		Del		P	R
	P	R	P	R	P	R		
blocks	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
driverlog	0.67	0.14	0.33	0.57	0.67	0.29	0.56	0.33
ferry	1.0	0.71	1.0	1.0	1.0	1.0	1.0	0.9
floortile	0.44	0.64	1.0	0.45	0.89	0.73	0.78	0.61
grid	0.63	0.59	0.67	0.57	0.63	0.71	0.64	0.62
gripper	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
hanoi	1.0	0.5	1.0	1.0	1.0	1.0	1.0	0.83
hiking	0.78	0.6	0.93	0.82	0.88	0.88	0.87	0.77
miconic	0.8	0.44	1.0	0.75	1.0	1.0	0.93	0.73
npuzzle	0.67	0.67	1.0	1.0	1.0	1.0	0.89	0.89
parking	0.56	0.36	0.5	0.33	0.5	0.33	0.52	0.34
satellite	0.6	0.21	0.8	0.8	1.0	0.5	0.8	0.5
transport	1.0	0.3	1.0	1.0	1.0	0.6	1.0	0.63
visitall	0.67	1.0	1.0	1.0	1.0	1.0	0.89	1.0
zenotravel	1.0	0.43	0.67	0.57	1.0	0.43	0.89	0.48

Table 3: Precision and recall values obtained when computing the $f_{P\&R}$ mapping with the reference model.

	No Static								Static							
	Pre		Add		Del		P	R	Pre		Add		Del		P	R
	P	R	P	R	P	R			P	R	P	R	P	R		
Blocks	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Driverlog	1.0	0.36	0.75	0.86	1.0	0.71	0.92	0.64	0.9	0.64	0.56	0.71	0.86	0.86	0.78	0.73
Ferry	1.0	0.57	1.0	1.0	1.0	1.0	1.0	0.86	1.0	0.57	1.0	1.0	1.0	1.0	1.0	0.86
Floortile	0.52	0.68	0.64	0.82	0.83	0.91	0.66	0.80	0.68	0.68	0.89	0.73	1.0	0.82	0.86	0.74
Grid	0.62	0.47	0.75	0.86	0.78	1.0	0.71	0.78	0.79	0.65	1.0	0.86	0.88	1.0	0.89	0.83
Gripper	1.0	0.67	1.0	1.0	1.0	1.0	1.0	0.89	1.0	0.67	1.0	1.0	1.0	1.0	1.0	0.89
Hanoi	1.0	0.50	1.0	1.0	1.0	1.0	1.0	0.83	0.75	0.75	1.0	1.0	1.0	1.0	0.92	0.92
Miconic	0.75	0.33	0.50	0.50	0.75	1.0	0.67	0.61	0.89	0.89	1.0	0.75	0.75	1.0	0.88	0.88
Satellite	0.60	0.21	1.0	1.0	1.0	0.75	0.87	0.65	0.82	0.64	1.0	1.0	1.0	0.75	0.94	0.80
Transport	1.0	0.40	1.0	1.0	1.0	0.80	1.0	0.73	1.0	0.70	0.83	1.0	1.0	0.80	0.94	0.83
Visitall	1.0	0.50	1.0	1.0	1.0	1.0	1.0	0.83	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Zenotravel	1.0	0.36	1.0	1.0	1.0	0.71	1.0	0.69	1.0	0.64	0.88	1.0	1.0	0.71	0.96	0.79
	0.88	0.50	0.88	0.92	0.95	0.91	0.90	0.78	0.90	0.74	0.93	0.92	0.96	0.91	0.93	0.86

Table 4: Precision and recall scores for learning tasks from labeled plans without (left) and with (right) static predicates.

evaluation is done only over half of the actions. This occurs, for instance, in the precondition *recall* of domains such as *Floortile*, *Gripper* or *Satellite*.

Remarkably, the overall *precision* is now 0.98, which means that the contents of the learned models is highly reliable. The value of *recall*, 0.87, is an indication that the learned models still miss some information (preconditions

	No Static			Static		
	Total	Preprocess	Length	Total	Preprocess	Length
Blocks	0.04	0.00	72	0.03	0.00	72
Driverlog	0.14	0.09	83	0.06	0.03	59
Ferry	0.06	0.03	55	0.06	0.03	55
Floortile	2.42	1.64	168	0.67	0.57	77
Grid	4.82	4.75	88	3.39	3.35	72
Gripper	0.03	0.01	43	0.01	0.00	43
Hanoi	0.12	0.06	48	0.09	0.06	39
Miconic	0.06	0.03	57	0.04	0.00	41
Satellite	0.20	0.14	67	0.18	0.12	60
Transport	0.59	0.53	61	0.39	0.35	48
Visitall	0.21	0.15	40	0.17	0.15	36
Zenotravel	2.07	2.04	71	1.01	1.00	55

Table 5: Total planning time, preprocessing time and plan length for learning tasks from labeled plans without/with static predicates.

are again the component more difficult to be fully learned). Overall, the results confirm the previous trend: the more input knowledge of the task, the better the models and the less planning time. Additionally, the solution plans required for this task are smaller because it is only necessary to program half of the actions (the other half are included in the input knowledge Ξ_0). *Visitall* and *Hanoi* are excluded from this evaluation because they only contain one action schema.

	Pre		Add		Del		P	R
	P	R	P	R	P	R		
Blocks	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Driverlog	1.0	0.71	1.0	1.0	1.0	1.0	1.0	0.90
Ferry	1.0	0.67	1.0	1.0	1.0	1.0	1.0	0.89
Floortile	0.75	0.60	1.0	0.80	1.0	0.80	0.92	0.73
Grid	1.0	0.67	1.0	1.0	1.0	1.0	0.84	0.78
Gripper	1.0	0.50	1.0	1.0	1.0	1.0	1.0	0.83
Miconic	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Satellite	1.0	0.57	1.0	1.0	1.0	1.0	1.0	0.86
Transport	1.0	0.75	1.0	1.0	1.0	1.0	1.0	0.92
Zenotravel	1.0	0.67	1.0	1.0	1.0	0.67	1.0	0.78
	0.98	0.71	1.0	0.98	1.0	0.95	0.98	0.87

Table 6: Precision and recall scores for learning tasks with partially specified action models.

9. Conclusions

Acknowledgment

This work is supported by the Spanish MINECO project TIN2017-88476-C2-1-R. Diego Aineto is partially supported by the *FPU16/03184* and Sergio Jiménez by the *RYC15/18009*, both programs funded by the Spanish government.

References

- [1] M. Ghallab, D. Nau, P. Traverso, *Automated Planning: theory and practice*, Elsevier, 2004.
- [2] M. Ramírez, *Plan recognition as planning*, Ph.D. thesis, Universitat Pompeu Fabra (2012).
- [3] H. Geffner, B. Bonet, *A concise introduction to models and methods for automated planning* (2013).
- [4] S. Kambhampati, *Model-lite planning for the web age masses: The challenges of planning with incomplete and evolving domain models*, in: National Conference on Artificial Intelligence, AAAI-07, 2007.
- [5] R. S. Michalski, J. G. Carbonell, T. M. Mitchell, *Machine learning: An artificial intelligence approach*, Springer Science & Business Media, 2013.
- [6] R. E. Fikes, N. J. Nilsson, *Strips: A new approach to the application of theorem proving to problem solving*, Artificial Intelligence 2 (3-4) (1971) 189–208.
- [7] B. Bonet, H. Palacios, H. Geffner, *Automatic derivation of memoryless policies and finite-state controllers using classical planners*, in: International Conference on Automated Planning and Scheduling (ICAPS), 2009.

	Total time	Preprocess	Plan length
Blocks	0.07	0.01	54
Driverlog	0.03	0.01	40
Ferry	0.06	0.03	45
Floortile	0.43	0.42	55
Grid	3.12	3.07	53
Gripper	0.03	0.01	35
Miconic	0.03	0.01	34
Satellite	0.14	0.14	47
Transport	0.23	0.21	37
Zenotravel	0.90	0.89	40

Table 7: Time and plan length learning for learning tasks with partially specified action models.

- [8] J. Segovia, S. Jiménez, A. Jonsson, Generalized planning with procedural domain control knowledge, 2016.
- [9] J. Segovia-Aguas, S. Jiménez, A. Jonsson, Hierarchical finite state controllers for generalized planning, in: International Joint Conference on Artificial Intelligence, IJCAI-16, AAAI Press, 2016, pp. 3235–3241.
- [10] J. Segovia-Aguas, S. Jiménez, A. Jonsson, Generating context-free grammars using classical planning, in: International Joint Conference on Artificial Intelligence, ICAPS-17, 2017.
- [11] D. Aineto, S. Jiménez, E. Onaindia, Learning strips action models with classical planning.
- [12] W. Shen, H. A. Simon, Rule creation and rule learning through environmental exploration, in: International Joint Conference on Artificial Intelligence, IJCAI-89, 1989, pp. 675–680.
- [13] X. Wang, Learning by observation and practice: An incremental approach for planning operator acquisition, in: International Conference on Machine Learning, 1995, pp. 549–557.
- [14] T. J. Walsh, M. L. Littman, Efficient learning of action schemas and web-service descriptions, in: National Conference on Artificial Intelligence, 2008, pp. 714–719.
- [15] Q. Yang, K. Wu, Y. Jiang, Learning action models from plan examples using weighted max-sat, Artificial Intelligence 171 (2-3) (2007) 107–143.
- [16] E. Amir, A. Chang, Learning partially observable deterministic action models, Journal of Artificial Intelligence Research 33 (2008) 349–402.
- [17] K. Mourão, L. S. Zettlemoyer, R. P. A. Petrick, M. Steedman, Learning STRIPS operators from noisy and incomplete observations, in: Conference on Uncertainty in Artificial Intelligence (UAI), 2012, pp. 614–623.
- [18] S. N. Cresswell, T. L. McCluskey, M. M. West, Acquiring planning domain models using LOCM, The Knowledge Engineering Review 28 (02) (2013) 195–213.
- [19] S. Cresswell, P. Gregory, Generalised domain model acquisition from action traces, in: International Conference on Automated Planning and Scheduling, ICAPS-11, 2011.
- [20] P. Gregory, S. Cresswell, Domain model acquisition in the presence of static relations in the LOP system., in: International Conference on Automated Planning and Scheduling, ICAPS-15, 2015, pp. 97–105.
- [21] D. McDermott, M. Ghallab, A. Howe, C. Knoblock, A. Ram, M. Veloso, D. Weld, D. Wilkins, PDDL – The Planning Domain Definition Language (1998).
- [22] M. Fox, D. Long, PDDL2.1: An extension to PDDL for expressing temporal planning domains., Journal of Artificial Intelligence Research 20 (2003) 61–124.
- [23] J. Slaney, S. Thiébaux, Blocks world revisited, Artificial Intelligence 125 (1-2) (2001) 119–153.
- [24] S. Jiménez, T. De La Rosa, S. Fernández, F. Fernández, D. Borrajo, A review of machine learning for automated planning, The Knowledge Engineering Review 27 (04) (2012) 433–467.
- [25] M. Fox, D. Long, The automatic inference of state invariants in TIM, Journal of Artificial Intelligence Research 9 (1998) 367–421.
- [26] M. Ramírez, H. Geffner, Plan recognition as planning, in: International Joint conference on Artificial Intelligence, 2009, pp. 1778–1783.
- [27] B. Bonet, H. Geffner, Planning as heuristic search, Artificial Intelligence 129 (1-2) (2001) 5–33.
- [28] C. Muise, Planning. domains, ICAPS system demonstration.
- [29] J. Rintanen, Madagascar: Scalable planning with sat, Proceedings of the 8th International Planning Competition (IPC-2014).
- [30] J. Davis, M. Goadrich, The relationship between precision-recall and ROC curves, in: International Conference on Machine learning, ACM, 2006, pp. 233–240.
- [31] A. Fern, S. W. Yoon, R. Givan, Learning domain-specific control knowledge from random walks., in: International Conference on Automated Planning and Scheduling, ICAPS-04, 2004, pp. 191–199.