

Discovering State and Action Abstractions for Generalized Task and Motion Planning

Aidan Curtis, Tom Silver, Joshua B. Tenenbaum, Tomás Lozano-Pérez, Leslie Kaelbling

MIT Computer Science and Artificial Intelligence Laboratory
 {curtisa, tslvr, jbt, tlp, lpk}@mit.edu

Abstract

Generalized planning accelerates classical planning by finding an algorithm-like policy that solves multiple instances of a task. A generalized plan can be learned from a few training examples and applied to an entire domain of problems. Generalized planning approaches perform well in discrete AI planning problems that involve large numbers of objects and extended action sequences to achieve the goal. In this paper, we propose an algorithm for learning features, abstractions, and generalized plans for continuous robotic task and motion planning (TAMP) and examine the unique difficulties that arise when forced to consider geometric and physical constraints as a part of the generalized plan. Additionally, we show that these simple generalized plans learned from only a handful of examples can be used to improve the search efficiency of TAMP solvers.

Introduction

A shared goal in the robotics and artificial intelligence communities is to build intelligent autonomous robotic agents that can solve long-horizon tasks in arbitrary human environments. Two of the major roadblocks to building such general-purpose intelligent systems are *scalability* to settings with large numbers of objects and *generalizability* to a wide array of novel environments and goals.

A scalable and generalizable intelligent robotic agent must consider objects' geometric, physical, visual, and semantic attributes to select and execute goal-directed actions. For example, a task such as “pack all my camping gear into my backpack” might require reasoning about the location, size and shape, mass and deformability, ownership, and category designation of each item in the environment. The agent must also execute actions that progress to the specified goal without violating physical and geometric constraints induced by the environment.

One typical approach to solving this type of problem is task and motion planning (TAMP) (Garrett et al. 2020). TAMP solvers work by integrating discrete symbolic planning with geometric motion planning where the results from each of these two subsystems can affect the other's feasibility. Because of this mutual dependence between the symbolic and

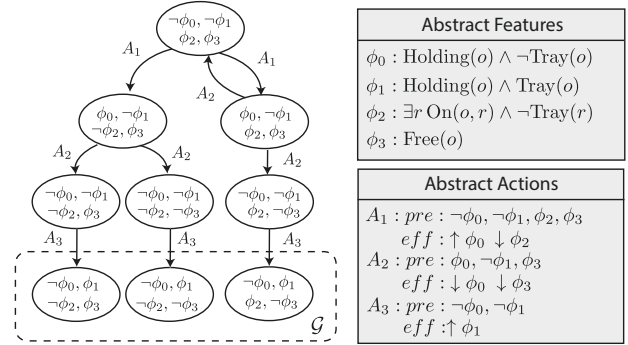


Figure 1: An abstract policy for the working LOAD task example. The generalized plan that solves this LOAD task picks an object (A_1) and places it on the tray (A_2) until either the tray is full ($\neg\phi_3$) or no objects are remaining on the table ($\neg\phi_2$). After one of these two features evaluates to false, the tray is grasped (A_3). Here $\text{Free}(o)$ is used as a shorthand for the larger feature $\exists p, r \forall o_2, p_2 : \text{Tray}(r) \wedge \neg \text{On}(o, p, r) \wedge \text{On}(o_2, p_2, r) \Rightarrow \neg \text{CFree}(o, p, o_2, p_2, r)$

geometric constraints, TAMP solvers usually rely on iterative procedures that alternate between geometric and discrete symbolic planning until a geometrically feasible plan is found that reaches the goal. The resulting iterative planning algorithms are much more computationally complex than classical planning algorithms. While TAMP solvers are fully generalizable to novel environments and tasks, they are limited in scalability due to the large combinatorial and continuous search problem that exists for each problem.

Generalized planning is a method for tackling exactly this scalability problem in classical planning. Generalized planners find abstract high-level plans that apply to a large, often infinite, class of problem instances with a shared structure. For example, “While there are objects in the box, pick up a reachable object in the box and place it on the table.” is a generalized plan that applies to settings with any number of objects. Solving each problem instance independently with standard planning approaches can be highly inefficient as planners are often slow and become exponentially slower with increasing solution length. The motivation of generalized planning is to avoid search in large state-space problem

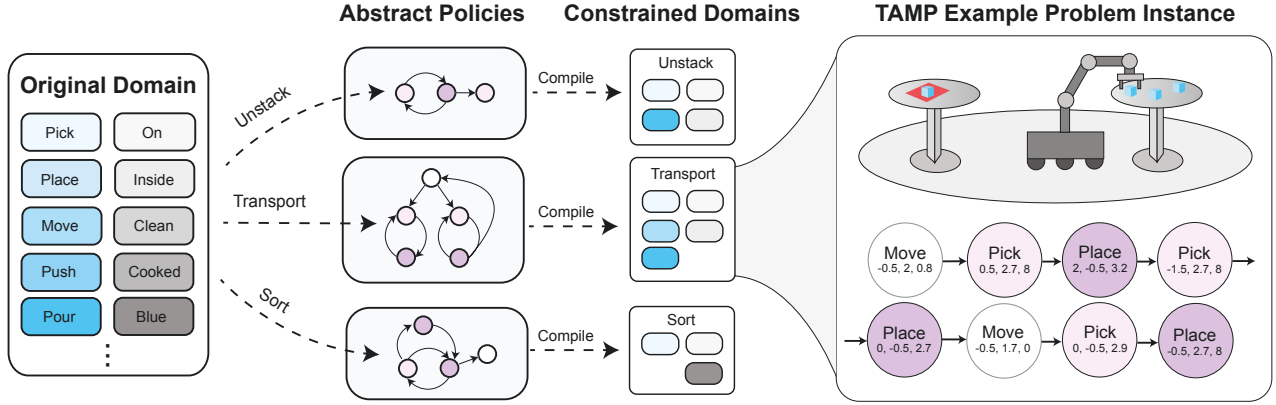


Figure 2: GENTAMP conceptual overview. Several robotics tasks are defined within a single domain. For each task, a policy is learned from a small set of example plans and compiled into a set of action and ordering constraints added to the original domain to create a task-specific domain. This task-specific domain forces TAMP search to follow the abstract policy, thus reducing the search space leading to faster planning.

instances by identifying and exploiting structure in a few example problems with smaller state spaces. In addition, because languages for expressing generalized plans often include looping constructs, the resulting generalized plans are often more compact than traditional plans for large domains. While discrete generalized planning methods help in problems with large numbers of objects, they do not incorporate continuous features. As such, they would not be capable of discovering generalized plans that rely on continuous features of the environment, such as the remaining capacity of a backpack, the water level of a camping thermos, or more subtle collision constraints present in any robotic setting. While there are various approaches to generalized planning for classical AI planning domains, our method is the first to apply these approaches to continuous robotics domains in which the generalized plan requires complex continuous features.

We propose GENTAMP, a framework for learning and executing generalized plans in mixed discrete-continuous environments. The novel contributions of this paper are as follows. First, we formulate the problem of generalized task and motion planning. Second, we outline a new algorithm for learning a generalized plan from examples that requires evaluation of complex continuous features with high-arity predicates and nested quantification. Third, we introduce an approach for executing generalized plans on unseen task and motion planning problems. Lastly, we evaluate our algorithm’s ability to scale to large unseen problems and compare planning speeds with a standard TAMP solver.

Background

Classical AI planning A classical planning problem is typically defined by the tuple $P = \langle S, I, A, G \rangle$, where S is a set of state variables that can take a finite set of values, I is the initial state, an initial assignment of values to each of the state variables, A is a set of actions, and G is

a subset of S describing the goal state. A classical planning domain is defined by a deterministic successor state function $f : S \times A \rightarrow S$. The objective of classical planning is to find a sequence of actions and their resulting states $(a_1, s_1), (a_2, s_2), \dots, (a_n, s_n)$ such that $a_i \in A$, $s_{i+1} = f(s_i, a_{i+1})$, and $s_n \in G$.

Such classical planning problems typically have an associated *lifted* task representation, which allows the domain description to abstract away from the particular set of entities in any problem instance. This description is parameterized by an entity set \mathcal{O} , and consists of a number of N-ary predicates $p(\bar{o}) : \mathcal{O}^N \rightarrow \{0, 1\}$ and a set of action schemas, $A_i(\bar{o})$ with entity-parameterized positive and negative predicate preconditions (pre^+, pre^-) and effects (eff^+, eff^-) such that action $A_i(\bar{o})$ is applicable in state s if its positive preconditions are satisfied in s and its negative ones are not: $pre^+(A_i(\bar{o})) \subseteq s \wedge pre^-(A_i(\bar{o})) \cap s = \emptyset$. The resulting effect of applying the entity-parameterized action schema on a state s removes the negative effects and adds the positive ones: $f(s, A_i(\bar{o})) = (s \setminus eff^-(A_i(\bar{o}))) \cup eff^+(A_i(\bar{o}))$. Given a particular set of entities \mathcal{O} for a problem instance, before planning, this lifted representation is usually compiled down to the standard state-variable representation above in a process called *grounding*, in which each state variable corresponds to a *ground literal*, which is an application of a predicate to a tuple of entities.

Generalized planning. The generalized planning problem builds on this formulation and establishes a meta-problem $Q = \langle P_1, P_2, \dots \rangle$ encompassing a (possibly infinite) set of classical planning problems. The objective of generalized planning is to find a deterministic policy $\pi : S \rightarrow A$ that, when successively applied starting from an initial state, results in a feasible action sequence that eventually reaches the goal for as many $P_i \in Q$ as possible. We model our problem formulation on the one established by Bonet and Geffner (2018), in which problem instances have no shared

state or action representation $P_i = \langle \mathcal{S}_i, I_i, A_i, G_i \rangle$.

Task and motion planning. The generalized planning methods in this paper build on the ability to solve TAMP problems; we use the PDDLSTREAM (Garrett, Lozano-Pérez, and Kaelbling 2018) problem definition language and associated algorithms as the basis for generalized planning in our implementations, but the overall approach is compatible with other TAMP methods. This TAMP problem definition amends the entity set to include *sampled entities* such as grasps, collision-free pose trajectories, and object placement samples. These entities are sampled from high-dimensional nonlinear continuous spaces by *streams*. Streams are continuous samplers with both procedural and declarative components. The procedural component is a function $g(\bar{o})$ that takes in a tuple of objects and generates sampled entities from some continuous space conditioned on the inputs. The declarative component specifies the semantics of these sampled entities and how they relate to existing entities in the domain through *certified predicates* which have, as arguments, the entities sampled by the stream and are only referenced in initial states, goals, and the preconditions of actions. Algorithms that use the PDDLSTREAM problem definition typically start with an insufficient set of sampled entities and certified predicates. During planning, they sample additional entities, adding corresponding certified predicates until the goal is achievable by a classical planner.

Generalized Task and Motion Planning

In this work, we propose the generalized TAMP problem setting. Typical robotics applications that rely on TAMP algorithms have a single *domain* that details the preconditions and effects of every action available to the agent and every possible property of objects and their relationships but does not specify the universe of objects, a concrete initial state, or a goal. Generalized TAMP additionally specifies a set of *tasks* $\{Q_1, \dots, Q_N\}$ where each Q is associated with a tuple $\langle I_Q, G_Q \rangle$. I_Q is a set of possible initial states in Q (e.g. “all red objects are on the table”), and G_Q is a set of possible goals for Q (e.g. “all red objects are on the red mat”). Because these are infinite sets, they are described by first-order logical expressions that quantify over objects in the state and can therefore be applied to states containing arbitrary numbers of entities. Infinitely many *problems* can be derived from any task with each problem instance P_i containing a unique ground initial state $I_i \in I_Q$ and goal $G_i \in G_Q$. The exact logical description of these tasks is not known a priori but is conveyed through task-specific demonstrations in the form of example plans.

A generalized plan for TAMP can be understood as an abstract “program” for solving the problem, which specifies the order and behavior of high-level operations but does not directly specify all of the continuous parameters. Unlike generalized symbolic plans, we cannot execute this program step-by-step, choosing parameters “greedily” as we go, because of additional constraints imposed by the geometry and physics of the environment. For example, greedily sampling and executing placements of objects in a packing task may lead to situations where objects placed earlier in a plan obstruct objects’ placements later in the plan. For this reason,

the objective of a generalized task and motion planner is to construct a new constrained domain for each task that, when solved using a TAMP solver, finds a ground plan most efficiently in terms of wall-clock computation time for any problem instance conforming to the task specification. (Figure 2)

This formulation is notably different from that of generalized planning in a few important ways. First, consistent with TAMP robotics applications, all tasks are specified under the same lifted domain containing the same action sets, streams, and properties describing the world. Second, goals and initial states are now defined in a logical language and involve continuous geometric and physical parameters (e.g., a pose is within a region.) Lastly, because direct execution of a generalized TAMP plan is impossible, we instead evaluate search speed across several testing problem instances.

Working example. To build intuition for generalized TAMP, we describe a simple working example of a generalized TAMP problem in terms of the specification described above. First, we establish an example overarching robotics domain containing a “free-flying” robotic gripper agent that can pick and place movable objects.

This domain has predicates $\text{Holding}(o)$, $\text{On}(o_1, p, o_2)$, $\text{Movable}(o)$, $\text{Tray}(o)$, and HandEmpty . The domain also has certified predicates $\text{Pose}(o, p, s)$ for possible object placement and $\text{CFree}(o_1, o_2, p_1, p_2)$ indicating that two objects at certain poses on the same surface are not in collision. Additionally, the domain will have the following action schemas:

```
Pick( $o, p, s$ )
  pre: HandEmpty, On( $o, p, s$ ), Pose( $o, p, s$ )
  eff: Holding( $o$ ),  $\neg\text{On}(o, p, s)$ ,  $\neg\text{HandEmpty}$ 

Place( $o, p, s$ )
  pre: Holding( $o$ ), Pose( $o, p, s$ ),
        $\forall o_2, p_2 : \text{CFree}(o, p, o_2, p_2)$ 
  eff:  $\neg\text{Holding}(o)$ , On( $o, p, s$ ), HandEmpty
```

```
Pick( $o, p, s$ )
  pre: HandEmpty, On( $o, p, s$ ), Pose( $o, p, s$ )
  eff: Holding( $o$ ),  $\neg\text{On}(o, p, s)$ ,  $\neg\text{HandEmpty}$ 

Place( $o, p, s$ )
  pre: Holding( $o$ ), Pose( $o, p, s$ ),
        $\forall o_2, p_2 : \text{CFree}(o, p, o_2, p_2)$ 
  eff:  $\neg\text{Holding}(o)$ , On( $o, p, s$ ), HandEmpty
```

Lastly, the domain has two streams that generate sampled entities and certified predicates. The first stream $\text{SamplePlacement}(o, s)$ takes in a surface and an object, samples a pose entity p , and adds $\text{Pose}(o, p, r)$ to the initial state. The second stream $\text{CheckCollision}(o_1, p_1, o_2, p_2)$ tests for collisions and adds the certified predicate $\text{CFree}(o_1, p_1, o_2, p_2)$ if none were detected.

We now define the LOAD task, one of many possible tasks within this domain. The goal of this task (G_{LOAD}) is for all objects to be on the tray or for the tray to be full, and the initial state (I_{LOAD}) is all objects on the table. The LOAD task consists of problem instances $\{P_N, \dots\}$ where N refers to the number of objects. The initial entity set for any problem within this task consists of a robotic gripper agent r , a set

Algorithm 1: GENTAMP Training

Input: $P = \{P_1, P_2, \dots, P_{N-1}\} \subseteq Q$ **Output:** π_ϕ

```

 $\mathcal{D} = \{\}, \text{pool} \leftarrow \{\}$ 
for  $P_i \in P$  do
   $\mathcal{D} \leftarrow \mathcal{D} \cup \text{TAMP}(P_i)$ 
end
for  $c \in \{1 \dots \text{max\_complexity}\}$  do
   $\text{pool} \leftarrow \text{PRUNE}(\text{pool} \cup \text{GENERATE}(\text{pool}))$ 
end
 $\mathcal{T} \leftarrow \text{ENCODESAT}(\mathcal{D}, \text{pool})$ 
 $F, \bar{A} \leftarrow \text{SATSOLVE}(\mathcal{T})$ 
 $\bar{I}, \bar{G} \leftarrow F(P)$ 
 $\pi_\phi \leftarrow \text{FONDPLAN}(F, \bar{A}, \bar{I}, \bar{G})$ 
return  $\pi_\phi, F$ 

```

of N objects b_1, \dots, b_N and their corresponding initial poses bp_1, \dots, bp_N , a tray t with an initial pose tp , and a table src . The literals in the initial state of a problem instance are HandEmpty , $\text{Tray}(t)$, $\text{On}(t, tp, src)$, $\text{Pose}(t, tp, src)$, $\text{On}(b_i, bp_i, src)$, $\text{Pose}(b_i, bp_i, src)$, and $\text{Movable}(b_i)$ for all b_i . An optimal policy for this simple task is to repeatedly pick up an object and place it on the tray until the tray is full and then pick up the tray. This task is a subtask of the TRANSPORT task, which is described and evaluated in the Experiments section.

GENTAMP Training

In this section, we describe our proposed method for learning a deterministic policy from very few small TAMP example plans. In the first step, we use example plans to extract a pool of unique features of states in the domain. Next, we combine these features with a set of constraints to form a SAT theory which, when solved, yields a subset of features from the feature pool that comprise an *approximately valid abstraction*. We then use these selected features and an initial state to construct a fully observable nondeterministic (FOND) planning problem. The solution to the FOND problem is a policy that deterministically returns an abstract action when queried with an evaluation of the selected features. Details of the policy learning pipeline are in Algorithm 1. This strategy for generalized planning was introduced by Bonet, Francès, and Geffner (2018). In the following sections we describe a novel version of this framework that applies to problem domains with both continuous state and action spaces.

State and action abstractions. To define a policy with inputs conforming to a common representation across problem instances within a task, we establish a transformation of the problem-instance state space. This can be done using a set of Boolean features $\phi_b : \mathcal{S}_i \rightarrow \{0, 1\}$ for any instance state space \mathcal{S}_i and numeric features $\phi_n : \mathcal{S}_i \rightarrow \mathbb{N}_0$ and their corresponding qualitative projections $\phi_{\bar{n}} : \mathcal{S}_i \rightarrow \{0, >0\}$ for any instance state space \mathcal{S}_i . These features are defined as relations over sets of entities and can therefore be applied to states with varying numbers of entities and proper-

ties. For example, the numeric feature $\phi_n(s) \equiv |\{x \in \mathcal{O}_s : \text{Blue}(x)\}|$ takes in a state s and returns the number of blue objects in that state. Such a feature can be applied to any state regardless of the number of objects in that state.

F denotes the set of Boolean and qualitative numeric features. In general, these features are insufficient to reconstruct the state, and so yield an abstract state representation $\phi(s)$, which is simply an evaluation of each $\phi_b, \phi_{\bar{n}}$ for a state from any of the problem instances in the generalized planning problem. This abstract transformation can be applied to the initial and goal states of a problem instance to obtain an abstract initial state \bar{I} and goal state \bar{G} . Abstract actions \bar{A} are defined by preconditions and effects that are Boolean features or qualitative evaluations of numeric features. Abstract actions are easily derived from the features $\phi_b, \phi_{\bar{n}}$. Specifically, the effects of a derived abstract action are the qualitative effects over the features, and the preconditions of a derived action are the features that agree qualitatively in the states where that qualitative effect was observed. These abstract actions are no longer deterministic since actions that decrease a numeric feature can have two possible effects on the qualitative abstract state: reducing the feature value to 0 or leaving it greater than 0.

An abstraction for a task is the tuple $\langle F, \bar{A}, \bar{I}, \bar{G} \rangle$. Given an abstraction, an abstract policy $\pi_\phi(s)$ can be found using generate-and-test methods (Srivastava et al. 2011b) or, more efficiently, using a non-deterministic planner (Muisse, McIlraith, and Beck 2012). This abstract policy can be applied to novel problem instances containing an arbitrary number of objects.

Our objective is to learn ϕ_b and ϕ_n , and the requisite abstract actions, for a task Q from data representing transitions sampled from problem instances of the task Q . The learned feature sets and resulting abstract actions should satisfy three critical properties: soundness, completeness, and goal-distinguishability (see appendix for definitions).

These three properties of the abstraction are proven to be *valid*, in the sense that they are sufficient to guarantee that a policy in the space of abstract actions and states will always be refinable to a concrete plan in any problem instance of the domain (Bonet et al. 2019).

Collecting training data. Generating valid abstractions requires an exhaustive search through the state space to verify soundness, completeness, and goal distinguishability on every transition. However, due to the continuous, high-dimensional, and nonlinear nature of TAMP problems, it is impossible to enumerate every reachable state and transition in the environment. For this reason, we focus on constructing *approximately valid* abstractions from a subset of possible transitions in the environment.

Our approach uses a TAMP solver to find example plans for each training problem instance and uses the transitions from those plans to create approximately valid abstractions. The constituent states of the plan consist of all symbolic predicates and entities along with any sampled entities and certified predicates that are in the preimage of the plan.

Generating a feature pool. To select a set of features that constitute an approximately valid abstraction, we generate a pool of complexity-bounded candidate features and select a

subset of those features that jointly yield a valid abstraction. To accommodate high-arity certified predicates with sampled entities, we design a generative grammar that enables the use of features containing implications and mixed existential and universal quantifiers.

A TAMP domain comes with a set of primitives predicates $p(\bar{o})$. We additionally establish a set of named variables $\{x_1, \dots, x_M\}$ where M is the maximum predicate arity. These predicates and named variables are used as terminals in forming a set of concepts $C(\bar{z})$ through primitive negation, conjunction, and a single implication. Concepts are combined to form quantified concepts $C_q(x, \bar{z}_1, \bar{z}_2)$ where x is a free variable, \bar{z}_1 is a tuple of bound variables, and \bar{z}_2 is a tuple of variables that have yet to be bound. Formula containing only a single free variable are formed by quantifying all but one of the free variables using chained universal and existential quantifiers. Lastly, features are formed from the quantified concepts that contain only a single free variable. The resulting features are functions that take in a state and return a natural number that specifies the number of entities that can be plugged into the remaining free variable such that the feature's concepts holds. The specifics of this generative grammar are in Table 1.

Each feature in this grammar is accompanied with a *feature complexity* score that is defined as the maximum of the number of generative grammar rules applied to create the feature and the number of arguments of the feature. This grammar is implemented in a bottom-up fashion by combining primitives, concepts, and quantified concepts to create a full feature and discarding features when they pass a fixed complexity limit (5 in our experiments).

Arguments
$z_m \rightarrow x_1 \mid x_2 \mid \dots \mid x_M$
Concepts
$C_r(\bar{z}) \rightarrow p(\bar{z})$
$C_r(\bar{z}) \rightarrow \neg p(\bar{z})$
$C_r(\bar{z}_1 \circ \bar{z}_2) \rightarrow C_{r_1}(\bar{z}_1) \wedge C_{r_2}(\bar{z}_2)$
$C \rightarrow C_r(\bar{z})$
$C \rightarrow C_{r_1}(\bar{z}_1) \Rightarrow C_{r_2}(\bar{z}_2)$
Quantified Concepts
$C_q(x, \emptyset, \bar{z}) \rightarrow C(\{x\} \circ \bar{z})$
$C_q(x, \bar{q} \circ \{z_1\}, \bar{z}_{2:N}) \rightarrow \exists z_1. C_q(x, \bar{q}, \bar{z}_{1:N})$
$C_q(x, \bar{q} \circ \{z_1\}, \bar{z}_{2:N}) \rightarrow \forall z_1. C_q(x, \bar{q}, \bar{z}_{1:N})$
Features
$\phi_n \rightarrow f : s \mapsto \{x \in \mathcal{O}_s : C_q(x, \bar{z}, \emptyset)\} $

Table 1: Generative grammar for features that take in a state and generate per-entity first-order logical expressions.

This grammar specification has two important properties necessary for TAMP domains. First, we use arbitrary-arity features instead of binary features to accommodate the high-arity predicates present in most TAMP domains. To handle

these higher arity predicates, our grammar includes alternating quantifiers evaluated in Prolog (Clocksin and Melish 2003). Second, we incorporate certified predicates and sampled entities into the grammar through the terminals (p) so that the resulting features are mixtures of geometric and symbolic properties. These changes greatly increase the total number of features and thus require more aggressive pruning at each generation step. We prune concepts and quantified concepts in each step by *uniqueness* and *equivalence*. Uniqueness pruning looks at each collected transition and checks to ensure no lower-complexity feature has the same value on each transition. Equivalence pruning uses standard logical inference rules to test equivalence between logical formulas.

Listed below are some example quantified concepts from our working LOAD task, including quantified concepts composed from both symbolic and geometric primitives from the TAMP specification.

- $f_1(o) \equiv \text{Holding}(o) \wedge \neg \text{Tray}(o)$
- $f_2(o) \equiv \text{Holding}(o) \wedge \text{Tray}(o)$
- $f_3(o) \equiv \exists p, r : \text{On}(o, p, r) \wedge \neg \text{Tray}(r)$
- $f_4(o) \equiv \exists p, r \forall o_2, p_2 : \text{Tray}(r) \wedge \neg \text{On}(o, p, r) \wedge \text{On}(o_2, p_2, r) \Rightarrow \neg \text{CFree}(o, p, o_2, p_2, r)$

When these quantified concepts are converted to features, we get numeric evaluations of each concept. Using the final concept as an example, the feature corresponding to this concept returns the number of objects that can be placed on a tray without collision with existing objects on the tray. This feature combines sampled continuous properties of the state, such as collision constraints, with symbolic properties such as `On` and `Tray`. It is easy to imagine how this feature might be useful in a generalized task and motion plan for the LOAD task.

Abstraction Learning. Given a feature pool generated by the above grammar and the set of collected transitions, we now try to find a subset of that feature pool that satisfies the soundness, completeness, and goal-distinguishability criteria for the example plans in the training dataset. We encode the desired constraints (outlined in the appendix) as a SAT theory, and then apply a SAT solver to generate an approximately valid abstraction. Our implementation uses the OpenWBO (Martins, Manquinho, and Lynce 2014) weighted Max-SAT solver and sets weights on the features that correspond the complexity of that feature in the grammar. After applying the SAT solver to our working LOAD task to enforce our criteria over the set of selected features, we find that the lowest-weight set of features satisfying the correctness criteria is exactly those specified in the previous section; the extracted actions are in Figure 1. This is typically the computationally slowest part of the policy learning pipeline because constructing the SAT theory requires evaluation of every numeric feature on all example transitions (see Table 1 in the appendix).

Because of the non-closed world assumption of sampled entities and certified predicates, we cannot guarantee the correctness of any feature evaluation on the state that quantifies over continuous state variables. For example, we cannot assert the correctness of the Boolean feature “All object

grasps are not reachable” because there are an infinite number of possible grasps. However, assuming the sampled entities are drawn uniformly from their respective continuous spaces, the features will be evaluated correctly in the limit of infinite samples. Because this uniform assumption is violated when extracting sampled entities in the preimage of TAMP plans, we add additional stream samples in each plan prior to abstraction learning. Details of this procedure are in the appendix.

Qualitative Numeric Planning. Given an abstraction, we apply the feature transformation to the first and last states of the training instances to get abstract representations of our initial state and goal. This abstract fully observable non-deterministic planning (FOND) problem is then passed to a standard non-deterministic planner to find a generalized plan. In this work, we use state-of-the-art PRP (Muisse, McIlraith, and Beck 2012) as our FOND planner. Figure 1 shows the resulting generalized plan for the working LOAD task given our learned features.

Using a Generalized Task and Motion Plan

We now discuss how the learned generalized plan can be concretized (or executed) in a novel TAMP problem instance within the same task. Typical approaches to generalized planning execute learned policies by selecting randomly from a set of ground actions that 1.) are feasible in the state and 2.) induce the same effects on the abstract state as the abstract action from the policy. However, direct execution of a generalized task and motion plan is impossible due to additional geometric and physical constraints induced by the environment.

One common structure for TAMP planners is to have an outer loop that searches over plan “skeletons”, which consist of operator instances with the discrete parameters bound, but with the continuous parameters still free, and a set of constraints that must be satisfied in order to make the plan valid. Then, an inner loop searches for a satisfying assignment of the continuous parameters. To embed our learned policy into the standard TAMP framework, we construct a new set of planning operators that encodes the action-ordering constraints and abstract action preconditions and effects of the generalized plan, and feed these new operators into PDDL-STREAM. The resulting constrained domain can perform a more efficient concurrent search through the joint space of skeletons and continuous-parameter bindings, thus reducing the overall search time of the TAMP solver. The details of this algorithm are in Algorithm 1 of the appendix.

Using our LOAD task as an example, while a traditional TAMP solver would explore actions that place objects on the table, this would be strictly infeasible in the constrained task-specific problem specification because the abstract LOAD policy specifies that if an object is being held, the next action must increase the number of objects on the tray (Figure 1).

Experiments

We demonstrate GENTAMP on four continuous robotic TAMP tasks described below and visualized in Figure 3. We

chose tasks that are physical instantiations of widely used tasks in generalized planning literature (Srivastava, Immerman, and Zilberstein 2011; Bonet, Francès, and Geffner 2018; Illanes and McIlraith 2019) or classically difficult TAMP domains with large numbers of objects. Here we describe the goal of each of the tasks and the qualitative behavior of the policies discovered by GENTAMP.

We compare **GENTAMP** with two baseline approaches. **Unguided TAMP** is simply an execution of a state-of-the-art task and motion planner (the same one used for data collection). **Partial skeleton** uses the same task and motion planner but provided with an input skeleton that contains a known feasible action schema ordering for each task. The actions in the skeleton do not have populated object parameters, reducing the problem to constraint satisfaction. We consider four tasks within a domain containing a PR2 robot with full kinematic constraints.

Unstack: The objective of this task is to remove all of the stacked blocks and place them on the surface of the table. The policy discovered by GENTAMP is to find a block with no other block stacked on top of it, pick it up, and place it on the table until no stacked blocks remain.

Sort: The objective of this task is to place all of the objects on the table into their similarly colored regions. The policy discovered by GENTAMP is to place all the green objects in the green region, followed by placing all red objects on the red region. This policy is discovered despite none of the example plans sorting in this color-specific order.

Clutter: The goal is to pick up the target object, but this is made difficult by obstruction from many distractor objects. The policy discovered by GENTAMP is to identify the objects it collides with when reaching for the target object and move those objects out of the way by placing them elsewhere on the table until it can reach the target object. This task is known to be difficult for TAMP solvers and has previously required significant engineering, including collision error feedback, to be feasible in large problems (Srivastava et al. 2014). GENTAMP avoids such engineering by learning features from small examples that guide search toward moving objects that block trajectories.

Transport: The objective of this task is to carry all of the objects on the source platform to the destination platform. The agent has access to a tray that it can use to carry over multiple objects to the destination region. The strategy discovered by GENTAMP is to place objects onto the tray until there is no more physical space on the tray, carry the tray to the destination location, and unload the tray until no objects remain on the source location. Because objects are different sizes, the generalized TAMP plan reasons about geometry to determine if more objects can be placed on the tray.

As seen in Figure 3, our experiments show that execution speed from the GENTAMP policy greatly outperforms a state-of-the-art TAMP algorithm. The difference in planning speed increases exponentially with object count. Visualization of the topology for each of the tasks and the learned abstractions along with further experimental details can be found in the appendix.

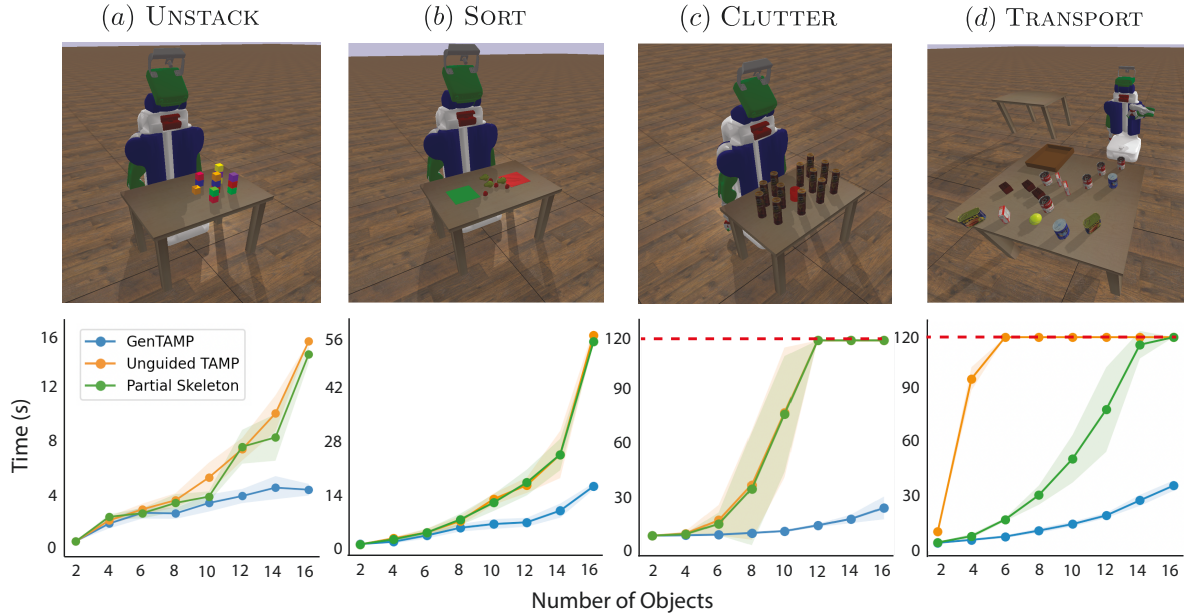


Figure 3: The top panels show sample initial states for our continuous TAMP problems. The bottom four panels show planning speed comparisons between our approach, standard TAMP, and partial skeleton. Each trial contains six seeds used to randomize object type, placement, size, and color. The plot error bars show a 95% confidence interval. The red line indicates a timeout.

Related Work

Generalized planning for purely discrete domains has been the subject of much prior work (Jiménez, Segovia-Aguas, and Jonsson 2019). One of the earliest approaches to generalized planning was a generate-and-test formulation where plans are found for single problem instances and tested on others (Levesque 2005; Hu and Levesque 2009). An alternative approach is to iteratively merge example plans into a single generalized plan (Srivastava, Immerman, and Zilberstein 2011; Srivastava et al. 2011a). Generalized planning has also been accomplished using heuristic-based solvers that search through the entire space of generalized plans (Lotinac et al. 2016; Aguas, Celorrio, and Jonsson 2016). Some more recent work finds generalized plans in symbolic domains by searching through a space of learned features and abstractions to find successful policies (Bonet et al. 2019; Bonet, Francès, and Geffner 2018; Francès, Bonet, and Geffner 2021). Our GENTAMP approach most closely follows this strategy (Bonet et al. 2019) with a focus on applications to continuous geometric domains through the use of a continuous TAMP solver.

There have been various attempts at using offline reinforcement learning and imitation learning to discover useful policies for robotic agents from example traces when no domain description is available. However, reinforcement learning typically struggles with sample complexity and generalizability to domains not seen during training. While some work has been done on finding generalizable policies, these generalization capabilities are limited to visual changes or changes to the dynamics of the simulated environments (Lee et al. 2019; Tobin et al. 2017; Peng et al. 2018). To our

knowledge, there is no existing imitation learning or offline reinforcement learning method that generates policies for robotics domains that can learn from a few examples and generalize to arbitrary geometry and objects.

The problem of accelerating TAMP or planning in general by learning from previous planning solutions is an active area of research. One class of methods accelerates TAMP by reducing the size of the search space in various ways. Such methods include identifying goal-irrelevant objects during search (Silver et al. 2021), focusing the search on certain regions of the state space by learning constraints (Chitnis et al. 2020), performing learned partial grounding (Gnad et al. 2019) of operators. Other methods attempt to accelerate TAMP by biasing the sampling procedures embedded within the continuous TAMP algorithm by learning scoring functions over the sample space (Kim et al. 2018). We aim to speed up TAMP in tasks where all objects, operators, and state-space regions are potentially relevant without biasing the internal sampling strategies.

Conclusion

In this paper, we introduced a new generalized task and motion planning problem definition inspired by a desire to find generalizable and scalable policies for continuous robotics tasks. We designed an algorithm for solving this generalized TAMP problem by learning task-specific abstract policies that apply to problems with arbitrary numbers of objects and initial states from only a few example plans. Lastly, we showed that these simple policies increase the search speed of TAMP solvers in large domains, leading to more scalable and generalizable long-horizon robotic planning.

Acknowledgments

We gratefully acknowledge support from NSF grant 1723381; from AFOSR grant FA9550-17-1-0165; from ONR grant N00014-18-1-2847; from the Honda Research Institute; and from MIT-IBM Watson Lab. Aidan Curtis and Tom Silver are supported by NSF GRFP fellowships. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of our sponsors.

References

- Aguas, J. S.; Celorrio, S. J.; and Jonsson, A. 2016. Hierarchical Finite State Controllers for Generalized Planning. In *IJCAI*.
- Bonet, B.; Francès, G.; and Geffner, H. 2018. Learning Features and Abstract Actions for Computing Generalized Plans. *CoRR*, abs/1811.07231.
- Bonet, B.; Fuentetaja, R.; E-Martín, Y.; and Borrajo, D. 2019. Guarantees for Sound Abstractions for Generalized Planning (Extended Paper). *CoRR*, abs/1905.12071.
- Bonet, B.; and Geffner, H. 2018. Features, Projections, and Representation Change for Generalized Planning. *CoRR*, abs/1801.10055.
- Chitnis, R.; Silver, T.; Kim, B.; Kaelbling, L. P.; and Lozano-Perez, T. 2020. CAMPs: Learning Context-Specific Abstractions for Efficient Planning in Factored MDPs. arXiv:2007.13202.
- Clocksin, W. F.; and Mellish, C. S. 2003. *Programming in Prolog*. Berlin: Springer, 5 edition. ISBN 978-3-540-00678-7.
- Francès, G.; Bonet, B.; and Geffner, H. 2021. Learning General Policies from Small Examples Without Supervision. arXiv:2101.00692.
- Garrett, C. R.; Chitnis, R.; Holladay, R.; Kim, B.; Silver, T.; Kaelbling, L. P.; and Lozano-Pérez, T. 2020. Integrated Task and Motion Planning. arXiv:2010.01083.
- Garrett, C. R.; Lozano-Pérez, T.; and Kaelbling, L. P. 2018. STRIPStream: Integrating Symbolic Planners and Blackbox Samplers. *CoRR*, abs/1802.08705.
- Gnad, D.; Torralba, ; Domínguez, M.; Areces, C.; and Bustos, F. 2019. Learning How to Ground a Plan – Partial Grounding in Classical Planning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01): 7602–7609.
- Hu, Y.; and Levesque, H. 2009. Planning with loops:1 Some new results. In *ICAPS Workshop on Generalized Planning*, 37.
- Illanes, L.; and McIlraith, S. A. 2019. Generalized Planning via Abstraction: Arbitrary Numbers of Objects. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01): 7610–7618.
- Jiménez, S.; Segovia-Aguas, J.; and Jonsson, A. 2019. A review of generalized planning. *The Knowledge Engineering Review*, 34: e5.
- Kim, B.; Wang, Z.; Kaelbling, L. P.; and Lozano-Pérez, T. 2018. Learning to guide task and motion planning using score-space representation. *CoRR*, abs/1807.09962.
- Lee, K.; Lee, K.; Shin, J.; and Lee, H. 2019. A Simple Randomization Technique for Generalization in Deep Reinforcement Learning. *CoRR*, abs/1910.05396.
- Levesque, H. 2005. Planning with Loops. In *IJCAI*.
- Lotinac, D.; Aguas, J. S.; Celorrio, S. J.; and Jonsson, A. 2016. Automatic Generation of High-Level State Features for Generalized Planning. In *IJCAI*.
- Martins, R.; Manquinho, V.; and Lynce, I. 2014. Open-WBO: A Modular MaxSAT Solver,. In Sinz, C.; and Egly, U., eds., *Theory and Applications of Satisfiability Testing – SAT 2014*, 438–445. Cham: Springer International Publishing. ISBN 978-3-319-09284-3.
- Muise, C.; McIlraith, S. A.; and Beck, J. 2012. Improved Non-Deterministic Planning by Exploiting State Relevance. In *ICAPS*.
- Peng, X. B.; Andrychowicz, M.; Zaremba, W.; and Abbeel, P. 2018. Sim-to-Real Transfer of Robotic Control with Dynamics Randomization. *2018 IEEE International Conference on Robotics and Automation (ICRA)*.
- Silver, T.; Chitnis, R.; Curtis, A.; Tenenbaum, J.; Lozano-Perez, T.; and Kaelbling, L. P. 2021. Planning with Learned Object Importance in Large Problem Instances using Graph Neural Networks. In *AAAI*.
- Srivastava, S.; Fang, E.; Riano, L.; Chitnis, R.; Russell, S.; and Abbeel, P. 2014. Combined task and motion planning through an extensible planner-independent interface layer. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, 639–646.
- Srivastava, S.; Immerman, N.; and Zilberstein, S. 2011. A new representation and associated algorithms for generalized planning. *Artif. Intell.*, 175: 615–647.
- Srivastava, S.; Immerman, N.; Zilberstein, S.; and Zhang, T. 2011a. Directed Search for Generalized Plans Using Classical Planners. In *ICAPS*.
- Srivastava, S.; Zilberstein, S.; Immerman, N.; and Geffner, H. 2011b. Qualitative Numeric Planning. In *Proceedings of the Twenty-Fifth Conference on Artificial Intelligence*, 1010–1016. San Francisco, California.
- Tobin, J.; Fong, R.; Ray, A.; Schneider, J.; Zaremba, W.; and Abbeel, P. 2017. Domain Randomization for Transferring Deep Neural Networks from Simulation to the Real World. *CoRR*, abs/1703.06907.