

A Categorical Representation Language and Computational System for Knowledge-Based Robotic Task Planning

Angeline Aguinaldo^{1,2}, Evan Patterson³, James Fairbanks⁴, William Regli¹, Jaime Ruiz⁴

¹ University of Maryland, College Park

² Johns Hopkins University Applied Physics Laboratory

³ Topos Institute

⁴ University of Florida

{aguinal, regli}@cs.umd.edu, evan@topos.institute, {fairbanksj, jaime.ruiz}@ufl.edu

Abstract

Classical planning representation languages based on first-order logic have preliminarily been used to model and solve robotic task planning problems. Wider adoption of these representation languages, however, is hindered by the limitations present when managing implicit world changes with concise action models. To address this problem, we propose an alternative approach to representing and managing updates to world states during planning. Based on the category-theoretic concepts of C-sets and double-pushout rewriting (DPO), our proposed representation can effectively handle structured knowledge about world states that support domain abstractions at all levels. It formalizes the semantics of predicates according to a user-provided ontology and preserves the semantics when transitioning between world states. This method provides a formal semantics for using knowledge graphs and relational databases to model world states and updates in planning. In this paper, we conceptually compare our category-theoretic representation with the classical planning representation. We show that our proposed representation has advantages over the classical representation in terms of handling implicit preconditions and effects, and provides a more structured framework in which to model and solve planning problems.

Introduction

In robotic task planning, tracking all the implicit effects and relationships in the world state can be very challenging, especially when working in complex environments. As a result, planning systems often rely on heuristics and simplifying assumptions, which can lead to suboptimal or even incorrect solutions (Wilkins and DesJardins 2001; Gil 1990). To address these challenges, researchers have sought to combine techniques from the fields of knowledge representation and automated planning which has formed the subfield called *knowledge-based planning* (Wilkins and DesJardins 2001). The goal of this research is to develop more structured and efficient representations of the world state that can capture the complex relationships and dependencies that arise in real-world planning problems. For example, some approaches use structured knowledge representations, such as ontologies (Kang and Choi 2009; Beetz et al. 2018;

Tenorth and Beetz 2017) and knowledge graphs (Amiri, Chandan, and Zhang 2022; Galindo et al. 2008; Miao, Jia, and Sun 2023; Agia et al. 2021), to model the world state. It is natural to conclude that knowledge-based planning, as it stands, may be better suited to solve the problem of tracking implicit preconditions and effects. On the contrary, knowledge-based planning has not been widely adopted because it requires translating these rich and complex representations into propositional facts so that they can be interpreted using a classical planning languages like PDDL (Knoblock et al. 1998). Because of this, they become computationally intensive search problems in the absence of search space reduction methods.

To address these concerns, we propose a world state representation based on the category-theoretic concepts of C-sets (Patterson, Lynch, and Fairbanks 2021) and double-pushout (DPO) rewriting (Brown et al. 2021). Our representation not only manages structured knowledge about the world state, but also formalizes the semantics of predicates according to a user-provided ontology, ensuring that the semantics of the world state are preserved and implicit preconditions and effects are handled when transitioning between states.

Related Work

We foresee the nearest application of our approach as being robotic task planning using scene graphs. Scene graphs are a specialized version of knowledge graphs that restrict its elements, attributes, and relations to facts obtained through vision-based perception and inference (Chang et al. 2023). In scene graphs, ontologies are often used to align scene data to class hierarchies. In planning, these ontologies can be used to enrich facts in the planning domain; however, these ontologies are often integrated in the planning decisions in ad hoc ways. For example, Galindo et al present a two-part knowledge representation system, which includes (i) spatial information about the robot environment in the form of a scene graph, and (ii) an ontology that describes the hierarchical relationships between concepts (Galindo et al. 2008). A function mapping elements in the scene graph to concepts in the ontology is defined. Both facts obtained through the scene graph and facts obtained from the ontology are translated into propositions in the domain. This approach results in an explosion of facts that requires a pruning step in order to be tractable for classical planning approaches.

This problem of scene graphs being too complex, with numerous vertices and edges, is long standing. To mitigate this, planners can employ procedures that determine which attributes of the scene are most relevant while also preserving the semantics provided by the class hierarchy and item features. An example of a planner designed for this purpose is the SCRUB planner (Agia et al. 2021). SCRUB is a planner-agnostic procedure that prunes the state space to include only the relevant facts within a scene graph. It is paired with SEEK (Agia et al. 2021), a planner-agnostic procedure that scores items in the scene based on an importance score produced by a graph neural network (Silver et al. 2021). All elements that are ancestors to the relevant elements, according to some threshold, are preserved as facts in the state. Both the SCRUB and SEEK procedures dramatically reduce the number of facts needed to characterize the world state. The facts in the world are translated into binary predicates and passed to a classical planner. This provides a heuristic-based measure for identifying relevant facts which can, like most heuristics, produce inaccurate approximations. The combinatorial approach we propose uses the existing semantic structure to determine relevance.

Miao et al take an alternate approach to using scene graphs to describe world states and action models (Miao, Jia, and Sun 2023). In their approach, action operators are specified in terms of an initial state subgraph, a final state subgraph, and an intermediate subgraph. For each element and relation in these subgraphs, the global scene graph is updated by adding elements into the scene graph that are discussed in the action model. If elements are referenced in the final state subgraph that are not present in the scene, they are introduced as isolated vertices. These vertices are connected to the global scene graph by consulting an external knowledge base that contains a type hierarchy. Their procedure searches for a matching element type, identifies a parent type that exists in the scene graph, and defines an edge from that element to the existing type. This is a fragile approach to resolving changes in the world state because it relies on an external knowledge base to be correctly and completely instantiated in order for new information to be properly integrated.

Overall, using scene graphs as a representation for world states during planning has gained attention. These methods satisfy the need to support rich and complex representations of the world using ontologies but still suffer in their ability to integrate ontological information in a way that does not cause a state explosion. In our approach, the ontology (schema) is an integral part of the formalism and with it comes specialized tooling for manipulating such data.

Preliminaries

In this section, we explain the mathematical model for planning as a state transition system and briefly discuss the use of category theory.

Planning as a State Transition System

The state transition system model of planning is a formal representation of a planning problem that describes the state

space of the problem and the possible actions that can be taken to move between states. In this model (Ghallab, Nau, and Traverso 2004), a *planning problem* can be defined as a tuple $P = \langle S, A, \gamma \rangle$, where:

- The *state space* $S = \{s_0, s_1, s_2, \dots\}$ is the set of all possible states. A state $s \in S$ represents a snapshot of the world at a particular point in time. It ideally includes all relevant information about the state of the world, such as the location of items and their properties.
- The *action space* $A = \{a_0, a_1, a_2, \dots\}$ is the set of all possible actions. An action represents a transition from one state to another state.
- The *transition function* $\gamma : A \times S \rightarrow S$ is a partial function that, where it is defined, maps an action and a state to the next state.

An action $a \in A$ is *applicable* at state $s \in S$ if $\gamma(a, s)$ is defined. A *plan*, $\pi = \langle a_1, a_2, \dots, a_n \rangle$, $a_i \in A$, is any sequence of actions. It is a *solution* to the planning problem if it transitions from the initial state s_0 to the goal state s_g , i.e., $\gamma(a_i, s_{i-1}) = s_i$ for $i = 1, \dots, n$ and $s_n = s_g$.

The state transition system model of planning provides a structured and formal way to represent planning problems and reason about the possible sequences of actions that can be taken to achieve a goal. The manner in which states, actions, and transition functions are represented both classically and categorically is the focus of this paper.

Category Theory

Category theory is a mathematical framework that aids in understanding the relationships between different mathematical structures. In recent years, it has found important applications in a variety of fields, including systems engineering and design (Breiner, Pollard, and Subrahmanian 2019; Breiner, Subrahmanian, and Sriram 2016; Censi 2016), robotics and planning (Aguinaldo et al. 2021; Aguinaldo and Regli 2021; Master et al. 2020), and physics (Baez and Stay 2011; Abramsky and Coecke 2008). For the purposes of planning, category theory provides an alternative mathematical language for representing world states and action models.

Classical Representations

The literature on classical and neoclassical planning encompasses a variety of approaches to modeling the state of the world. Among these, the most widely employed domain-independent representations are the set-theoretic, classical, and state-variable representations (Ghallab, Nau, and Traverso 2004).

Existing Representations for World States, S The classical representation of world states is based on a restricted form of first-order logic in which the world states are represented as a conjunction of *literals*, where a literal is an atomic proposition or its negation. These literals can encode facts such as the location of items, the state of various sensors, and other relevant information. Logical operators such as AND, OR, and NOT are used to combine these literals into more complex expressions that can be used to describe

the relationships between different parts of the world state. World states can be lifted to abstract states consisting of a conjunction of *predicates*, where predicates are n -ary relations between variable symbols. For example, the predicate $\text{on}(x, y)$ might represent the fact that something represented by x is on top of another thing represented by y or, equally plausible, the converse. It is important to note that the semantics of the predicates must be established informally, say through external documentation or word of mouth. A predicate becomes a grounded literal when the variable symbols, like x and y , are assigned to symbols that are constant. For instance, the grounded literal $\text{on}(b_1, b_2)$ could represent a block called b_1 that is on top of a block called b_2 .

Existing Representations for Actions, A The classical representation defines action schemas in terms of preconditions, effects, and parameters. Preconditions describe the conditions that must be true for the action to be applicable, while effects describe the changes that the action makes to the state of the world. Preconditions and effects are defined in terms of predicates where their variable symbols, also known as parameters, can take on different values depending on the context in which the action is used. The classical planning representation assumes that actions are deterministic, meaning that their effects are completely specified and do not depend on any probabilistic factors. Additionally, the classical representation assumes that actions can be arranged in any order, allowing for a wide variety of possible plans to be generated. In practice these action schemas can be augmented with types; however, types are not a formal feature of the logical system.

Applicability In classical planning, the applicability of actions to a world state is determined by a set of preconditions specified for each action. In the classical representation, preconditions are represented as conjunctions of literals. To check whether an action is applicable in a particular world state, the planner examines whether the preconditions of the action are satisfied by the current state. To do this, a planner typically uses logical inference techniques within propositional logic, first-order logic, or other logical formalisms. Once grounded, conjunction of predicates can be represented as sets of ground literals. The planner can then check whether the precondition is a subset of the world state, which means that the preconditions are satisfied in the world state. In the case of first-order logic, the world state and the preconditions of the action can be represented as sentences in a formal language. The planner can then use logical inference algorithms to determine whether the preconditions logically entail the world state, which means that the preconditions are satisfied in the world state. Such logical inference algorithms for the classical representation are PSPACE-complete (Russell et al. 2021); therefore, heuristics and other optimizations are often employed to make finding solutions more tractable. It is possible to model domain-specific implications, called derived predicates in PDDL (Thiébaux, Hoffmann, and Nebel 2005), to incorporate implicit preconditions when determining applicability of an action; however, derived predicates do not make any changes to the world state.

Transition, γ A state transition occurs when an action is applied to a world state, yielding a new world state. Similar to the preconditions, the world state can be interpreted as a set of literals. Therefore, when an action is applied, the new state is identified by applying set operations to the set representing the current world state. Positive effects add ground literals to the set and negative effects, those prefixed with NOT, subtract ground literals from the set to produce the new world state.

Categorical Representation

To effectively manage world states during planning and plan execution, we propose to adopt categorical logic as a logical formalism. An essential feature of categorical logic is the ability to capture the relationship between logical syntax (theories) and semantics (models) using functors, through a paradigm known as *functorial semantics*. This differs from the standard formalism of first order logic because it gives the syntax of the language status as an algebraic object independent of its semantics.

World States as C-sets

In our approach, we assume world states in robot task planning domains can be effectively modeled using scene graphs, which are structured representations of the elements, attributes, and relations based on a domain-specific ontology. Recall that scene graphs are a specialized form of knowledge graphs that focus on the visual aspects of a scene, with nodes representing items and edges representing the spatial relationships between them (Chang et al. 2023). These graphs can be used to capture a wide range of information about a world state, including the physical and semantic properties of items, their interactions with each other, and the context in which they exist.

Our proposed representation provides a sort of denotational semantics for scene graphs using the category C-Set. Let C denote a small category, called a *schema*. A *C-set*¹ is a functor² from C to the category Set. The schema is a category whose objects we interpret as types and whose morphisms describe “is-a” and other functional relationships between types. The category Set is the category of sets and functions. Thus, a C-set is a functor that sends types to sets and type relationships to functions. On this interpretation, C-sets are a simple but useful model of relational databases (Spivak and Kent 2011).

A more formal definition is provided below.

Definition 0.1 (Category of C-sets, C-Set). For a given schema C , the *category of C-sets* is the functor category $C\text{-Set} := \text{Set}^C$, whose objects are functors from C to Set and whose morphisms are natural transformations between those.

In practice, the schema is a category that is finitely presented by generators and relations. As an example, let $\text{Gr} =$

¹A C-set also known as a *copresheaf* on C .

²We direct the unfamiliar reader to (Leinster 2016) for the definitions of categories, functors, natural transformations, limits, and colimits.

```

rdf(Sam::Person, s::favoritePet, cat::Animal)
rdf(Molly::Person, s::favoritePet, dog::Animal)

```

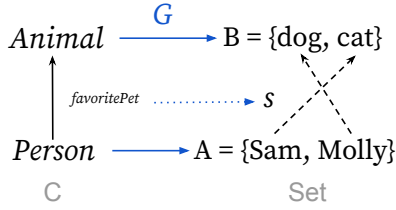


Figure 1: An example C-set, G , that stores data about people’s favorite pet. The category of elements contains triples analogous to RDF triples.

$\{E \rightrightarrows V\}$ be the category freely generated by two objects, E and V , and two parallel morphisms, $\text{src}, \text{tgt} : E \rightarrow V$. Then a Gr-set is a graph, which would be called by graph theorists a “directed multigraph.” So, another interpretation of C-sets is that they are a generalization of graphs to a broad class of combinatorial data structures.

The *category of elements* (Riehl 2016) construction³ of a C-set X packages the data of X into a category resembling a knowledge graph. Specifically, a morphism in the category of elements can be interpreted as a Resource Description Framework (RDF) triple (Patterson 2017), which is a common text-based serialization format for knowledge and scene graphs. A C-set is shown in this style in Figure 1.

Actions as Spans in C-Set

Action rules are specified as spans in C-Set. Action rules are made up of components similar to those of action schemas in classical representation. Specifically, actions rules are spans ($I \leftarrow K \rightarrow O$) in C-Set that consists of the *precondition*, I , on the left-hand side, the *effects*, O , on the right-hand side, and the *glue*, K , in the middle, which gives the data that remains unchanged between the input and the output.

In our approach, we present spans of C-sets as colimits of representable functors. A (covariant) representable functor (MacLane 1971), $C(A, -)$, for a given category C and object $A \in C$ maps objects, $X \in C$, to the set of morphisms that go from A to X . In other words, for a given representable functor pertaining to A , objects $X \in C$ are mapped to the set of edges that go from A to X . A colimit taken across functors like these glues all of the relevant structures together. This conversion ensures that implicit substructure is taken into account when an object, such as A , is explicitly identified.

The categorical rule specification differs from the classical one in that it takes a declarative approach by not articulating what atoms should be added and removed from the state, but rather discussing what should be in the state and resolving conflicts using the double-pushout (DPO) rewriting procedure which is discussed in a later section.

³The category of elements construction of X may be expressed using $\int X$ in other texts.

Applicability Using Monomorphisms

Recall that in classical planning, a precondition is satisfied by a world state when its literals are a subset of the world state or if there exists a logical entailment between the precondition and the world state. In the category-theoretic context, these notions are generalized via monomorphisms.

Monomorphisms generalize the concept of an injective function to arbitrary categories. In Set, monomorphisms are precisely injective functions. In C-Set, monomorphisms are natural transformations such that every component is an injective function, e.g., a monomorphism between graphs is a graph homomorphism such that the vertex and edge maps are both injective. More specifically, the monic condition of a monomorphism checks that two entities in the precondition cannot be mapped to the same entity in the world state.

Transition Using Double-Pushout (DPO) Rewriting

The action rules are exactly double-pushout (DPO) rewriting rules. Double-pushout (DPO) rewriting is a type of graph rewriting that is particularly well-suited for algebraic approaches to graph transformation. In fact, DPO rewriting generalizes directly from graph rewriting to C-set rewriting (Brown et al. 2021). The DPO method, described below, is used to compute all possible matches of the preconditions, and to determine which matches are compatible with the effects. The result is a set of transformation steps that can be applied to the target graph.

DPO rewriting relies on the fundamental concept of a pushout. A pushout is a colimit of a diagram having the shape of a span ($\bullet \leftarrow \bullet \rightarrow \bullet$). Given a span $X \leftarrow Z \rightarrow Y$, a pushout produces an object that is the union of X and Y joined along Z , $(X \cup Y)/Z$. A pushout in C-Set is computed by taking the disjoint union of the sets being pushed out, adding the functions between sets based on C , and quotienting by Z .

Pseudocode for the DPO rewriting procedure is given in Algorithm 1. The first step is to find a monomorphism, m , that matches I in X , as described in the previous subsection. The pushout complement, f , is computed provided the morphisms l and m . A pushout complement is a map that manages the deletion of entities that form the complement K/I . Because i is a monomorphism and we assume the identification and dangling conditions (Brown et al. 2021) are met, the pushout complement exists and is unique up to isomorphism. Having constructed the three sides of the square, l, m, f , the pushout square can be completed by a unique map g . Then, to compute the new world state, the right pushout square is computed provided f and r .

$$\begin{array}{ccccc}
 I & \xleftarrow{l} & K & \xrightarrow{r} & O \\
 m \downarrow & & f \downarrow & & \downarrow \\
 X & \xleftarrow{g} & Z & \longrightarrow & Y
 \end{array}$$

The time complexity of the `FindHomomorphism()` subroutine is $O(n^k)$ where k is the size of I and n is the size of the relevant substructure in X which is dictated by the objects in I (Brown et al. 2021). The time complexity of the remaining subroutines is the same as that of computing

Algorithm 1: Double-Pushout (DPO) Rewriting

Require: (action rule) $I \leftrightarrow K \rightarrow O \in \mathbf{C}\text{-Set}$

Require: (world) $X \in \mathbf{C}\text{-Set}$

$m = \text{FindHomomorphism}(I, X)$

$f = \text{ComputePushoutComplement}(l, m)$

$g = \text{CompletePushout}(l, f, m)$

$Y = \text{ComputePushout}(f, r)$

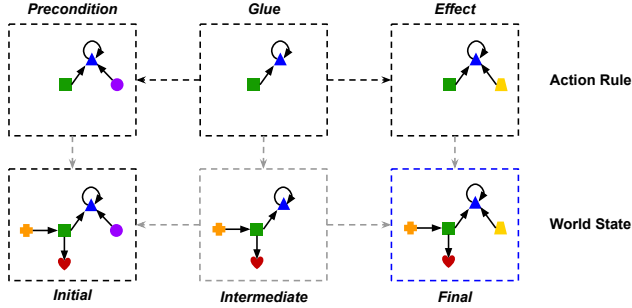


Figure 2: An illustration of how DPO rewriting is executed on a C-set where \mathbf{C} defines the shapes and the arrows between them. Each shape in the figure is assigned to a color to help with readability.

pushouts in \mathbf{Set} which is $O(p)$ where p is the sum of the sizes of the sets involved in the span.

As an example, Figure 2 shows how an action rule changes the initial state to the final state. In this figure, the rule states that the circle in the initial state should be replaced by a trapezoid and the square and triangle should persist. The initial state, bottom-left, satisfies the precondition because it contains a matching pattern involving the square, triangle, and circle. This means that a monomorphism can be identified from the precondition to the initial state. The intermediate state is the result of identifying a pattern that, when joined with the precondition along the glue, produces the initial state. This removes the circle from the initial state. The final state is then constructed by taking the pushout for the span ($\text{Intermediate} \leftarrow \text{Glue} \rightarrow \text{Effect}$). This produces a pattern where a trapezoid is added in place of the original circle. The example shown in Figure 2 demonstrates that rules can involve both the creation and destruction of entities in the world if the precondition contains an entity and the effect does not contain that entity. This allows for a non-monotonicity in updates of the world state.

Double-pushout (DPO) rewriting is reminiscent of Combinatorial Categorical Grammar (CCG) parsing which has been applied to problems in planning, specifically for plan recognition (Geib and Goldman 2009; Geib and Kantharaju 2018). Both methods have components to their rules that align syntax with semantics, however, DPO rewrite rule application can be better thought of as running a program that transform states while adhering to certain ontological properties. CCG, on the other hand, is a term-rewriting system that performs trivial substitutions when rules are applied.

Comparison

In this section, we discuss the conceptual differences between the classical and the categorical representations. The comparison is summarized in Table 1. We also walk through an example, shown in Figure 3, that highlights the limitations of the classical representation in tracking implicit effects. In this example domain, a bread loaf (bread loaf) and slices of the loaf (slice_0, slice_1, slice_2) are on a countertop (countertop). The goal is to move the bread loaf, and implicitly, all its slices, from the countertop to the kitchen table (kitchentable).

Handling Structured Knowledge

This example presents a few noteworthy semantic features. A planning representation in this domain must be able to encode the following facts, explicitly or implicitly, at different points in time:

- (a) the bread slices are part of the bread loaf
- (b) the bread loaf is on the countertop
- (c) the bread slices are on the countertop
- (d) the bread loaf is on the kitchen table
- (e) the bread slices are on the kitchen table

In the case of the categorical representation, fact (a) is captured by the morphism $\text{partOf} : \text{BreadSlices} \rightarrow \text{BreadLoaf}$ in the schema. Fact (b) about the bread loaf being on the countertop is reified through an abstract type called `Object`. This is done so that the morphism $\text{on} : \text{Object} \rightarrow \text{Object}$ can represent the general notion of an item being on top of another item, instead of the more specific relation of a bread loaf being on a countertop. Fact (c) is then captured by the composite morphism $\text{on} \circ \text{partOf}$ is a $\text{partOf} : \text{BreadSlices} \rightarrow \text{Object}$ ⁴. In the classical representation, fact (a) is captured by the propositions $\text{partOf}(\text{slice}_n, \text{bread})$ for $n = 0, 1, 2$. Fact (b) and (d) are captured by the propositions $\text{on}(\text{bread}, \text{countertop})$ and $\text{on}(\text{bread}, \text{kitchentable})$. Fact (c) and (e) are captured by the propositions $\text{on}(\text{slice}_n, \text{countertop})$ and $\text{on}(\text{slice}_n, \text{kitchentable})$. Intuitively, because the bread loaf is on the countertop, the slices that make up the loaf are also on the countertop. In the categorical representation, this is captured using a composite morphism. In the classical representation, this is done by explicitly stating for each slice, that it is on the countertop.

Handling Applicability of Actions

In the categorical representation, applicability of an action is determined by the existence of a monomorphism in $\mathbf{C}\text{-Set}$ from the rule input to the world state. Recall that a representable functor maps an object, $x \in \mathbf{C}$, to the set of morphisms that involve x . When you present an action using colimits of representables, there are both explicit conditions given by the representables and implicit conditions that appear when the representable is computed. This provides a mechanism for having implicit conditions represented in the action schemas. In this example, the initial

⁴The operator \circ represents a composite function which can be read as "f following g" or $(f \circ g)(x) = f(g(x))$.

Representation	Categorical	Classical
State, S	C-set	Conjunction of literals
Action, A	Spans in C-Set containing: Preconditions, Glue, Effects	Action model containing: Parameters, Preconditions, Effects
Applicability	Monomorphisms in C-Set	Subset inclusion
Transition, γ	DPO rewriting	Set-based addition and subtraction

Table 1: A summary of the differences between the classical representation and the categorical representation aligned to the state transition system model for planning.

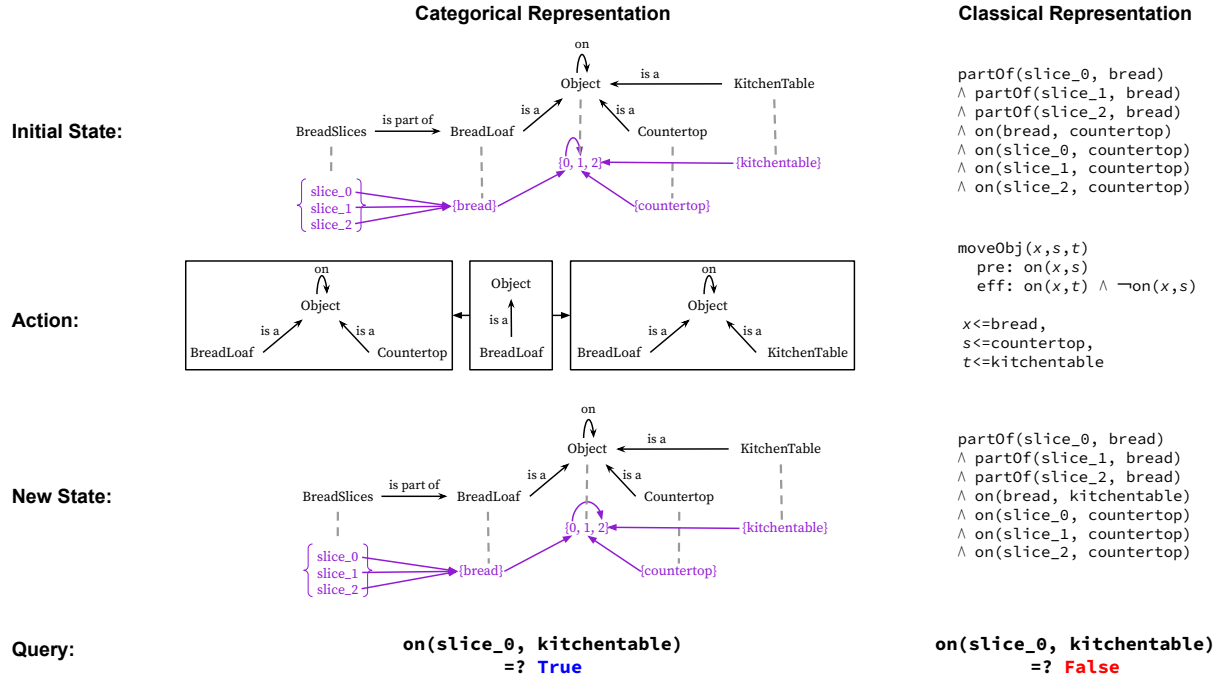


Figure 3: A comparison of states, actions, and inferences made between the categorical representation and the classical representation for an example that moves a loaf of bread from a countertop to a kitchen table. This example illustrates a failure of the classical representation to preserve the global semantics of the world when actions act on only a part of the world. Using the double-pushout method of C-set rewriting, the categorical representation is able to do so.

state satisfies the input of the action rule because it contains BreadLoaf, Object, and Countertop and its related morphisms. In the classical representation, applicability of the action $\text{moveObj}(\text{bread}, \text{countertop}, \text{kitchentable})$ is determined by whether or not the world state contains the element $\text{on}(\text{bread}, \text{countertop})$ in the set of propositions.

Handling Implicit Effects

Now that we have determined that this action is applicable in both representations, we can apply the action. In the categorical representation, this action is modeled by a span in C-Set whose left foot includes knowledge about the bread loaf being on the countertop and whose right foot includes knowledge about the bread loaf being on the kitchen table. The apex of the span states that the bread loaf itself is preserved throughout this change. In the classical representation, the action schema describes the generic action of mov-

ing items x from one location, s , to another, t . The action operator is grounded by assigning x to the breadloaf, s to the countertop, and t to the kitchentable. Once this action is applied, a desirable outcome would be for the new state to account for the movement of the bread slices from the countertop to the kitchen table because they are part of the breadloaf. In the categorical representation, the same composite morphism that existed in the initial state exists in the final state; however, the target of the morphism has changed from the countertop to the kitchentable. This captures the implicit change that occurred to the bread slice locations. In the classical representation, the new state captures the fact that the breadloaf is on the kitchen table, but does not capture the fact that the bread slices are on the kitchen table. This error would likely cause a planner to instruct an agent to move each bread slice individually due to an inconsistency in the world state.

Conclusion

The limitations of classical planning representation languages in tracking implicit preconditions and effects have motivated us to propose an alternative world state representation based on the category-theoretic concepts of C-sets and DPO rewriting. Our categorical representation accommodates structured knowledge about the world state and formalizes a model of the application domain using a user-provided ontology. This method provides formal semantics for using knowledge graphs and relational databases to model world states and updates in planning. Our comparison between the classical and categorical planning representation languages demonstrates that our proposed representation is more structured and has advantages over the classical one in handling complex planning scenarios. Future work include designing mechanisms for transferring plans between domains and extending this system to handle non-combinatorial data, such as numeric data. We believe that our proposed representation has the potential to significantly enhance the effectiveness and efficiency of planning systems in robotic task planning domains.

Acknowledgments

This work was partially funded by the U.S. Defense Advanced Research Projects Agency under contract #HR00112220004. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect these agencies' views. We thank Kristopher Brown and Owen Lynch from the Topos Institute for heavily contributing to the development of DPO rewriting and C-sets in AlgebraicJulia. We also thank the students of the Ruiz HCI Lab—Alexander Barquero, Rodrigo Calvo, Niriksha Regmi, Daniel Delgado, Andrew Tompkins—who built the task guidance platform that inspired the examples in this paper. Lastly, we thank Dana Nau for valuable feedback about this paper.

References

- Abramsky, S.; and Coecke, B. 2008. Categorical Quantum Mechanics. *Handbook of Quantum Logic and Quantum Structures vol II*, 1–63.
- Agia, C.; Jatavallabhula, K. M.; Khodeir, M.; Miksik, O.; Vineet, V.; Mukadam, M.; Paull, L.; and Shkurti, F. 2021. TASKOGRAPHY: Evaluating robot task planning over large 3D scene graphs. *5th Conference on Robot Learning*, 1–18.
- Aguinaldo, A.; Bunker, J.; Pollard, B.; Shukla, A.; Canedo, A.; Quiros, G.; and Regli, W. 2021. RoboCat: A Category Theoretic Framework for Robotic Interoperability Using Goal-Oriented Programming. *IEEE Transactions on Automation Science and Engineering*, 1–9.
- Aguinaldo, A.; and Regli, W. 2021. A Graphical Model-Based Representation for PDDL Plans using Category Theory. In *ICAPS 2021 Workshop XAIP*.
- Amiri, S.; Chandan, K.; and Zhang, S. 2022. Reasoning with Scene Graphs for Robot Planning under Partial Observability. *IEEE Robotics and Automation Letters*, 7(2): 5560–5567.
- Baez, J.; and Stay, M. 2011. Physics, topology, logic and computation: A Rosetta Stone. *Lecture Notes in Physics*, 813: 95–172.
- Beetz, M.; Bessler, D.; Haidu, A.; Pomarlan, M.; Bozcuoglu, A. K.; and Bartels, G. 2018. Know Rob 2.0 - A 2nd Generation Knowledge Processing Framework for Cognition-Enabled Robotic Agents. *Proceedings - IEEE International Conference on Robotics and Automation*, 512–519.
- Breiner, S.; Pollard, B.; and Subrahmanian, E. 2019. Functorial Model Management. *Proceedings of the Design Society: International Conference on Engineering Design*, 1(1): 1963–1972.
- Breiner, S.; Subrahmanian, E.; and Sriram, R. D. 2016. Modeling the Internet of Things: A foundational approach. *ACM International Conference Proceeding Series*, Part F1271: 38–41.
- Brown, K.; Patterson, E.; Hanks, T.; and Fairbanks, J. 2021. Computational category-theoretic rewriting. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*.
- Censi, A. 2016. A Mathematical Theory of Co-Design. Technical report, Laboratory for Information and Decision Systems, MIT.
- Chang, X.; Ren, P.; Xu, P.; Li, Z.; Chen, X.; and Hauptmann, A. 2023. A Comprehensive Survey of Scene Graphs: Generation and Application. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1): 1–26.
- Galindo, C.; Fernández-Madrigal, J. A.; González, J.; and Saffiotti, A. 2008. Robot task planning using semantic maps. *Robotics and Autonomous Systems*, 56(11): 955–966.
- Geib, C. W.; and Goldman, R. P. 2009. A probabilistic plan recognition algorithm based on plan tree grammars. *Artificial Intelligence*, 173(11): 1101–1132.
- Geib, C. W.; and Kantharaju, P. 2018. Learning combinatory categorial grammars for plan recognition. *32nd AAAI Conference on Artificial Intelligence, AAAI 2018*, 3007–3014.
- Ghallab, M.; Nau, D.; and Traverso, P. 2004. *Automated Planning: Theory and Practice*. Morgan Kaufmann. ISBN 9781558608566.
- Gil, Y. 1990. Description Logics and Planning. *AI Magazine*, 26(2): 1.
- Kang, D.; and Choi, H.-j. 2009. Hierarchical Planning through Operator and World Abstraction using ontology for home service robots. *2009 11th International Conference on Advanced Communication Technology*, 2209–2214.
- Knoblock, C.; Barrett, A.; Christianson, D.; Friedman, M.; Kwok, C.; Golden, K.; Penberthy, S.; Smith, D. E.; Sun, Y.; and Weld, D. 1998. PDDL — The Planning Domain Definition Language. Technical report, Yale Center for Computational Vision and Control.
- Leinster, T. 2016. *Basic Category Theory*. Cambridge University Press.
- MacLane, S. 1971. *Categories for the Working Mathematician*. New York: Springer-Verlag. Graduate Texts in Mathematics, Vol. 5.

- Master, J.; Patterson, E.; Yousfi, S.; and Canedo, A. 2020. String Diagrams for Assembly Planning. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*.
- Miao, R.; Jia, Q.; and Sun, F. 2023. Long-term robot manipulation task planning with scene graph and semantic knowledge. *Robotic Intelligence and Automation*.
- Patterson, E. 2017. Knowledge Representation in Bicategories of Relations. *CoRR*, abs/1706.00526.
- Patterson, E.; Lynch, O.; and Fairbanks, J. 2021. Categorical Data Structures for Technical Computing. *Compositionality*, 4(5): 1–27.
- Riehl, E. 2016. *Category theory in context*. ISBN 9780486809038.
- Russell, S.; Norvig, P.; Chang, M.-W.; Devlin, J.; Dragan, A.; Forsyth, D.; Goodfellow, I.; Malik, J.; and Mansinghka, V. 2021. *Artificial Intelligence: a modern approach*. Pearson, 4th edition.
- Silver, T.; Chitnis, R.; Curtis, A.; Tenenbaum, J.; Lozano-Pérez, T.; and Kaelbling, L. P. 2021. Planning with Learned Object Importance in Large Problem Instances using Graph Neural Networks. *35th AAAI Conference on Artificial Intelligence, AAAI 2021*, 13B: 11962–11971.
- Spivak, D. I.; and Kent, R. E. 2011. Ologs: a categorical framework for knowledge representation. *PLOS ONE*.
- Tenorth, M.; and Beetz, M. 2017. Representations for robot knowledge in the KNOWROB framework. *Artificial Intelligence*, 247: 151–169.
- Thiébaux, S.; Hoffmann, J.; and Nebel, B. 2005. In defense of PDDL axioms. *Artificial Intelligence*, 168(1-2): 38–69.
- Wilkins, D. E.; and DesJardins, M. 2001. A call for knowledge-based planning. *AI Magazine*, 22(1): 99–115.