Motivation

Young children engage in forms of intuitive scientific discovery, building structured causal theories of their environment using many of the principles that underpin professional science [1, 2]. Despite progress in modeling many of these ideas, automatic discovery of models of realistic phenomena from observation and interaction remains largely out of reach.

Objective

Facilitate progress towards automatic scientific discovery, first by introducing a representation of causal models that is expressive enough to succinctly capture the complexities of real world phenomena, and second by presenting a corpus of domains and accompanying benchmark challenge.

Contribution

Causal Inductive Synthesis Corpus A manually constructed collection of interactive domains, which abstract core causal concepts present in real-world mechanisms and environments.

Autumn

The AUTumn language is a Turing-complete language for specifying causal probabilistic models. It allows succinct expression for models that vary dynamically through time, respond to external input, have internal state and memory, exhibit probabilistic non-determinism, and have complex causal dependencies between variables.

The Autumn Language

AUTumn is a language for probabilistic causal models. It is designed to express models that vary as a function of time or external input. Many constructs are standard; these include the use of \( \_ \) for variable assignment, and giving variables types by declaring them as.

```
and giving variables types by declaring them as
```

```
prev
```

```
the
```

Events Beyond \texttt{init next}, temporal events may also be specified using the construct \texttt{on}, with the pattern on \texttt{event intervention}. An \texttt{event} is any sequence of type \texttt{Bool}, and an \texttt{intervention} is a modification to a value.

Sequences

Values in an AUTumn program represent sequences that vary with time. A value \( \eta \) at time \( t \) may be (i) time invariant, i.e., \( \eta_t = \eta \) for some constant \( \eta \), (ii) stateless and time varying, i.e, \( \eta_t = f(t) \) for some function \( f \), or (iii) stateful / recurrent sequences defined in terms of previous values, i.e., \( \eta_t = f(\eta_{t-1}) \), \( \eta_t \) is defined using the \texttt{init next} construct, which defines an initial value for a variable and an update rule to be performed on subsequent time steps. Previous values of a variable can be accessed using the \texttt{prev} primitive; for example, \texttt{prev x} gives the value of \( x \) at time \( t-1 \).

Events Beyond \texttt{init next}, temporal events may also be specified using the construct \texttt{on}, with the pattern on \texttt{event intervention}. An \texttt{event} is any sequence of type \texttt{Bool}, and an \texttt{intervention} is a modification to a value.

## The Causal Inductive Synthesis Corpus

### Specification

Let \( \mathcal{D} \) denote the set of all AUTumn models. CISC is a dataset \( \mathcal{D} = \{ m_1, m_2, \ldots, m_N \} \) of \( N \) AUTumn models, i.e., \( m_i \in \mathcal{D} \). For each model \( m_i \in \mathcal{D} \), there is also a collection of \( M_{\text{train}}^{m_i} \) test trajectories \( T_1^{m_i}, \ldots, T_{M_{\text{train}}}^{m_i} \) and \( M_{\text{test}}^{m_i} \) train trajectories \( T_1^{m_i}, \ldots, T_{M_{\text{test}}}^{m_i} \). A trajectory is a pair \( (\mathcal{A}, \mathcal{O}) \), where \( \mathcal{A} \) and \( \mathcal{O} \) are finite sequences (of identical length) of actions and observations respectively. The action space \( \mathcal{A} \) allows for selecting a grid-cell, pressing an arrow, performing no action, or stopping a simulation. The observation space \( \mathcal{O} \) is a colored grid of cells.

### Motivation

Young children engage in forms of intuitive scientific discovery, building structured causal theories of their environment using many of the principles that underpin professional science [1, 2]. Despite progress in modeling many of these ideas, automatic discovery of models of realistic phenomena from observation and interaction remains largely out of reach.

### Objective

Facilitate progress towards automatic scientific discovery, first by introducing a representation of causal models that is expressive enough to succinctly capture the complexities of real world phenomena, and second by presenting a corpus of domains and accompanying benchmark challenge.

### Motivation

Young children engage in forms of intuitive scientific discovery, building structured causal theories of their environment using many of the principles that underpin professional science [1, 2]. Despite progress in modeling many of these ideas, automatic discovery of models of realistic phenomena from observation and interaction remains largely out of reach.

### Objective

Facilitate progress towards automatic scientific discovery, first by introducing a representation of causal models that is expressive enough to succinctly capture the complexities of real world phenomena, and second by presenting a corpus of domains and accompanying benchmark challenge.