Presented at The Twelfth Annual Conference on Advances in Cognitive Systems,
Atlanta, Georgia, USA, 13-15 October 2025

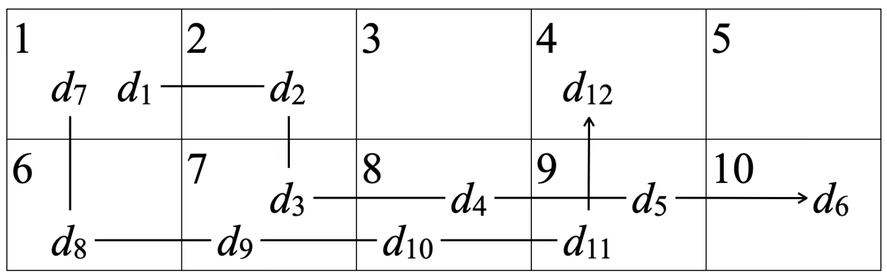
A Data-Transparent Probabilistic Model of Temporal Propositional Abstraction

Hiroyuki Kido
Cardiff University, Park Place, Cardiff, CF10 3AT, United Kingdom

ABSTRACT Standard probabilistic models face fundamental challenges such as data scarcity, a large hypothesis space, and poor data transparency. To address these challenges, we propose a novel probabilistic model of data-driven temporal propositional reasoning. Unlike conventional probabilistic models where data is a product of domain knowledge encoded in the probabilistic model, we explore the reverse direction where domain knowledge is a product of data encoded in the probabilistic model. This more data-driven perspective suggests no distinction between maximum likelihood parameter learning and temporal propositional reasoning. We show that our probabilistic model is equivalent to a highest-order, i.e., full-memory, Markov chain, and it can also be viewed as a hidden Markov model requiring no distinction between hidden and observable variables. We discuss that limits provide a natural and mathematically rigorous way to handle data scarcity, including the zero-frequency problem. We also discuss that a probability distribution over data generated by our probabilistic model helps data transparency by revealing influential data used in predictions. The reproducibility of this theoretical work is fully demonstrated by the included proofs.

MOTIVATING EXAMPLE

We want to predict the location of a robot moving around the building with ten rooms shown below.

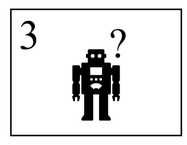


Only information we can use is the twelve time-series data, $d_1, d_2, \ldots, d_{12}$, collected by the robot.

Question
The robot moved through Rooms 2, 3, and 8.
Which room is the robot likely to be two time steps later — Room 4 or Room 10?

Existing solutions
Probabilistic models, e.g., Markov chains and hidden Markov models

FUNDAMENTAL PROBLEMS



✦ **Data scarcity, e.g., zero frequency**
No answer is straightforward as the robot has never been in Room 3.
Data smoothing works only when probabilistic models are simple enough.

✦ **Huge hypothesis spaces**
The three time-series data $d_{2\text{-}4}$ best match the given locations. This leads to the prediction of Room 10. This requires a 4th-order Markov chain with 9×10^4 parameters.

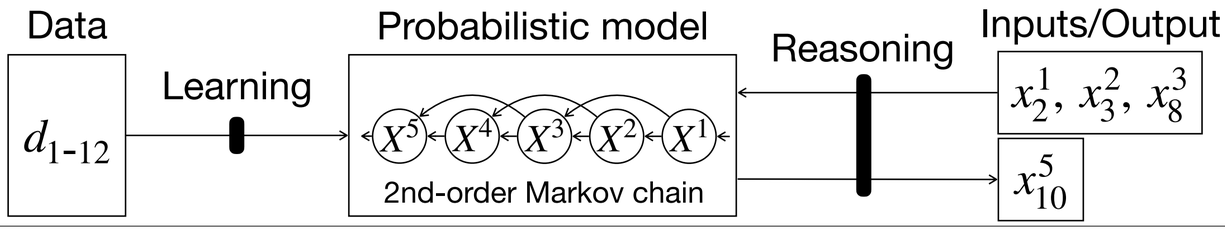
Room at Time t given Rooms for the past 4 time steps

$$p(X^t \mid X^{t-4:t-1}) = \begin{pmatrix} p(x_1^t \mid x_1^{t-4:t-1}) & \cdots & p(x_{10}^t \mid x_1^{t-4:t-1}) \\ & \vdots & \\ p(x_1^t \mid x_{10}^{t-4:t-1}) & \cdots & p(x_{10}^t \mid x_{10}^{t-4:t-1}) \end{pmatrix}$$

$10^4 \times 10$ matrix

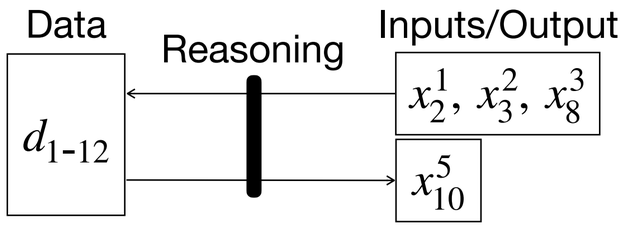
Room 10 at Time t given Room 1 for the past 4 time steps

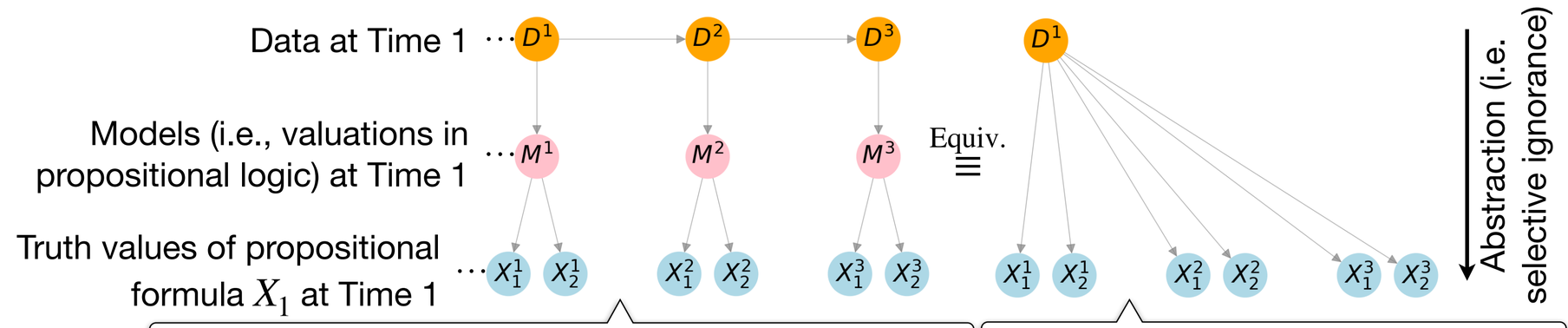
✦ **Poor data transparency**
The prediction of Room 10 cannot be grounded in $d_{2\text{-}4}$. Learning is the process of exploiting data to adjust the parameters of probabilistic models, whereas reasoning is the process of using the *parameters*, not the *data* itself, to make predictions.



OUR SOLUTION

Reasoning is the process of using *data* for predictions.





$$p(D^{1:T}, M^{1:T}, X_{1:I}^{1:T}) \overset{\text{Thm.1}}{=} \prod_{t=1}^{T} \left[p(D^t \mid D^{t-1}) p(M^t \mid D^t) \prod_{i=1}^{I} p(X_i^t \mid M^t) \right]$$

$$p(X_{1:I}^{1:T}) \overset{\text{Thm.2}}{=} \frac{1}{K} \sum_{k=1}^{K} \prod_{t=1}^{T} \prod_{i=1}^{I} p(X_i^t \mid m(n^{t-1}(d_k)))$$

Definition (Data distribution)
Let $n : Data \to Data$ be a function that maps each data point to the one at the next time step.

$$p(d^t \mid d^{1:t-1}, m^{1:t-1}, x_{1:I}^{1:t-1})$$

$$\overset{\text{Def.1}}{=} \begin{cases} \frac{1}{|Data|} & \text{if } t = 1 \\ 1 & \text{if } t \neq 1 \text{ and } d^t = n(d^{t-1}) \\ 0 & \text{otherwise} \end{cases}$$

$$\overset{\text{Prop.1}}{=} p(d^t \mid d^{t-1})$$

Namely, data changes deterministically.

Definition (Model distribution)
Let $m : Data \to Models$ be a function that maps each data point to the corresponding model, i.e., valuation in propositional logic.

$$p(m^t \mid d^{1:t}, m^{1:t-1}, x_{1:I}^{1:t-1})$$

$$\overset{\text{Def.2}}{=} \begin{cases} 1 \text{ if } m^t = m(d^t) \\ 0 \text{ otherwise} \end{cases} \overset{\text{Prop.2}}{=} p(m^t \mid d^t)$$

Namely, each data point supports a single model deterministically.

Definition (Knowledge distribution)
Let $[\![X_i]\!]_{m^t}$ be the truth value of X_i^t in the model m^t and $\mu \in [0.5, 1]$.

$$p(x_i^t \mid d^{1:t}, m^{1:t}, x_{1:I}^{1:t-1}, x_{1:i-1}^t)$$

$$\overset{\text{Def.3}}{=} \begin{cases} \mu & \text{if } x_i^t = [\![X_i]\!]_{m^t} \\ 1 - \mu & \text{otherwise} \end{cases} \overset{\text{Prop.3}}{=} p(x_i^t \mid m^t)$$

Namely, the truth values of formulas obey the semantics of propositional logic.

EVALUATIONS

A discrete-time, discrete-space localisation problem in a 7×7 grid. A robot retains the last ten visited locations, and it moves randomly while avoiding those in memory. When all adjacent accessible locations are stored, it remains in place.

Theoretical result
Given $\mu = 1$, our solution (TA: temporal abstraction) is equivalent to Markov chains (MC) with maximum likelihood parameters.

✦ **Data scarcity**
Given $\mu \neq 1$, TA outperforms MC with simple Laplace smoothing, which assigns an equal probability to zero frequency events.

✦ **Huge hypothesis spaces**
TA has essentially linear complexity due to the deterministic data trajectory and deterministic data–model support relation.

✦ **Poor data transparency**
Probability distributions over data ground predictions in data. In fact, the prediction of Room 5 is grounded in d_2 at Time 1.