

Graph Dreamer: Temporal Graph World Models for Sample-Efficient and Generalisable Reinforcement Learning

Keywords: Graph Neural Networks, Model-Based RL, Dynamics Models, Sample Efficiency, Generalisation, Transfer Learning, Physical Systems, HVAC Control

Many real-world control problems involve systems with inherent spatial structure and temporal dynamics, from thermal networks in buildings to mechanical linkages in robotics. However, most existing RL approaches handle temporal sequences (RNNs) and spatial relationships (GNNs) in isolation, failing to capture the coupled spatial-temporal evolution governing physical systems. Current model-based RL methods like DreamerV3 [1] learn from image sequences but cannot leverage structural relationships in graph-structured environments.

We introduce Graph Dreamer, the first world model architecture designed specifically for variable-size, heterogeneous graph environments. While existing world models operate on fixed-dimensional inputs like images, we develop a new approach that explicitly learns the inherent structural relationships governing environmental dynamics through latent graph representations, enabling generalisation across environments with different topologies and scales. Graph Dreamer uses a Graph Recurrent State-Space Model (Graph-RSSM) with a time-then-space update: a per-node temporal GRU followed by relation-aware message passing that propagates interactions across the graph. The model performs variational-inference with per-node stochastic latents (posterior and prior), uses an encoder with single-hop message passing and decoders to reconstruct node features. Reward and continuation signals are predicted from a pooled graph summary which is concatenated with exogenous inputs (e.g. weather variables). Actions are selected by a graph actor that produces node-wise actions with a single inter-actuator coordination pass. A distributional critic pools per-node value features. The actor and critic are trained entirely in imagination using Dreamer-style λ -returns and a slow target value network. This design results in permutation-equivariant dynamics, parameter sharing across node and edge types and scalability to variable number of actuators. We expect several practical benefits from the Graph Dreamer model, including increased sample efficiency compared to traditional RL (both model-free and model-based), improved robustness under partial observability and better generalisation to out-of-distribution environments. We plan to evaluate Graph Dreamer on several structural control tasks, such as heating, ventilation and air conditioning (HVAC) control in multi-zone buildings.

Graph Dreamer addresses a fundamental RL limitation: inability to leverage structural knowledge about system evolution. By unifying spatial and temporal modeling within imagination-based RL, our framework enables sample-efficient learning for graph-structured physical systems. This opens possibilities for smart buildings, robotics, network control and other structured control environments where sample efficiency is critical.

References: [1] Mastering Diverse Domains through World Models (Hafner et al., 2024)