Dynamic Feature Representation via Policy Attention for Dynamic Path Planning in Urban Road Networks

Kai Zhang[1,#], **Jingjing Gu**[1,*], **Qiuhong Wang**[1]

1 College of Computer Science and Technology, Nanjing, China

Poster presenter, * Corresponding author

zhangkainone@nuaa.edu.cn

ICLR 14th International Conference on Learning Representations

Nanjing University of Aeronautics and Astronautics

SPACE INTELLIGENT COMPUTING AND SYSTEM LABORATORY

1. Introduction

Backgrounds. Recent studies have shown that time-varying traffic conditions in urban road networks invalidate the pre-planned routes, making Dynamic Path Planning (DPP) a critical problem [5]. Deep Reinforcement Learning (DRL) provides a promising paradigm by learning adaptive routing policies without explicit traffic prediction [1, 2]. However, its effectiveness depends on how traffic dynamics are represented in the state [3].

Challenges. Existing methods 1) either encode global dynamics with high computational cost [4] or local observations that miss critical information , failing to capture decision-relevant features [6]; 2) do not identify which parts of the dynamic environment influence decisions, leading to redundancy, weakened Markov property, and degraded learning efficiency [7].

Contributions. 1) We propose a hierarchical Dynamics Feature Representation (DFR) framework that constructs a task-related subgraph by filtering global dynamics through a policy attention mechanism. 2) We introduce an n-hop neighborhood method to refine the subgraph into node-centric one, resulting in a decision-relevent state. 3) We demonstrate that DFR enables efficient and near-optimal policy learning by providing a low-dimensional, decision-sufficient state representation for DRL-based DPP.

2. Methodology

2.1 Problem Definition

The DPP problem is mapped into the finite-horizon and stochastic MDP model as a 6-tuple $\mathcal{M} = (\mathcal{S}, \mathcal{A}, T, R, \gamma, H)$:

- **State space** $\mathcal{S}$ denotes the set of all possible states, and a state $s_t \in \mathcal{S} = \{v^t, v_g, f_t\}$ consists of the current node v^t, the goal node $v_g = v_j$, and the *dynamics features* f_t.
- **Action space** $\mathcal{A} \subseteq \mathbb{R}^m$ is the set of possible actions taken by the agent, and the action $a_t = \{0, 1, \ldots, n_a\}$ is designed to move to one of n_a neighbors of v_a^t in the graph.
- **Transition function** $T : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \to [0, 1]$ defines the probability of transitioning from current state s_t to next state s_{t+1} after taking action a_t, that is, $T(s_{t+1}|s_t, a_t)$.
- **Reward function** $R : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \to \mathbb{R}$ guides the agent to achieve the goal by assigning a numerical reward to its each transition.

$$R(s_t, a_t, s_{t+1}) = -c(v^t, v^{t+1}; W_t) + b \cdot \mathbb{I}(v^{t+1} == v_g) \tag{1}$$

where $b \in \mathbb{R}^+$ is a constant reward for reaching the goal node, and $c(\cdot, \cdot)$ is the traffic cost of two nodes.

- **Discounted factor** $\gamma \in [0, 1]$ is used to balance between the future reward ($\gamma \to 1$) and immediate reward ($\gamma \to 0$).
- **Time horizon** $H \in \mathbb{N}^+$ is the maximum number of steps, that is, $t \leq H$. If the agent cannot reach the goal within H steps, the task ends in failure.

2.2 Dynamic Feature Representation

Given a DPP task $\tau = (v_s, v_g; \mathbf{W}_{:T})$, and the state $s_t = \{v^t, v_g; f_t\}$. DFR is a three-level process, refining the global information into a compact, decision-relevant state.

$$\mathbf{W}_{:T} \xmapsto{\tau,\Psi} \mathbf{W}'_{:T} \xmapsto{v^t,\Phi} \mathbf{W}''_{:T} \tag{2}$$

- **Global dynamics features** $\mathbf{W}$**.** The full edge weights W_t serve as the dynamics feature f_t, i.e., $s_t = (v^t, v_g; W_t)$, where v^t is the current node. W_t is high-dimensional and redundant.
- **Task-related key features** $\mathbf{W}'$**.** A key subset $W'_t(\tau) = \Psi(\tau, W_t)$, where Ψ is a function that extracts partial dynamics from W_t. Formally, for task τ, W'_t is sufficient if the optimal policy conditioned on it approximates the policy conditioned on W_t:

$$\pi^*(v^t, v_g; W'_t) \approx \pi^*(v^t, v_g; W_t) \tag{3}$$

- **Node-related local features** $\mathbf{W}''$**.** $W''_t(v^t) = \Phi(W'_t, v^t)$, where Φ is a function that maps W'_t to a lower-dimensional subset W''_t associated with the current node v^t.

$$\pi^*(v^t, v_g; W''_t) \approx \pi^*(v^t, v_g; W'_t) \tag{4}$$

- **Theoretical basis.** Predictive State Representations (PSR) provide a theoretical foundation [8]. PSR posits that the state of a system can be defined by predictions of future observable outcomes given possible action sequences, without resorting to latent variables. PSR ensures that the optimal policy based on W''_t approximates the policy based on $\mathbf{W}_{:T}$:

$$\pi^*(v^t, v_g; W''_t) \approx \pi^*(v^t, v_g; \mathbf{W}_{:T}) \tag{5}$$

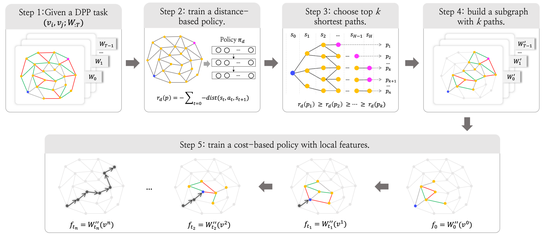

Figure 1: *The diagram of dynamics feature using policy attention and n-hop neighborhood method. The colors of the edges in the graph represent different dynamics.*

2.3 Dynamics Feature Extraction

2.3.1 Policy Attention for Ψ.

The policy attention leverages a distance-based optimal policy π_d^* to compute the shortest paths from v^t to v_g. The paths derived from π_d^* are ranked by length, and the top-k shortest paths are selected to form a subgraph $G' = (V', E')$, where the parameter k controls the trade-off between completeness and compactness: a smaller k may omit critical paths, whereas a larger k may introduce redundant information.

2.3.2 n-Hop Neighborhood for Φ

Let $\mathcal{N}^i(v^t)$ denote the i-th order neighbors of v^t, and V' and E' are the node set and egde set of G', respectively. The local node set of v^t is then defined as

$$V_l(v^t) = \mathcal{N}^0(v^t) \cup \mathcal{N}^1(v^t) \cup \cdots \cup \mathcal{N}^n(v^t) \cap V'$$

and the corresponding edges form the node-related subgraph. The weights of these edges define the local dynamics feature

$$f_t = W''_t(v^t) = \{w(v_i, v_j; t) \in E' \mid v_i, v_j \in V_l(v^t)\}$$

n determines the spatial scale of W''_t: smaller n captures highly localized dynamics but may overlook broader context, while larger n expands coverage but increases dimensionality and computational cost.

3. Results and Discussions

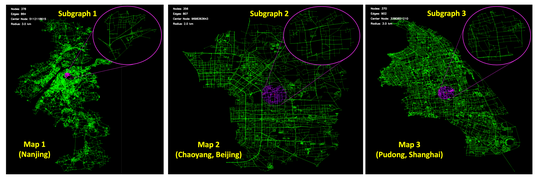

Figure 2: *Urban road network of three cities or districts. For each region, a center node and radius are specified, and this range is extracted to form a subgraph (magenta).*

Three baseline RL algorithms: **DQN** [9], **PPO** [10], and **GCN+DQN** [1], which are evaluated both with and without DFR. Four metrics: **Mean GAP**, **Success Rate (SR)** , **Compactness Rate (CR)**, and **Planning Time (PT)**.

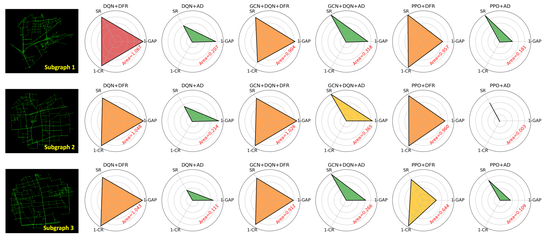

Figure 3: *Performance comparison of different algorithms across three regions. "DFR" stands for our framework, while "AD" refers to the use of All Dynamics features.*

Main results. As shown in Figure 3, our DFR framework achieves the fastest planning among RL baselines while maintaining high performance, by reducing feature dimensionality and the computational overhead for collecting dynamics features. The average path planning time is 8.18 ± 1.74 ms for DQN/PPO and 27.26 ± 6.8 ms for GCN+DQN, which are reduced by $85.59\%, 46.08\%, 79.32\%$, respectively.

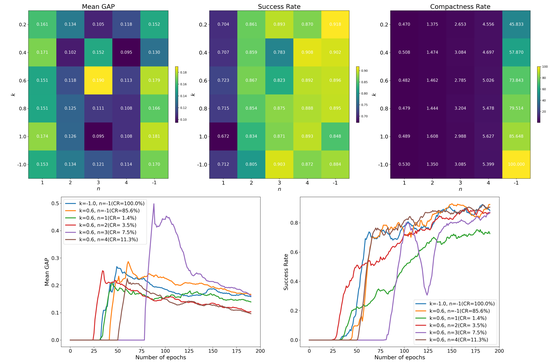

Figure 4: *Top: Performance headmaps of the models under various combinations of k and n. Bottom: Training curves under $k = 0.6$ with varying n.*

Ablation study. Compared with n, k has a more complex and less predictable impact on model performance, which poses greater challenges for parameter tuning. Therefore, based on these conclusions on small-scale subgraphs, it is recommend that in large-scale graph deployment, configurations with moderate k and smaller n should be preferred. n increases until the aggregation boundary of n is found, and k can be further explored.

References

[1] Kan Guo, et al. Contrastive learning for traffic flow forecasting based on multi graph convolution network. *IEEE TITS*, 18(2):290–301, 2024.

[2] Dang Viet Anh Nguyen, et al. Robustness of reinforcement learning-based traffic signal control under incidents: A comparative study. *arXiv preprint arXiv:2506.13836*, 2025.

[3] Chao Chen, et al. curl: A generic framework for bi-criteria optimum path-finding based on deep reinforcement learning. *IEEE TITS*, 24(2):1949–1961, 2023.

[4] Haiyang Liu, et al. Global-aware enhanced spatial-temporal graph recurrent networks: A new framework for traffic flow prediction. *arXiv preprint arXiv:2401.04135*, 2024.

[5] Runjia Du, et al. Dynamic urban traffic rerouting with fog-cloud reinforcement learning. *Computer-Aided Civil and Infrastructure Engineering*, 39(6):793–813, 2024a.

[6] Shengli Du, et al. Real-time local path planning strategy based on deep distributional reinforcement learning. *Neurocomputing*, 599: 128085, 2024b.

[7] Lawrence Francis, et al. Optimizing traffic signal control using high-dimensional state representation and efficient deep reinforcement learning. In *2025 IST-Africa Conference (IST-Africa)*, pp. 1–11. IEEE, 2025.

[8] Michael Littman and Richard S Sutton. Predictive representations of state. *NeurIPS*, 14, 2001.

[9] Volodymyr Mnih, et al. Human-level control through deep reinforcement learning. *nature*, 518(7540):529–533, 2015.

[10] John Schulman, et al. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.