Research Proposal: Constrained Conditional Generative Modeling with Diffusion Transformers for Sequential Decision-Making

1 Introduction

Reinforcement Learning (RL) has made substantial advancements in solving sequential decision-making problems. The Decision Transformer (DT) [2], introduced as a sequence modeling approach to RL, showed that transformer-based sequence models trained via supervised methods on trajectories could directly predict optimal actions. Complementary to this, recent work [1] demonstrates that conditional generative models, especially diffusion-based models, provide an effective alternative framework for sequential decision-making.

This proposal explores constrained generation within a diffusion transformer framework, leveraging the strengths of diffusion-based conditional generative models to enhance transformer-based RL agents. Specifically, we seek to impose explicit constraints during diffusion-based generation to satisfy safety, resource, or policy constraints frequently encountered in practical applications.

2 Motivation and Related Work

Traditional RL methods optimize expected returns through dynamic programming and value functions. The Decision Transformer [2] approaches RL as a conditional sequence modeling task, predicting actions conditioned on desired future returns, which simplifies the training process considerably. Recently, diffusion models [1] have shown impressive results in modeling complex conditional distributions, making them natural candidates for decision-making tasks.

By introducing constraints explicitly into diffusion-based generation, we can tailor action policies to adhere to safety-critical requirements, budget constraints, or other domain-specific conditions. This approach not only improves reliability but also enhances interpretability.

3 Mathematical Formulation

We model sequential decision-making using state-action-reward sequences:

$$\tau = \{(s_t, a_t, r_t)\}_{t=1}^{T}.$$

We assume trajectories are sampled from some behavior policy π_b. The diffusion transformer models the joint conditional distribution of action sequences given state and constraints via the denoising diffusion probabilistic model (DDPM):

$$p_\theta(a_{1:T}|s_{1:T}, c) = \int p_\theta(a_{1:T}^{(0:N)}|s_{1:T}, c) da_{1:T}^{(1:N)} \tag{1}$$

where c represents constraints, N is the diffusion step, and

$$p_\theta(a^{(n-1)}|a^{(n)}, s, c) = \mathcal{N}(a^{(n-1)}; \mu_\theta(a^{(n)}, s, c, n), \Sigma_\theta(a^{(n)}, s, c, n)). \tag{2}$$

Constraints can be expressed as inequality or equality conditions:

$$g_i(s_t, a_t) \leq 0, \quad h_j(s_t, a_t) = 0, \quad \forall t, i, j. \tag{3}$$

We incorporate constraints via penalty methods in the diffusion objective:

$$L(\theta) = \mathbb{E}_{\tau\sim\pi_b}\left[-\log p_\theta(a_{1:T}|s_{1:T}, c) + \lambda \sum_{t,i,j} \left(||\max(g_i(s_t, a_t), 0)||^2 + ||h_j(s_t, a_t)||^2\right)\right]. \tag{4}$$

4 Proposed Algorithm

We outline a generalized pseudocode algorithm for training and inference with constrained diffusion transformers.

Algorithm 1 Constrained Diffusion Transformer Training

Require: Dataset $D = \{\tau_i\}_{i=1}^{M}$ with trajectories $\tau_i = \{(s_t, a_t, r_t)\}_{t=1}^{T_i}$, constraint set c.

Initialize diffusion transformer parameters θ

repeat

  Sample trajectory batch $\tau \sim D$

  Sample diffusion timestep $n \sim \text{Uniform}(1, N)$ and noise $\epsilon \sim \mathcal{N}(0, I)$

  Compute noisy action $a^{(n)} = \sqrt{\alpha_n} a + \sqrt{1 - \alpha_n} \epsilon$

  Predict denoised action $\hat{a} = \mu_\theta(a^{(n)}, s, c, n)$

  Compute constraint penalty $\mathcal{C} = \sum_{i,j} \left(|| \max(g_i(s,a), 0)||^2 + ||h_j(s,a)||^2 \right)$

  Optimize diffusion loss with penalty:

$$L(\theta) = ||\hat{a} - a||^2 + \lambda \mathcal{C}$$

until convergence

Algorithm 2 Constrained Diffusion Transformer Inference

Require: Initial state sequence $s_{1:T}$, constraint set c

Sample initial noise $a_{1:T}^{(N)} \sim \mathcal{N}(0, I)$

for $n = N, \ldots, 1$ **do**

  Predict actions:

$$a_{1:T}^{(n-1)} = \mu_\theta(a_{1:T}^{(n)}, s_{1:T}, c, n)$$

  Project actions to satisfy constraints:

$$a_{1:T}^{(n-1)} = \arg\min_a \left(||a - a_{1:T}^{(n-1)}||^2 + \lambda' \sum_{i,j} \left(|| \max(g_i(s,a), 0)||^2 + ||h_j(s,a)||^2 \right) \right)$$

end for

Return $a_{1:T}^{(0)}$

5 Conclusion and Future Directions

This proposal lays the foundation for integrating constraint satisfaction explicitly into diffusion-based conditional generative modeling frameworks for decision-making tasks. Future research will explore applying this method to practical scenarios, assessing scalability, and evaluating performance against established baselines.

References

[1] Anurag Ajay, Yilun Du, Pulkit Agrawal, Joshua B. Tenenbaum, Chelsea Finn, Sergey Levine, and Igor Mordatch. Is conditional generative modeling all you need for decision-making? *arXiv preprint arXiv:2211.15657*, 2022.

[2] Lili Chen, Kevin Lu, Aravind Rajeswaran, Kimin Lee, Aditya Grover, Michael Laskin, Pieter Abbeel, Aravind Srinivas, and Igor Mordatch. Decision transformer: Reinforcement learning via sequence modeling. *Advances in Neural Information Processing Systems*, 34:15084–15097, 2021.