Learning to Scale Logits for Temperature-Conditional GFlowNets

We propose a simple technique to adjust the SoftMax logit temperature for temperature-conditional GFlowNets and demonstrate its useful applications in drug discovery.

Minsu Kim[1,2,*], Joohwan Ko[1,*], Taeyoung Yun[1,*] Dinghuai Zhang[2], Ling Pan[3]
Woo Chang Kim[1], Jinkyoo Park[1], Emmanuel Bengio[4], Yoshua Bengio[2]
[1] KAIST [2] Mila [3] HKUST [4] Valance Labs [*] Equal contribution

1. Preliminary

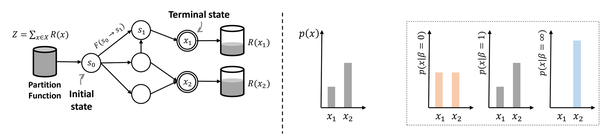


Illustration of GFlowNets: Flow Models on Directed Acyclic Graph

Generative Distribution of GFlowNets

Generative Distribution of Temperature Conditional GFlowNets

1.1 What are GFlowNets?

GFlowNets aim to convert a non-negative reward function $R(x)$ into a generative policy: $p(x) \propto R^{\beta}(x)$ with inverse temperature β which determine steepness of target distribution.

1.2 Temperature Conditional GFlowNets

Our aim is to learn the temperature-conditional GFlowNets (Temp-GFN):

$$p(x|\beta) \propto R^{\beta}(x)$$

The major benefit of temperature-conditional GFlowNets is the controllability of GFlowNets' exploration and exploitation through adjusting temperature.

2. Research Objective

Research Scope 1: Training stability of Temp-GFN

- Temp-GFN training is challenging due to varying target distribution steepness.
- Representing various temperature distributions can be numerically difficult and unstable.
- **Incorporating additional inductive bias into SoftMax temperature may stabilize Temp-GFN training.**

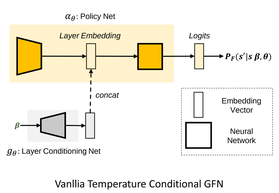
Research Scope 2: Usefulness of Temp-GFN

- Training GFN at low temperatures (steep distributions) is challenging.
- Generalizing to steep, high-reward regions is difficult with limited observations.
- **Training at various high temperatures (low β) and querying at low temperatures (high β) may generalize better than fixed low temperature training.**

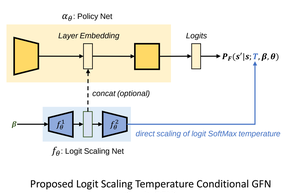
3. Methodology: Logit-GFN

3.1 Overall Idea

Introducing the Logit-Scaling Net: a novel auxiliary network that optimally scales the SoftMax temperature of policy network logits, enhancing Temp-GFN training and enabling effective generalization across various temperature distributions.



Vanllia Temperature Conditional GFN (Layer-GFN)



Proposed Logit Scaling Temperature Conditional GFN (Logit-GFN)

3.2 Logit Scaling

Objective: Facilitate training of temperature-conditional GFlowNets, $p(x|\beta) \propto R(x)^{\beta}$, over varying inverse temperatures β.

Logit-Scaling Trick: Uses a skip connection to adjust the SoftMax temperature T of P_F logits based on β. Defined as:

$$P_F(s'|s;\beta,\theta) := \frac{\exp\left(\alpha_\theta(s,s')/f_\theta(\beta)\right)}{\sum_{s'' \in \mathrm{Ch}(s)} \exp(\alpha_\theta(s,s'')/f_\theta(\beta))} \quad (1)$$

- $\alpha_\theta(s,s')$: Neural net independent of β
- $f_\theta(\beta)$: Logit-scaling net transforming β into SoftMax temperature T

Key Point: Logit-scaling is agnostic to the policy network's form, including layer-conditioning networks.

(Note that the logit-scaling can be applied to layer-conditioning networks (where α is also dependent on β), ensuring their **full expressive capacity** is maintained across different temperature regimes.)

3.3 Training Objective

Training Procedure. We minimize the trajectory balance (TB) loss with a replay buffer $\mathcal{D}$, but train GFlowNets over multiple $\beta \sim P_{\text{train}}(\beta)$:

$$\mathcal{L}(\theta;\mathcal{D}) = \mathbb{E}_{P_{\text{train}}(\beta)}\mathbb{E}_{P_{\mathcal{D}}(\tau)}\left[\left(\log \frac{Z_\theta(\beta)\prod_{t=1}^{n} P_F(\cdot)}{R(x)^{\beta}\prod_{t=1}^{n} P_B(\cdot)}\right)^2\right], \quad (2)$$

where $P_F(\cdot) = P_F(s_t|s_{t-1};\beta,\theta)$ and $P_B(\cdot) = P_B(s_{t-1}|s_t;\beta,\theta)$.

Implementation. We condition the partition function $Z_\theta(\beta)$ on β and use DNNs f_θ, Z_θ to map scalars, minimizing parameter overhead.

3.4 Online discovery algorithm with Logit-GFN

Objective. Discover diverse, high-reward candidates $x \in \mathcal{X}$ (e.g., molecules with high binding affinity) focusing on top rewards, diversity, and modes.

Method. Enhance GFlowNets exploration by sampling multiple β values, forming a diverse replay buffer $\mathcal{D}$:

$$\mathcal{D} \leftarrow \mathcal{D} \cup \{\tau_1, \ldots, \tau_M\} \quad (3)$$

$$\tau_1, \ldots, \tau_M \sim \int_{\beta} P_F(\tau|\beta) dP_{\text{exp}}(\beta). \quad (4)$$

Dynamic control policy $P_{\text{exp}}(\beta)$ varies exploration range.

Algorithm 1 Scientific Discovery with Temperature-Conditional GFlowNets

```
Set D ← ∅                                              ▷ Initialize dataset.
for t = 1, ..., T do                                   ▷ Training T rounds
    β_1, ..., β_M ~ P_exp(β)                           ▷ Sample temperatures from exploration query prior.
    for m = 1, ..., M do
        τ_m ~ P_F(τ|β = β_m; θ)                        ▷ Sample trajectories from Logit-GFN.
        D ← D ∪ {τ_m}
    end for
    for k = 1, ..., K do                               ▷ Training K epochs per each training rounds
        Use ADAM for gradually minimizing L(θ; D).
    end for
end for
Output: D
```

4. Experiments

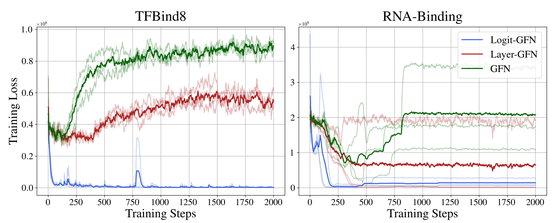
4.1 Tasks

We evaluate temperature-conditional GFlowNets on a toy grid world and four biochemical tasks: QM9, TFBind8, sEH, and RNA-binding.

- **Grid world**: The agent can move forward but not backward to reach a high-reward goal state. With multiple high-reward modes, the agent aims to cover all modes in the grid space.
- **QM9**: Generation of molecular graphs by sequentially adding atom components. The reward function is based on the Homo-Lumo gap.
- **TFBind8**: Generation of DNA sequences by bidirectional token addition. The reward function measures binding activity with human transcription factors.
- **sEH**: Generation of molecular graphs by sequentially adding predefined fragment components. The reward function is specified by enzymatic activity metrics.
- **RNA-binding**: Generation of RNA sequences by bidirectional token addition. The reward function is based on binding activity with human transcription factors.

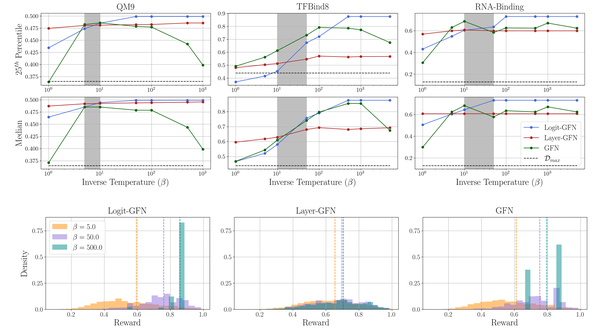
4.2 Training Stability

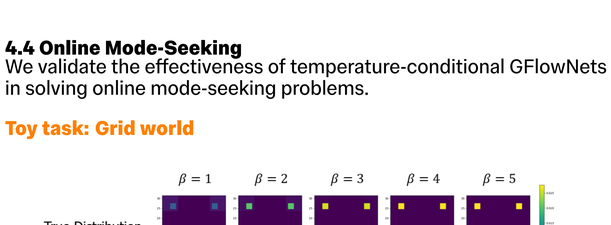
We assess the training stability of temperature-conditional GFlowNets.



4.3 Offline Generalization

We examine the controllability of temperature-conditional GFlowNets, i.e., $p(x|\beta) \propto R(x)^{\beta}$, in offline model-based optimization.

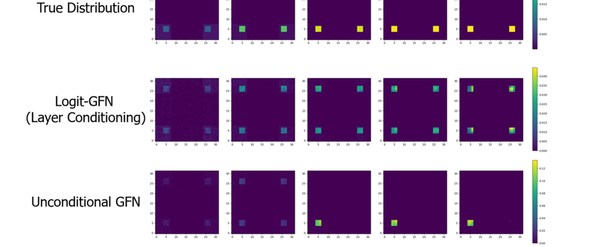




4.4 Online Mode-Seeking

We validate the effectiveness of temperature-conditional GFlowNets in solving online mode-seeking problems.

Toy task: Grid world



Biochemical discovery tasks