Stochastic Generative Flow Networks
(Supplementary material)

Ling Pan[*1,2] **Dinghuai Zhang**[*1,2] **Moksh Jain**[1,2] **Longbo Huang**[3] **Yoshua Bengio**[1,2,4]

[1]Mila - Québec AI Institute
[2]Université de Montréal
[3]Tsinghua University
[4]CIFAR AI Chair

A EXPERIMENTAL DETAILS

A.1 GRIDWORLD

The reward function for GridWorld is defined as in Eq. (1) following Bengio et al. [2021], where $R_0 = 2.0$, $R_1 = 0.5$, and $R_2 = 0.001$.

$$R(x) = R_0 + R_1 \prod_i \mathbb{I}\left(0.25 < |x_i/H - 0.5|\right) + R_2 \prod_i \mathbb{I}\left(0.3 < |x_i/H - 0.5| < 0.4\right) \tag{1}$$

We use a feedforward network that consists of two hidden layers with 256 hidden units and LeakyReLU activation. States are represented using one-hot embeddings. As for the environment model in Stochastic GFlowNet, it is also a feedforward layer consisting of two hidden layers with 256 hidden units and LeakyReLU activation. All models are trained for 20000 iterations, and we use a parallel of 16 rollouts in the environment at each iteration (which are then stored in the experience replay buffer). The GFlowNet model is updated based on the rollouts, and we train it based on the Adam [Kingma and Ba, 2015] optimizer using a learning rate of 0.001 (the learning rate for Z in TB is 0.1). We train the environment model using data sampled from the experience replay buffer with a batch size of 16, which is trained using the Adam optimizer with a learning rate of 0.0001. MCMC and PPO use the same configuration as in Bengio et al. [2021].

A.2 BIT SEQUENCES

We follow the same setup for the bit sequence generation task as in Malkin et al. [2022]. The GFlowNet model is a Transformer [Vaswani et al., 2017] that consists of 3 hidden layers with 64 hidden units and uses 8 attention heads. The exploration strategy is ϵ-greedy with $\epsilon = 0.0005$, while the sampling temperature is set to 1. It uses a reward exponent of 3. The learning rate for training the GFlowNet model is 5×10^{-3}, with a batch size of 16. As for the environment model in Stochastic GFlowNet, we use a feedforward network consisting of two hidden layers with 2048 hidden units and ReLU activation, which is trained using the Adam optimizer with a learning rate of 5×10^{-4}. It is trained using data sampled from the experience replay buffer with a batch size of 128. We train all models for 50000 iterations, using a parallel of 16 rollouts in the environment. MCMC, A2C, and SAC adopt the same configuration as in Malkin et al. [2022].

A.3 TFBIND-8

For the TFBind-8 generation task, we follow the same setup as in Jain et al. [2022]. The vocabulary consists of 4 nucleobases, and the trajectory length is 8. The GFlowNet model is a feedforward network that consists of 2 hidden layers with 2048 hidden units and ReLU activation. The exploration strategy is ϵ-greedy with $\epsilon = 0.001$, while the reward exponent is 3. The

[*]Equal contribution.

Accepted for the 39th Conference on Uncertainty in Artificial Intelligence (UAI 2023).

learning rate for training the GFlowNet model is 10^{-4}, with a batch size of 32. As for the environment model, we use a feedforward network consisting of two hidden layers with 2048 hidden units and ReLU activation, which is trained using the Adam optimizer with a learning rate of 10^{-5}. It is trained using data sampled from the experience replay buffer with a batch size of 16. We train all models for 5000 iterations. MCMC, A2C, and SAC baselines follow the same configuration as in Jain et al. [2022].

A.4 ANTIMICROBIAL PEPTIDE GENERATION

We follow the same setup for the antimicrobial peptide generation task as in Malkin et al. [2022]. The GFlowNet model is a Transformer [Vaswani et al., 2017] that consists of 3 hidden layers with 64 hidden units and uses 8 attention heads. The exploration strategy is ϵ-greedy with $\epsilon = 0.01$, while the sampling temperature is set to 1. It uses a reward exponent of 3. The learning rate for training the GFlowNet model is 0.001, with a batch size of 16. As for the environment model, we use a feedforward network consisting of two hidden layers with 128 hidden units and ReLU activation, which is trained using the Adam optimizer with a learning rate of 0.0005. It is trained using data sampled from the experience replay buffer with a batch size of 128. We train all models for 20000 iterations, using a parallel of 16 rollouts in the environment.

References

Emmanuel Bengio, Moksh Jain, Maksym Korablyov, Doina Precup, and Yoshua Bengio. Flow network based generative models for non-iterative diverse candidate generation. *Advances in Neural Information Processing Systems*, 34:27381–27394, 2021.

Moksh Jain, Emmanuel Bengio, Alex Hernandez-Garcia, Jarrid Rector-Brooks, Bonaventure F.P. Dossou, Chanakya Ekbote, Jie Fu, Tianyu Zhang, Micheal Kilgour, Dinghuai Zhang, Lena Simine, Payel Das, and Yoshua Bengio. Biological sequence design with GFlowNets. *International Conference on Machine Learning (ICML)*, 2022.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *International Conference on Learning Representations (ICLR)*, 2015.

Nikolay Malkin, Moksh Jain, Emmanuel Bengio, Chen Sun, and Yoshua Bengio. Trajectory balance: Improved credit assignment in GFlowNets. *Neural Information Processing Systems (NeurIPS)*, 2022.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Neural Information Processing Systems (NIPS)*, 2017.