Streaming Bayes GFlowNets

Tiago da Silva, Daniel Augusto de Souza, Diego Mesquita

Keywords — GFlowNets, Variational Bayesian inference

TL;DR
- we propose Streaming Bayes GFlowNets (SB-GFlowNets) as a general-purpose variational inference tool for streaming Bayesian inference in discrete spaces,
- in simple terms, SB-GFlowNet employs a GFlowNet as a surrogate prior to update the current posterior based on newly observed data, eliminating the need to process the entire dataset repeatedly,
- we introduce off-policy and on-policy algorithms for training SB-GFlowNets,
- we demonstrate that SB-GFlowNets are susceptible to catastrophic error propagation and discuss potential workarounds,
- we empirically verify that SB-GFlowNets can drastically reduce the training time of a GFlowNet in a streaming Bayes setting,

I. Background: GFlowNets

GFlowNets are amortized algorithms for sampling from distributions over compositional objects, i.e., over objects that can be sequentially constructed from an initial state through the application of simple actions (e.g., graphs via edge-addition).

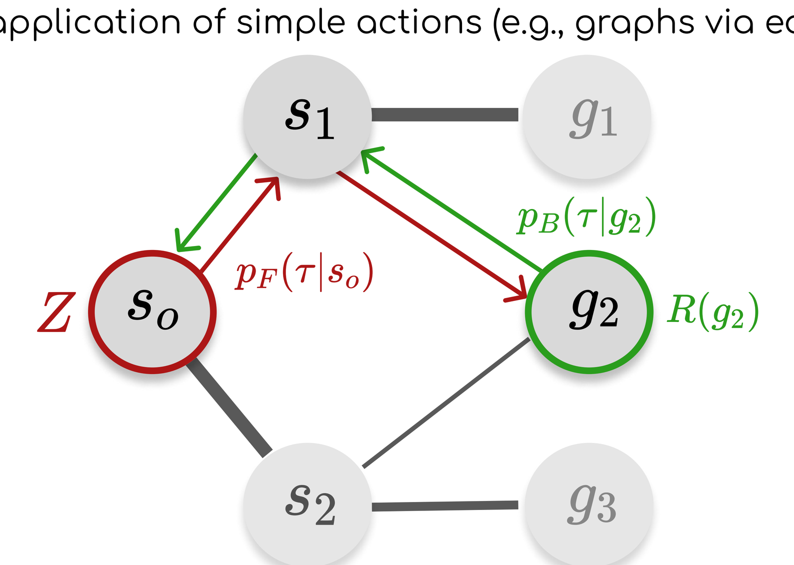

Figure 1: An illustration of the GFlowNet's state graph as a DAG on $\mathcal{S}$.

In a nutshell, a GFlowNet is composed of two three ingredients.
1. An extension $\mathcal{S}$ of the target distribution support's $\mathcal{X}$.
2. A measurable pointed DAG $\mathcal{G}$ on $\mathcal{S}$ dictating how the states in $\mathcal{S}$ are connected to one another. We refer to $\mathcal{G}$ as the state graph.
3. A forward and backward policies defining the stochastic transitions within $\mathcal{G}$.

Our objective is to learn a forward $p_F(\tau)$ and a backward $p_B(\tau|x)$ policies such that the marginal of $p_F(\tau)$ over $\mathcal{X}$ matches a given unnormalized density $r : \mathcal{X} \to \mathbb{R}_+$.

$$p_F(\tau) = \prod_{(s,s')\in\tau} p_F(s' \mid s) \text{ and } \int_{\mathcal{T}} 1_{\tau \rightsquigarrow x} p_F(\tau) \mathrm{d}\tau = r(x); \qquad (1)$$

$\mathcal{T}$ denotes the space of trajectories in $\mathcal{G}$ and $\tau \rightsquigarrow$, the event in which τ finishes on $x \in \mathcal{X}$.

TL;DR
- we propose Streaming Bayes GFlowNets (SB-GFlowNets) as a general-purpose variational inference tool for streaming Bayesian inference in discrete spaces,
- in simple terms, SB-GFlowNet employs a GFlowNet as a surrogate prior to update the current posterior based on newly observed data, eliminating the need to process the entire dataset repeatedly,
- we introduce off-policy and on-policy algorithms for training SB-GFlowNets,
- we demonstrate that SB-GFlowNets are susceptible to catastrophic error propagation and discuss potential workarounds,
- we empirically verify that SB-GFlowNets can drastically reduce the training time of a GFlowNet in a streaming Bayes setting,

 {tiago.henrique, diego.mesquita}@fgv.br, daniel.souza.21@ucl.ac.uk