GFlowNets embaraçosamente paralelas

Tiago da Silva, Luiz Max, Amauri Souza, Sami Kaski, Diego Mesquita

18 de novembro de 2024

Escola de Matemática Aplicada (EMAp)

Table of contents

Inferência Bayesiana aproximada com GFlowNets

Inferênia Bayesiana em espaços discretos

Seja $\mathcal{G} = \{G_1, \ldots, G_S\}$ um espaço discreto muito grande.

Exemplo: $\mathcal{G} =$ espaço de todos os grafos com 25 nós, $|\mathcal{G}| \approx 10^{90}$.

Seja $\mathbf{X} \sim f(\cdot|G)$ amostra extraída de distribuição indexada por $G \in \mathcal{G}$.

Seja $\pi\colon \mathcal{G} \to \mathbb{R}_+$ uma distribuição a priori sobre $\mathcal{G}$.

Objetivo

Estimar a posteriori sobre $\mathcal{G}$. Pela regra de Bayes,

$$\pi(G|\mathbf{X}) = \frac{\pi(G) f(\mathbf{X}|G)}{p(\mathbf{X})},$$

em que $p(\mathbf{X}) = \sum_{g \in \mathcal{G}} \pi(G) f(\mathbf{X}|G)$.

Por quê? Queremos estimar momentos de função m sobre G,

$$\mathbb{E}_{G \sim \pi(\cdot|\mathbf{X})}[m(G)] \tag{1}$$

Aplicações: descoberta de drogas, inferência causal, filogenética.

Inferênia Bayesiana em espaços discretos

Seja $\mathcal{G} = \{G_1, \ldots, G_S\}$ um espaço discreto muito grande.

Exemplo: $\mathcal{G} =$ espaço de todos os grafos com 25 nós, $|\mathcal{G}| \approx 10^{90}$.

Seja $\mathbf{X} \sim f(\cdot|G)$ amostra extraída de distribuição indexada por $G \in \mathcal{G}$.

Seja $\pi\colon \mathcal{G} \to \mathbb{R}_+$ uma distribuição a priori sobre $\mathcal{G}$.

Objetivo

Estimar a posteriori sobre $\mathcal{G}$. Pela regra de Bayes,

$$\pi(G|\mathbf{X}) = \frac{\pi(G)f(\mathbf{X}|G)}{p(\mathbf{X})},$$

em que $p(\mathbf{X}) = \sum_{g \in \mathcal{G}} \pi(G)f(\mathbf{X}|G)$.

Por quê? Queremos estimar momentos de função m sobre G,

$$\mathbb{E}_{G \sim \pi(\cdot|\mathbf{X})}\left[m(G)\right] \tag{1}$$

Aplicações: descoberta de drogas, inferência causal, filogenética.

Inferênia Bayesiana em espaços discretos

Seja $\mathcal{G} = \{G_1, \ldots, G_S\}$ um espaço discreto muito grande.

Exemplo: $\mathcal{G} =$ espaço de todos os grafos com 25 nós, $|\mathcal{G}| \approx 10^{90}$.

Seja $\mathbf{X} \sim f(\cdot|G)$ amostra extraída de distribuição indexada por $G \in \mathcal{G}$.

Seja $\pi \colon \mathcal{G} \to \mathbb{R}_+$ uma distribuição a priori sobre $\mathcal{G}$.

Objetivo

Estimar a posteriori sobre $\mathcal{G}$. Pela regra de Bayes,

$$\pi(G|\mathbf{X}) = \frac{\pi(G) f(\mathbf{X}|G)}{p(\mathbf{X})},$$

em que $p(\mathbf{X}) = \sum_{g \in \mathcal{G}} \pi(G) f(\mathbf{X}|G)$.

Por quê? Queremos estimar momentos de função m sobre G,

$$\mathbb{E}_{G \sim \pi(\cdot|\mathbf{X})}\left[m(G)\right] \tag{1}$$

Aplicações: descoberta de drogas, inferência causal, filogenética.

Inferênia Bayesiana em espaços discretos

Seja $\mathcal{G} = \{G_1, \ldots, G_S\}$ um espaço discreto muito grande.

Exemplo: $\mathcal{G} =$ espaço de todos os grafos com 25 nós, $|\mathcal{G}| \approx 10^{90}$.

Seja $\mathbf{X} \sim f(\cdot|G)$ amostra extraída de distribuição indexada por $G \in \mathcal{G}$.

Seja $\pi \colon \mathcal{G} \to \mathbb{R}_+$ uma distribuição a priori sobre $\mathcal{G}$.

Objetivo

Estimar a posteriori sobre $\mathcal{G}$. Pela regra de Bayes,

$$\pi(G|\mathbf{X}) = \frac{\pi(G) f(\mathbf{X}|G)}{p(\mathbf{X})},$$

em que $p(\mathbf{X}) = \sum_{g \in \mathcal{G}} \pi(G) f(\mathbf{X}|G)$.

Por quê? Queremos estimar momentos de função m sobre G,

$$\mathbb{E}_{G \sim \pi(\cdot|\mathbf{X})}\left[m(G)\right] \tag{1}$$

Aplicações: descoberta de drogas, inferência causal, filogenética.

Inferênia Bayesiana em espaços discretos

Seja $\mathcal{G} = \{G_1, \ldots, G_S\}$ um espaço discreto muito grande.

Exemplo: $\mathcal{G} =$ espaço de todos os grafos com 25 nós, $|\mathcal{G}| \approx 10^{90}$.

Seja $\mathbf{X} \sim f(\cdot|G)$ amostra extraída de distribuição indexada por $G \in \mathcal{G}$.

Seja $\pi\colon \mathcal{G} \to \mathbb{R}_+$ uma distribuição a priori sobre $\mathcal{G}$.

Objetivo

Estimar a posteriori sobre $\mathcal{G}$. Pela regra de Bayes,

$$\pi(G|\mathbf{X}) = \frac{\pi(G)f(\mathbf{X}|G)}{p(\mathbf{X})},$$

em que $p(\mathbf{X}) = \sum_{g\in\mathcal{G}} \pi(G)f(\mathbf{X}|G)$.

Por quê? Queremos estimar momentos de função m sobre G,

$$\mathbb{E}_{G\sim\pi(\cdot|\mathbf{X})}\left[m(G)\right] \tag{1}$$

Aplicações: descoberta de drogas, inferência causal, filogenética.

Inferência Bayesiana em espaços discretos

Objetivo

Estimar a posteriori sobre $\mathcal{G}$. Pela regra de Bayes,

$$\pi(G|\mathbf{X}) = \frac{\pi(G)f(\mathbf{X}|G)}{p(\mathbf{X})},$$

em que $p(\mathbf{X}) = \sum_{g\in\mathcal{G}} \pi(G)f(\mathbf{X}|G)$.

Desafio: Não podemos calcular a soma em $p(\mathbf{X})$ porque há muitos elementos em $\mathcal{G}$ — precisaríamos somar por muito tempo.

Solução: Gerar amostras $(g_1, \ldots, g_T)$ (aproximadas) de $\pi(\cdot|\mathbf{X})$ e utilizar um estimador de Monte Carlo p/ a quantidade

$$\mathbb{E}_{G\sim\pi(\cdot|\mathbf{X})}[m(G)] \approx \frac{1}{T} \sum_{1\leq t\leq T} m(g_t).$$

Inferência Bayesiana em espaços discretos

Objetivo

Estimar a posteriori sobre $\mathcal{G}$. Pela regra de Bayes,

$$\pi(G|\mathbf{X}) = \frac{\pi(G)f(\mathbf{X}|G)}{p(\mathbf{X})},$$

em que $p(\mathbf{X}) = \sum_{g \in \mathcal{G}} \pi(G)f(\mathbf{X}|G)$.

Desafio: Não podemos calcular a soma em $p(\mathbf{X})$ porque há muitos elementos em $\mathcal{G}$ — precisaríamos somar por muito tempo.

Solução: Gerar amostras $(g_1, \ldots, g_T)$ (aproximadas) de $\pi(\cdot|\mathbf{X})$ e utilizar um estimador de Monte Carlo p/ a quantidade

$$\mathbb{E}_{G \sim \pi(\cdot|\mathbf{X})}[m(G)] \approx \frac{1}{T} \sum_{1 \leq t \leq T} m(g_t).$$

Generative flow networks (GFlowNets): Visão geral

O objetivo agora é **aprender** um kernel Markoviano K_θ p/ ter **garantias não assintóticas** sobre a distribuição marginal de $(G^{(t)})_t$.

Método:

1. Resolvemos um problema de otimização, $\theta^* = \arg\min_\theta \mathcal{L}(\theta)$, p/ alguma função de perda $\mathcal{L}$
2. Escolhemos $G^{(o)} \in \mathcal{G}$ e construímos uma cadeia de Markov $(G^{(t)})_t$ com $G^{(t)} | G^{(t-1)} \sim K_{\theta^*}(G^{(t-1)}, \cdot)$

Garantias não assintóticas: Existe $T < \infty$ (relativamente pequeno e conhecido) tal que $\mathbb{P}[G^{(T)} = g] = \pi(g | \mathbf{X})$.

Generative flow networks (GFlowNets): Visão geral

O objetivo agora é **aprender** um kernel Markoviano K_θ p/ ter **garantias não assintóticas** sobre a distribuição marginal de $(G^{(t)})_t$.

Método:

1. Resolvemos um problema de otimização, $\theta^* = \arg\min_\theta \mathcal{L}(\theta)$, p/ alguma função de perda $\mathcal{L}$
2. Escolhemos $G^{(o)} \in \mathcal{G}$ e construímos uma cadeia de Markov $(G^{(t)})_t$ com $G^{(t)} | G^{(t-1)} \sim K_{\theta^*}(G^{(t-1)}, \cdot)$

Garantias não assintóticas: Existe $T < \infty$ (relativamente pequeno e conhecido) tal que $\mathbb{P}[G^{(T)} = g] = \pi(g|\mathbf{X})$.

Problema ilustrativo

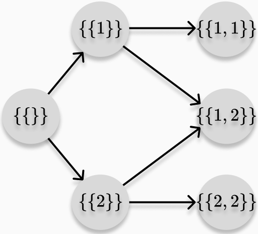
Seja $\mathcal{G}$ o conjunto de multiconjuntos de tamanho m com elementos extraídos de $\mathcal{D} = \{1, \ldots, d\}$. Como amostrar elementos de $\mathcal{G}$ proporcionalmente a uma função positiva $R \colon \mathcal{G} \to \mathbb{R}_+$?

GFlowNets: Detalhes

Problema ilustrativo

Seja $\mathcal{G}$ o conjunto de multiconjuntos de tamanho m com elementos extraídos de $\mathcal{D} = \{1, \ldots, d\}$. Como amostrar elementos de $\mathcal{G}$ proporcionalmente a uma função positiva $R \colon \mathcal{G} \to \mathbb{R}_+$?

Vamos definir *grafo de estado* — acíclio e direcionado — sobre uma extensão $\mathcal{S}$ do suporte original $\mathcal{G}$ ($\mathcal{S}$ = multiconjuntos de tamanho $\leq m$).

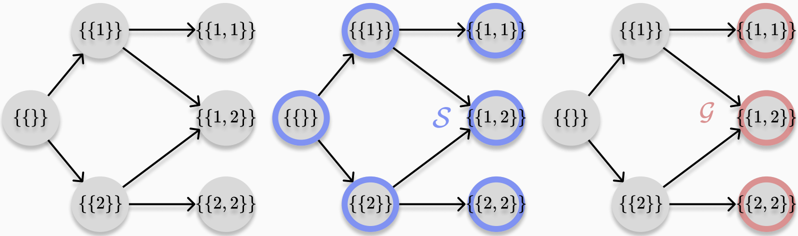


GFlowNets: Detalhes

Problema ilustrativo

Seja $\mathcal{G}$ o conjunto de multiconjuntos de tamanho m com elementos extraídos de $\mathcal{D} = \{1, \ldots, d\}$. Como amostrar elementos de $\mathcal{G}$ proporcionalmente a uma função positiva $R \colon \mathcal{G} \to \mathbb{R}_+$?

Vamos definir *grafo de estado* — acíclio e direcionado — sobre uma extensão $\mathcal{S}$ do suporte original $\mathcal{G}$ ($\mathcal{S}$ = multiconjuntos de tamanho $\leq m$).

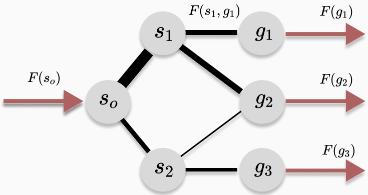


GFlowNets: Detalhes

Problema ilustrativo

Seja $\mathcal{G}$ o conjunto de multiconjuntos de tamanho m com elementos extraídos de $\mathcal{D} = \{1, \ldots, d\}$. Como amostrar elementos de $\mathcal{G}$ proporcionalmente a uma função positiva $R \colon \mathcal{G} \to \mathbb{R}_+$?

Vamos interpretar a rede construída como uma **rede de fluxo**.



Queremos que,

- nos estados finais $g \in \mathcal{G}$, $\sum_{s' \in \text{Pais}(g)} F(s', g) = R(g)$;
- nos estados intermediários $s \in \mathcal{S} \setminus \mathcal{G} \cup \{s_o\}$, balanço do fluxo:

$$\sum_{s' \in \text{Filhos}(s)} F(s, s') = \sum_{s' \in \text{Pais}(s)} F(s', s)$$

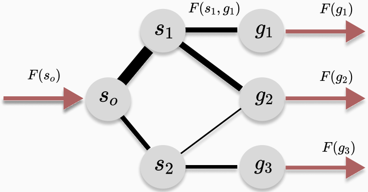
Se navegarmos esta rede conforme seu fluxo — partindo de s_o — chegaremos a cada $g \in \mathcal{G}$ com frequência proporcional a $R(g)$.

GFlowNets: Detalhes

Problema ilustrativo

Seja $\mathcal{G}$ o conjunto de multiconjuntos de tamanho m com elementos extraídos de $\mathcal{D} = \{1, \dots, d\}$. Como amostrar elementos de $\mathcal{G}$ proporcionalmente a uma função positiva $R \colon \mathcal{G} \to \mathbb{R}_+$?

Vamos interpretar a rede construída como uma **rede de fluxo**.



Queremos que,

- nos estados finais $g \in \mathcal{G}$, $\sum_{s' \in \text{Pais}(g)} F(s', g) = R(g)$;
- nos estados intermediários $s \in \mathcal{S} \setminus \mathcal{G} \cup \{s_o\}$, balanço do fluxo:

$$\sum_{s' \in \text{Filhos}(s)} F(s, s') = \sum_{s' \in \text{Pais}(s)} F(s', s)$$

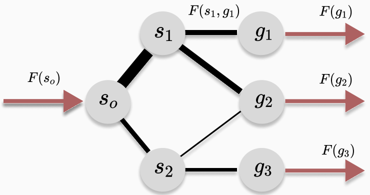
Se navegarmos esta rede conforme seu fluxo — partindo de s_o — chegaremos a cada $g \in \mathcal{G}$ com frequência proporcional a $R(g)$.

GFlowNets: Detalhes

Problema ilustrativo

Seja $\mathcal{G}$ o conjunto de multiconjuntos de tamanho m com elementos extraídos de $\mathcal{D} = \{1, \ldots, d\}$. Como amostrar elementos de $\mathcal{G}$ proporcionalmente a uma função positiva $R\colon \mathcal{G} \to \mathbb{R}_+$?

Vamos interpretar a rede construída como uma **rede de fluxo**.

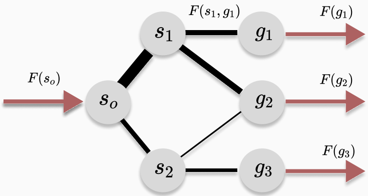


Queremos que,

- nos estados finais $g \in \mathcal{G}$, $\sum_{s' \in \text{Pais}(g)} F(s', g) = R(g)$;
- nos estados intermediários $s \in \mathcal{S} \setminus \mathcal{G} \cup \{s_o\}$, balanço do fluxo:

$$\sum_{s' \in \text{Filhos}(s)} F(s, s') = \sum_{s' \in \text{Pais}(s)} F(s', s)$$

Se navegarmos esta rede conforme seu fluxo — partindo de s_o — chegaremos a cada $g \in \mathcal{G}$ com frequência proporcional a $R(g)$.



Como estimar o fluxo F?

Nós definimos

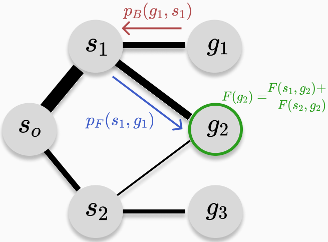
$$p_F(s, s') = \frac{F(s, s')}{\sum_{s''} F(s, s'')}, \qquad p_B(s', s) = \frac{F(s, s')}{\sum_{s''} F(s'', s')}, \tag{2}$$

e parameterizamos p_F e p_B com redes neurais (e.g., MLP).

Hoje de manhã, mostramos que a condição

$$\frac{p_F(\tau|s_o)}{p_B(\tau|G)R(G)} = \frac{p_F(\tau'|s_o)}{p_B(\tau'|G')R(G')},$$

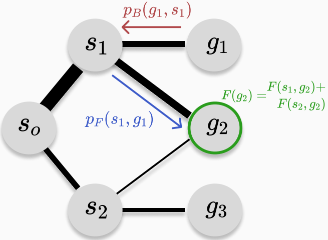
é necessária e suficiente para resolver este problema. Nós a chamamos de *condição de balanço contrastiva* (CB).



GFlowNets: Sumário

$$\frac{p_F(\tau|s_o)}{p_B(\tau|G)R(G)} = \frac{p_F(\tau'|s_o)}{p_B(\tau'|G')R(G')},$$

Para forçar a condição acima, resolvemos o problema de otimização estocástico (usando, e.g., Adam)



$$p_F^\star, p_B^\star = \underset{p_F, p_B}{\arg\min} \underset{\tau, \tau' \sim \xi}{\mathbb{E}} \left[\mathcal{L}_{CB}(\tau, \tau'; p_F, p_B, Z)\right],$$

em que

$$\mathcal{L}_{CB}(\tau, \tau'; p_F, p_B) = \left(\frac{p_F(\tau|s_o)}{p_B(\tau|G)R(G)} - \frac{p_F(\tau'|s_o)}{p_B(\tau'|G')R(G')}\right)^2.$$

Durante inferência, partimos de s_o e amostramos estados conforme $p_F^\star(\cdot|s_o)$ até chegarmos a algum $g \in \mathcal{G}$.

Inferência Bayesiana paralela

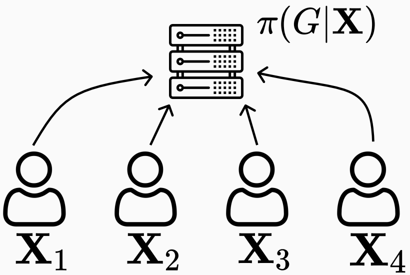
Cálculo distribuído da posteriori

Problema de inferência distribuída embaraçosa

Calcular a posteriori sobre um conjunto de dados particionado $\mathbf{X} = \{\mathbf{X}_1, \ldots, \mathbf{X}_N\}$ por razões de privacidade (federado) ou de eficiência,

$$\pi(G|\mathbf{X}) = \frac{1}{p(\mathbf{X})} \cdot \prod_{1 \leq n \leq N} f(\mathbf{X}_n|G)\pi(G)^{\frac{1}{N}}, \tag{3}$$

em uma **única etapa de comunicação**.

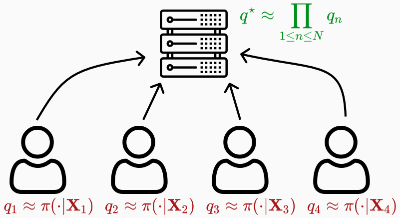


Problema de inferência distribuída embaraçosa

Calcular a posteriori sobre um conjunto de dados particionado $\mathbf{X} = \{\mathbf{X}_1, \ldots, \mathbf{X}_N\}$ por razões de privacidade (federado) ou de eficiência,

$$\pi(G|\mathbf{X}) = \frac{1}{p(\mathbf{X})} \cdot \prod_{1 \leq n \leq N} f(\mathbf{X}_n|G)\pi(G)^{\frac{1}{N}}, \tag{4}$$

em uma **única etapa de comunicação**.



Definimos a **subposteriori**

$$\pi(G|\mathbf{X}_n) \propto f(\mathbf{X}_n|G)\pi(G)^{\frac{1}{N}} \tag{5}$$

e consideramos uma **aproximação local** q_n de $\pi(G|\mathbf{X}_n)$.

Estimamos então uma **aproximação global** $q^\star$ de $\pi(\cdot|\mathbf{X})$ a partir de q_n's:

$$q^\star \propto \prod_{1 \leq n \leq N} q_n. \tag{6}$$

Exemplo: q_n é uma aproximação variacional de $\pi(G|\mathbf{X}_n)$ e $q^\star$ usa SGLD para amostrar de $\prod_n q_n$ com base em $\sum_n \nabla \log q_n$ (se $\nabla \log q_n$ existir).

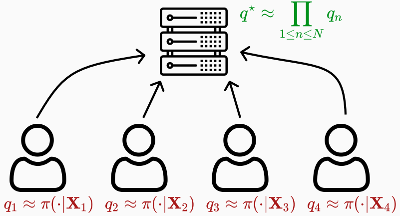
Inferência distribuída aproximada

Problema de inferência distribuída embaraçosa

Calcular a posteriori sobre um conjunto de dados particionado $\mathbf{X} = \{\mathbf{X}_1, \ldots, \mathbf{X}_N\}$ por razões de privacidade (federado) ou de eficiência,

$$\pi(G|\mathbf{X}) = \frac{1}{p(\mathbf{X})} \cdot \prod_{1 \le n \le N} f(\mathbf{X}_n|G)\pi(G)^{\frac{1}{N}}, \tag{4}$$

em uma **única etapa de comunicação**.



Definimos a **subposteriori**

$$\pi(G|\mathbf{X}_n) \propto f(\mathbf{X}_n|G)\pi(G)^{\frac{1}{N}} \tag{5}$$

e consideramos uma **aproximação local** q_n de $\pi(G|\mathbf{X}_n)$.

Estimamos então uma **aproximação global** $q^\star$ de $\pi(\cdot|\mathbf{X})$ a partir de q_n's:

$$q^\star \propto \prod_{1\le n\le N} q_n. \tag{6}$$

Exemplo: q_n é uma aproximação variacional de $\pi(G|\mathbf{X}_n)$ e $q^\star$ usa SGLD para amostrar de $\prod_n q_n$ com base em $\sum_n \nabla \log q_n$ (se $\nabla \log q_n$ existir).

GFlowNets embaraçosamente paralelas

GFlowNets: produto de distribuições

GFlowNets paralelas

Sejam $R_1, \ldots, R_N : \mathcal{G} \to \mathbb{R}_+$ funções não negativas. Seja $M_n = (p_F^{(n)}, p_B^{(n)})$ a GFlowNet treinada sobre R_n para $1 \leq n \leq N$. Queremos aprender uma GFlowNet $M^\star = (p_F^\star, p_B^\star)$ para aproximar

$$p_T^\star(G) = \sum_{\tau \rightsquigarrow G} p_F^\star(\tau | s_o) \propto \prod_{1 \leq n \leq N} R_n(G) \tag{7}$$

utilizando apenas $M_1, \ldots, M_N$.

No contexto de inferência Bayesiana paralela, $R_n(G) = f(\mathbf{X}_n | G)\pi(G)$ corresponde à **subposteriori não normalizada**.

GFlowNets: produto de distribuições

GFlowNets paralelas

Sejam $R_1, \ldots, R_N : \mathcal{G} \to \mathbb{R}_+$ funções não negativas. Seja $M_n = (p_F^{(n)}, p_B^{(n)})$ a GFlowNet treinada sobre R_n para $1 \leq n \leq N$. Queremos aprender uma GFlowNet $M^\star = (p_F^\star, p_B^\star)$ para aproximar

$$p_T^\star(G) = \sum_{\tau \rightsquigarrow G} p_F^\star(\tau | s_o) \propto \prod_{1 \leq n \leq N} R_n(G) \tag{7}$$

utilizando apenas $M_1, \ldots, M_N$.

No contexto de inferência Bayesiana paralela, $R_n(G) = f(\mathbf{X}_n | G)\pi(G)$ corresponde à **subposteriori não normalizada**.

GFlowNets: produto de distribuições

$M^\star$ terá que satisfazer a

$$\frac{p_F(\tau|s_o)}{p_B(\tau|G)\prod_{1\leq n\leq N} R_n(G)} = \frac{p_F(\tau'|s_o)}{p_B(\tau'|G')\prod_{1\leq n\leq N} R_n(G')}, \tag{8}$$

que não pode ser avaliada no servidor por impossibilidade de acessar R_n.

Contudo, sabemos que cada M_n satisfaz

$$\frac{p_F^{(n)}(\tau|s_o)}{p_B^{(n)}(\tau|G)R_n(G)} = \frac{p_F^{(n)}(\tau'|s_o)}{p_B^{(n)}(\tau'|G')R_n(G')}, \tag{9}$$

aproximadamente.

Substitutindo $\frac{R_n(G')}{R_n(G)}$ da Equação (8) na Equação (9), obtemos

$$\frac{p_F(\tau|s_o)}{p_B(\tau|G)} = \frac{p_F(\tau'|s_o)}{p_B(\tau'|G')} \prod_{1\leq n\leq N} \frac{p_F^{(n)}(\tau|s_o)}{p_B^{(n)}(\tau|G)R_n(G)} \cdot \frac{p_B^{(n)}(\tau'|G')}{p_F^{(n)}(\tau'|s_o)},.$$

Esta é a **condição de balanço agregado**.

GFlowNets: produto de distribuições

$M^\star$ terá que satisfazer a

$$\frac{p_F(\tau|s_o)}{p_B(\tau|G)\prod_{1\leq n\leq N} R_n(G)} = \frac{p_F(\tau'|s_o)}{p_B(\tau'|G')\prod_{1\leq n\leq N} R_n(G')}, \tag{8}$$

que não pode ser avaliada no servidor por impossibilidade de acessar R_n.

Contudo, sabemos que cada M_n satisfaz

$$\frac{p_F^{(n)}(\tau|s_o)}{p_B^{(n)}(\tau|G)R_n(G)} = \frac{p_F^{(n)}(\tau'|s_o)}{p_B^{(n)}(\tau'|G')R_n(G')}, \tag{9}$$

aproximadamente.

Substitutindo $\frac{R_n(G')}{R_n(G)}$ da Equação (8) na Equação (9), obtemos

$$\frac{p_F(\tau|s_o)}{p_B(\tau|G)} = \frac{p_F(\tau'|s_o)}{p_B(\tau'|G')} \prod_{1\leq n\leq N} \frac{p_F^{(n)}(\tau|s_o)}{p_B^{(n)}(\tau|G)R_n(G)} \cdot \frac{p_B^{(n)}(\tau'|G')}{p_F^{(n)}(\tau'|s_o)},.$$

Esta é a **condição de balanço agregado**.

GFlowNets: produto de distribuições

$M^\star$ terá que satisfazer a

$$\frac{p_F(\tau|s_o)}{p_B(\tau|G)\prod_{1\leq n\leq N} R_n(G)} = \frac{p_F(\tau'|s_o)}{p_B(\tau'|G')\prod_{1\leq n\leq N} R_n(G')}, \tag{8}$$

que não pode ser avaliada no servidor por impossibilidade de acessar R_n.

Contudo, sabemos que cada M_n satisfaz

$$\frac{p_F^{(n)}(\tau|s_o)}{p_B^{(n)}(\tau|G)R_n(G)} = \frac{p_F^{(n)}(\tau'|s_o)}{p_B^{(n)}(\tau'|G')R_n(G')}, \tag{9}$$

aproximadamente.

Substitutindo $\frac{R_n(G')}{R_n(G)}$ da Equação (8) na Equação (9), obtemos

$$\frac{p_F(\tau|s_o)}{p_B(\tau|G)} = \frac{p_F(\tau'|s_o)}{p_B(\tau'|G')} \prod_{1\leq n\leq N} \frac{p_F^{(n)}(\tau|s_o)}{p_B^{(n)}(\tau|G)R_n(G)} \cdot \frac{p_B^{(n)}(\tau'|G')}{p_F^{(n)}(\tau'|s_o)},.$$

Esta é a **condição de balanço agregado**.

GFlowNets: produto de distribuições

Como antes, esta condição de balanço induz a função de perda

$$\mathcal{L}_{AB}(\tau, \tau'; p_F, p_B) = \left(\log \frac{p_F(\tau|s_o)}{p_B(\tau|G)} \prod_{1 \leq n \leq N} \frac{p_B^{(n)}(\tau|G)}{p_F^{(n)}(\tau|s_o)} - \log \frac{p_F(\tau'|s_o)}{p_B(\tau'|G')} \prod_{1 \leq n \leq N} \frac{p_B^{(n)}(\tau'|G')}{p_F^{(n)}(\tau'|s_o)} \right)^2 .$$

Nós chamamos o modelo resultante da minimização de

$$\mathbb{E}_{\tau, \tau' \sim \xi}[\mathcal{L}_{AB}(\tau, \tau'; \theta^\star)]$$

de **GFlowNet Embaraçosamente Paralela** (EP-GFlowNet).

EP-GFlowNets: Resumo

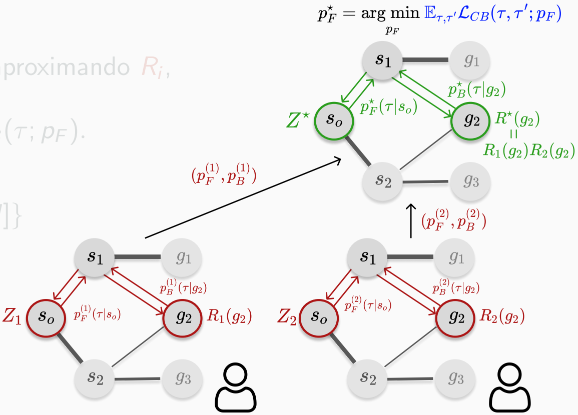
Objetivo. Obter uma GFlowNet para amostrar de $\prod_{n=1}^{N} R_n$ sem acesso simultâneo a todos R_i's.

1. Treinamos N modelos locais aproximando R_i,

$$p_F^{(i)} = \arg\min_{p_F} \mathbb{E}_{\tau} \mathcal{L}_{TB}(\tau; p_F).$$

2. Utilizamos $\{(p_F^{(i)}, p_B^{(i)}) \colon i \in [N]\}$ para estimar $p_F^{\star}$,

$$p_F^{\star} = \arg\min_{p_F} \mathbb{E}_{\tau,\tau'} \mathcal{L}_{AB}(\tau, \tau'; p_F).$$



Em seguida, consideramos amostras da marginal $p_T^{\star}$ como aproximações da distribuição induzida por $\prod R_n$.

EP-GFlowNets: Resumo

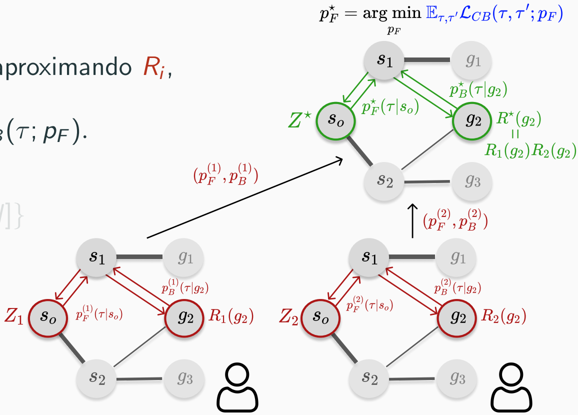
Objetivo. Obter uma GFlowNet para amostrar de $\prod_{n=1}^{N} R_n$ sem acesso simultâneo a todos R_i's.

1. Treinamos N modelos locais aproximando R_i,

$$p_F^{(i)} = \underset{p_F}{\arg\min}\, \mathbb{E}_\tau \mathcal{L}_{TB}(\tau; p_F).$$

2. Utilizamos $\{(p_F^{(i)}, p_B^{(i)}) : i \in [N]\}$ para estimar $p_F^\star$,

$$p_F^\star = \underset{p_F}{\arg\min}\, \mathbb{E}_{\tau,\tau'} \mathcal{L}_{AB}(\tau, \tau'; p_F).$$



Em seguida, consideramos amostras da marginal $p_T^\star$ como aproximações da distribuição induzida por $\prod R_n$.

EP-GFlowNets: Resumo

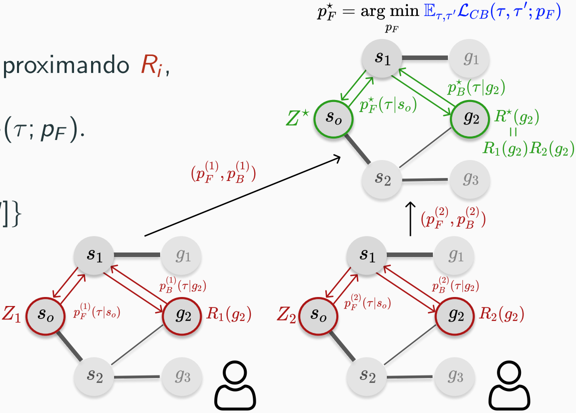
Objetivo. Obter uma GFlowNet para amostrar de $\prod_{n=1}^{N} R_n$ sem acesso simultâneo a todos R_i's.

1. Treinamos N modelos locais aproximando R_i,

$$p_F^{(i)} = \arg\min_{p_F} \mathbb{E}_\tau \mathcal{L}_{TB}(\tau; p_F).$$

2. Utilizamos $\{(p_F^{(i)}, p_B^{(i)}) \colon i \in [N]\}$ para estimar $p_F^\star$,

$$p_F^\star = \arg\min_{p_F} \mathbb{E}_{\tau,\tau'} \mathcal{L}_{AB}(\tau, \tau'; p_F).$$



Em seguida, consideramos amostras da marginal $p_T^\star$ como aproximações da distribuição induzida por $\prod R_n$.

EP-GFlowNets: Resumo

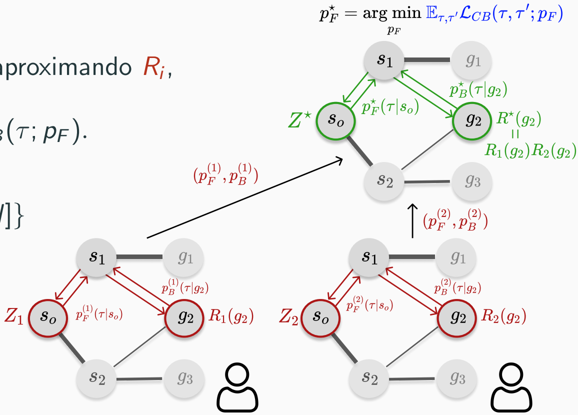
Objetivo. Obter uma GFlowNet para amostrar de $\prod_{n=1}^{N} R_n$ sem acesso simultâneo a todos R_i's.

1. Treinamos N modelos locais aproximando R_i,

$$p_F^{(i)} = \arg\min_{p_F} \mathbb{E}_\tau \mathcal{L}_{TB}(\tau; p_F).$$

2. Utilizamos $\{(p_F^{(i)}, p_B^{(i)}) \colon i \in [N]\}$ para estimar $p_F^\star$,

$$p_F^\star = \arg\min_{p_F} \mathbb{E}_{\tau,\tau'} \mathcal{L}_{AB}(\tau, \tau'; p_F).$$



Em seguida, consideramos amostras da marginal $p_T^\star$ como aproximações da distribuição induzida por $\prod R_n$.

Exemplo: Inferência federada com GFlowNets

Aprendizado de estruturas

Uma **rede Bayesiana** é um grafo acíclio e direcionado G especificando explicitamente as **distribuições e independências condicionais** de um grupo de variáveis aleatórias $\{X_1, \ldots, X_m\}$. Assim,

$$P(X_1, \ldots, X_m) = \prod_{1 \leq i \leq m} P(X_i | Pa(X_i)), \tag{10}$$

em que $Pa(X_i)$ é o conjunto de pais de X_i na rede Bayesiana.

Problema. Dado um conjunto de N realizações independentes de P, $\mathcal{D} \in \mathbb{R}^{N \times m}$, queremos inferir a rede Bayesiana subjacente.

Aprendizado federado de estrutura de redes Bayesianas

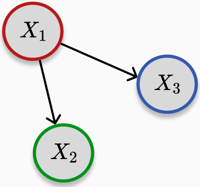
Aprendizado de estruturas

Uma **rede Bayesiana** é um grafo acíclio e direcionado G especificando explicitamente as **distribuições e independências condicionais** de um grupo de variáveis aleatórias $\{X_1, \ldots, X_m\}$. Assim,

$$P(X_1, \ldots, X_m) = \prod_{1 \leq i \leq m} P(X_i | Pa(X_i)), \tag{10}$$

em que $Pa(X_i)$ é o conjunto de pais de X_i na rede Bayesiana.

Problema. Dado um conjunto de N realizações independentes de P, $\mathcal{D} \in \mathbb{R}^{N \times m}$, queremos inferir a rede Bayesiana subjacente.



$$P(X_1, X_2, X_3) = P(X_1)P(X_2|X_1)P(X_3|X_1)$$

Aprendizado federado de estrutura de redes Bayesianas

Sejam $\mathcal{D}^{(1)}, \ldots, \mathcal{D}^{(K)}$ **conjuntos de dados pequenos e privados** possuídos por K diferentes clientes. A distribuição alvo é

$$R^\star\left(G \middle| \bigcup_k \mathcal{D}^{(k)}\right) \coloneqq \prod_{k=1}^{K} R(G|\mathcal{D}^{(k)}). \tag{11}$$

Intuitivamente, $R^\star$ atribui **alta probabilidade** a estruturas **consistentes com todos** os modelos locais $R_n \coloneqq R(\cdot|\mathcal{D}^{(n)})$.

Consideraremos que os dados seguem o modelo Gaussiano linear

$$X_i|X_{j\neq i}, G, \theta_i \sim \mathcal{N}(\theta_i^T X_{Pa(X_i)}, \sigma^2), \tag{12}$$

e basearemos nossa inferência na distribuição

$$R(G|\mathcal{D}) = \prod_{1\leq n\leq N} \max_{\theta_n} P(\mathcal{D}_n|G, \theta_n). \tag{13}$$

Aprendizado federado de estrutura de redes Bayesianas

Sejam $\mathcal{D}^{(1)}, \ldots, \mathcal{D}^{(K)}$ **conjuntos de dados pequenos e privados** possuídos por K diferentes clientes. A distribuição alvo é

$$R^{\star}\left(G \middle| \bigcup_{k} \mathcal{D}^{(k)}\right) \coloneqq \prod_{k=1}^{K} R(G|\mathcal{D}^{(k)}). \tag{11}$$

Intuitivamente, $R^{\star}$ atribui **alta probabilidade** a estruturas **consistentes com todos** os modelos locais $R_n \coloneqq R(\cdot|\mathcal{D}^{(n)})$.

Consideraremos que os dados seguem o modelo Gaussiano linear

$$X_i|X_{j\neq i}, G, \theta_i \sim \mathcal{N}(\theta_i^T X_{Pa(X_i)}, \sigma^2), \tag{12}$$

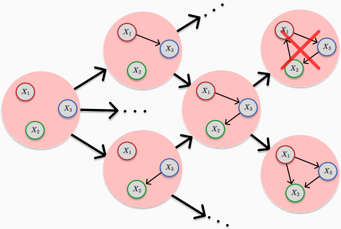
e basearemos nossa inferência na distribuição

$$R(G|\mathcal{D}) = \prod_{1\leq n\leq N} \max_{\theta_n} P(\mathcal{D}_n|G, \theta_n). \tag{13}$$

EP-GFlowNets: Aprendizado federado de estruturas

Há três etapas: geração, treinamento e comunicação.

1. Construir um **processo generativo iterativo** sobre GADs.



2. Treinar os **modelos locais** para aproximar R_k.

$$p_F^{(k)}, p_B^{(k)} = \arg\min_{p_F} \mathbb{E}_\tau \mathcal{L}_{CB}(\tau; p_F, p_B; R_k) \quad (14)$$

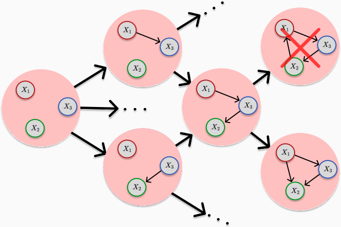
3. **Comunicar** $p_F^{(k)}$'s a um servidor e aprender o **modelo global**.

$$p_F^\star = \arg\min_{p_F} \mathbb{E}_{\tau,\tau'} \mathcal{L}_{CB}(\tau, \tau'; p_F; p_F^{(1)}, \ldots, p_F^{(K)}). \quad (15)$$

EP-GFlowNets: Aprendizado federado de estruturas

Há três etapas: geração, treinamento e comunicação.

1. Construir um **processo generativo iterativo** sobre GADs.



2. Treinar os **modelos locais** para aproximar R_k.

$$p_F^{(k)}, p_B^{(k)} = \arg\min_{p_F} \mathbb{E}_\tau \mathcal{L}_{CB}(\tau; p_F, p_B; R_k) \tag{14}$$

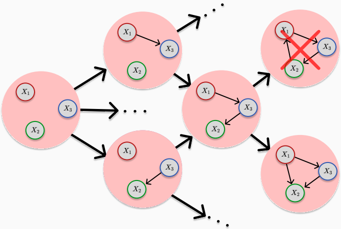
3. **Comunicar** $p_F^{(k)}$'s a um servidor e aprender o **modelo global**.

$$p_F^\star = \arg\min_{p_F} \mathbb{E}_{\tau,\tau'} \mathcal{L}_{CB}(\tau, \tau'; p_F; p_F^{(1)}, \ldots, p_F^{(K)}). \tag{15}$$

EP-GFlowNets: Aprendizado federado de estruturas

Há três etapas: geração, treinamento e comunicação.

1. Construir um **processo generativo iterativo** sobre GADs.



2. Treinar os **modelos locais** para aproximar R_k.

$$p_F^{(k)}, p_B^{(k)} = \arg\min_{p_F} \mathbb{E}_\tau \mathcal{L}_{CB}(\tau; p_F, p_B; R_k) \tag{14}$$

3. **Comunicar** $p_F^{(k)}$'s a um servidor e aprender o **modelo global**.

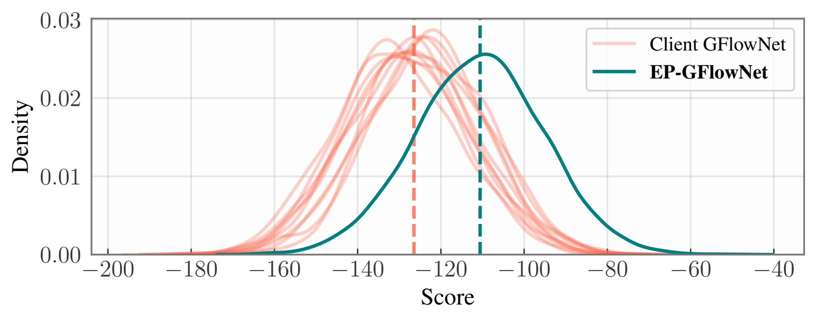
$$p_F^\star = \arg\min_{p_F} \mathbb{E}_{\tau,\tau'} \mathcal{L}_{CB}(\tau, \tau'; p_F; p_F^{(1)}, \ldots, p_F^{(K)}). \tag{15}$$

EP-GFlowNets: Aprendizado federado de estruturas

Máxima verossimilhança dos dados sob G:

$$\text{Score}(G) = \prod_{1 \leq i \leq N} \max_{\theta_i} P\left(\bigcup_{n=1}^{K} \mathcal{D}_i^{(n)} \middle| G, \theta_i\right) \tag{16}$$

Avaliação. A distribuição induzida pela EP-GFlowNet se concentra em grafos consistentes com $\bigcup_n \mathcal{D}^{(n)}$ mais que qualquer dos modelos locais.



Distribuição de Score(G) quando $G \sim p_F^{(n)}$ e quando $G \sim p_F^{\star}$. ↑

(resultados similares quando usamos um conjunto de teste)

Mensagem para casa

Mensagem para casa

GFlowNets são poderosos amostradores em espaços discretos

GFlowNets aprendem um processo generativo iterativo com protocolos de transição parametrizados por redes neurais para amostrar sobre distribuições discretas em espaços complexos (como grafos).

EP-GFlowNets ensejam aprendizado federado e distribuído

A composição multiplicativa de GFlowNets baseada na nossa condição de balanço contrastiva abre espaço para aplicações em inferência distribuída e federada e otimização multiobjetivo.

GFlowNets têm grande potencial em inferência Bayesiana

GFlowNets demonstraram resultados promissores em problemas de aprendizado de estrutura, processamento de linguagem natural baseado em LLMs e inferência filogenética.

Perguntas?

Comentários?

Obrigado!