Generative Flow Networks for Causal Discovery

-- Li Wenqian

Agenda

1. Motivation

2. GFlowNet Foundation

3. Causal Discovery

4. Research Proposal and Progress

Inefficient *Exploration* makes Reinforcement Learning project difficult to be applied in large-scale

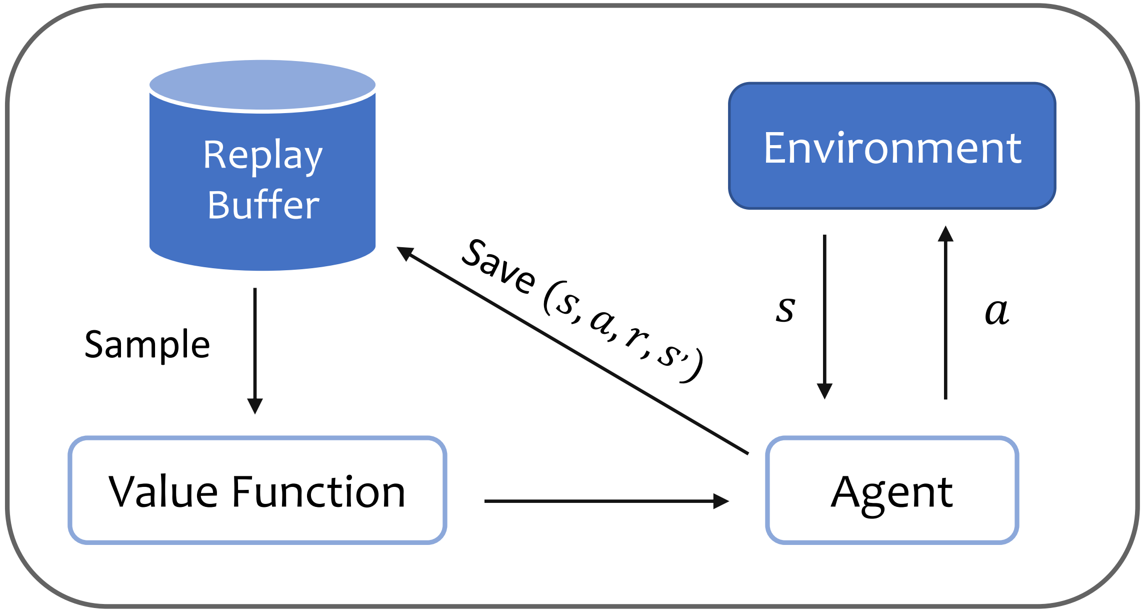
Inefficient Sampling makes the RL agent goes in a wrong direction. Most generated samples could not help us find a good trajectory / solution.

Exploration Strategy is important to the RL agent to find the optimal trajectory while the full exploration for each trajectory (especially for the case with long trajectory) will waste time.

Sub-optimal Result is possible since we could not guarantee the agent to have explored all potential trajectories.

The limitation of Reinforcement Learning

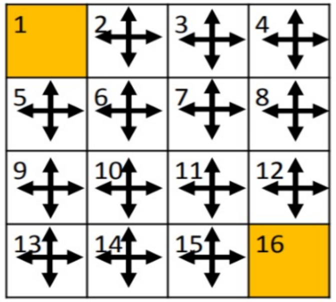
— *Random Policy in the Grid World environment makes the robot take the repeating path*



Working Flow of Reinforcement Learning

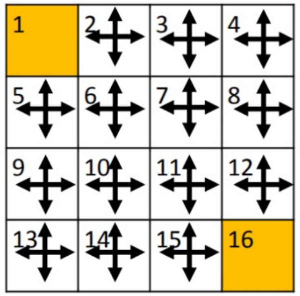
0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0
0.0	0.0	0.0	0.0

π_0



For the grid world environment experiment, we initialize the value function (left) with zero . The robot would loop in the agent-environment cycle until the terminal state would be achieved

π_0



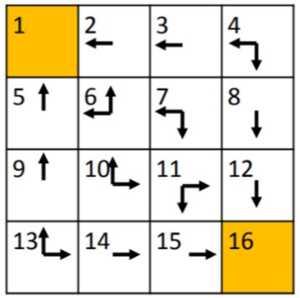
$\xrightarrow{E}$

v_{π_0}

0.0	-14.	-20.	-22.
-14.	-18.	-20.	-20.
-20.	-20.	-18.	-14.
-22.	-20.	-14.	0.0

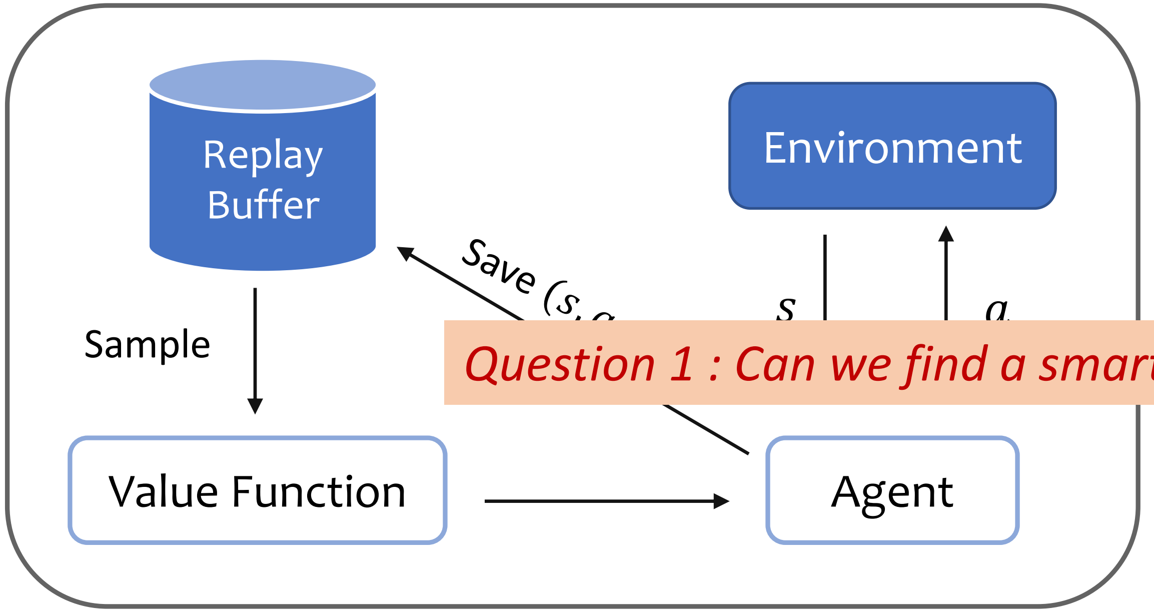
$\xrightarrow{I}$

π_1



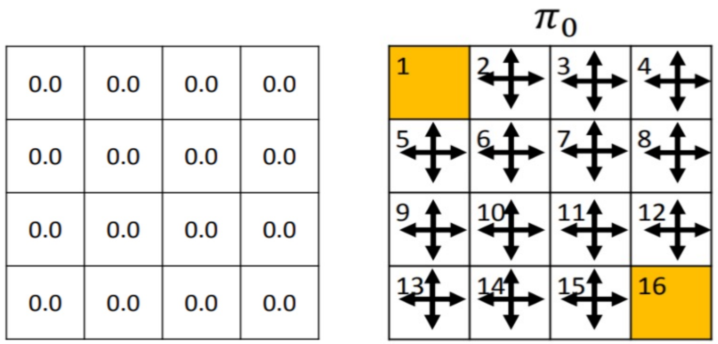
The limitation of Reinforcement Learning

— Random Policy in the Grid World environment makes the robot take the repeating path

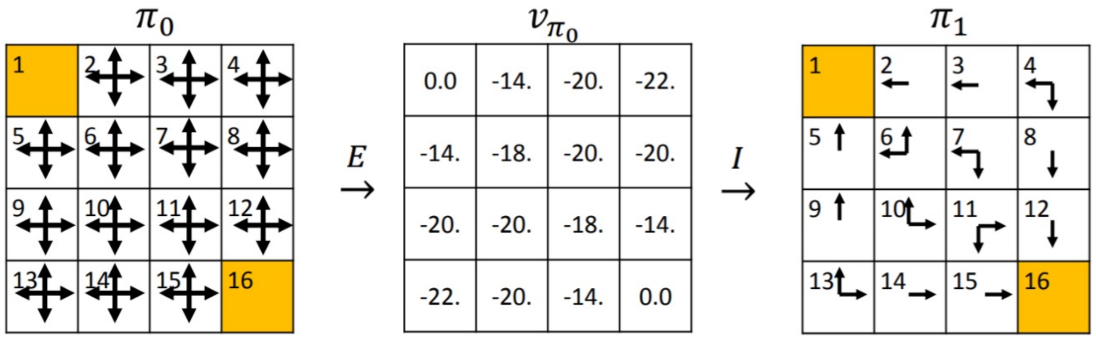


Working Flow of Reinforcement Learning

Question 1 : Can we find a smarter way to train the policy ?



For the grid world environment experiment, we … with zero . The robot … ment cycle until the terminal state would be achieved

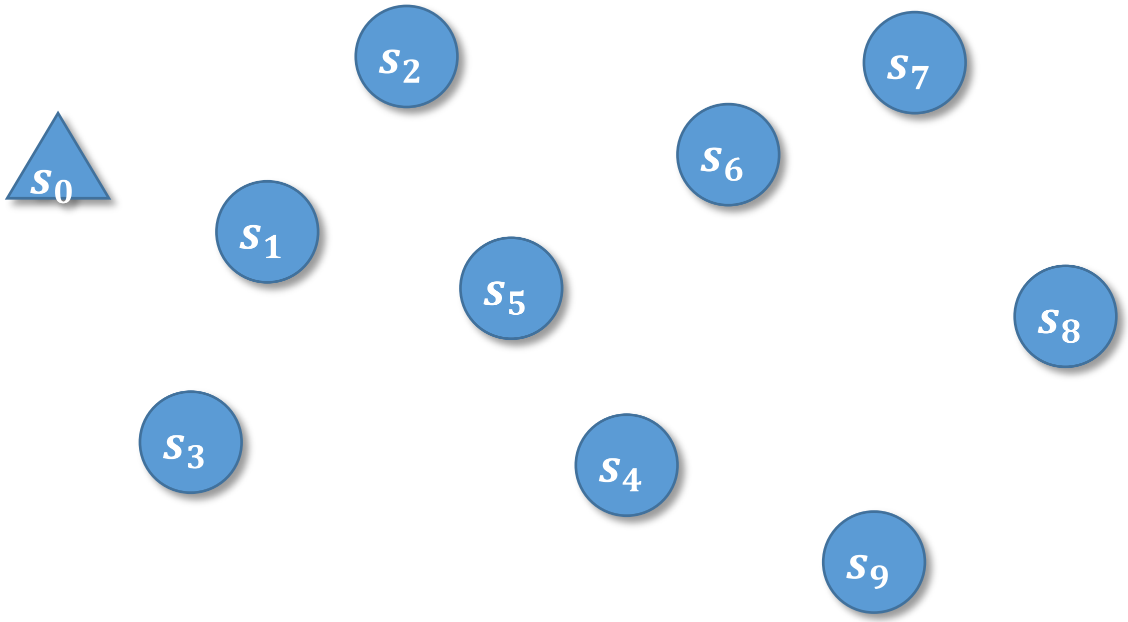


Motivation

The limitation of Reinforcement Learning

— For generating compositional object, the search space increases in exponential

For every current state s_i, we have (n-1) states for search choice

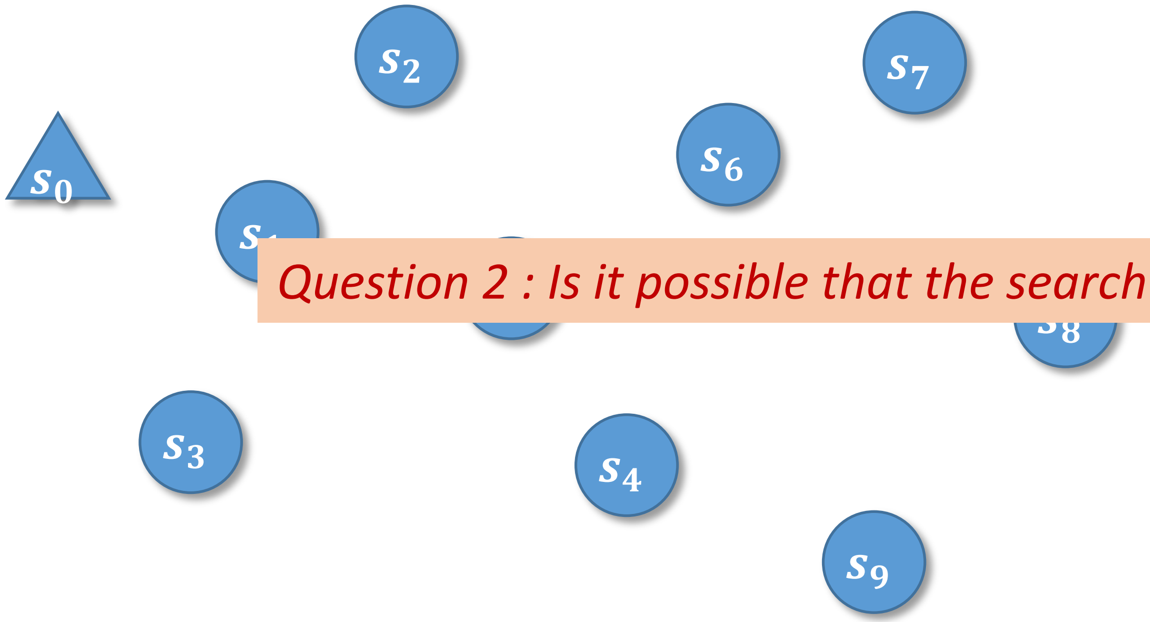


Go moves with exponential choices

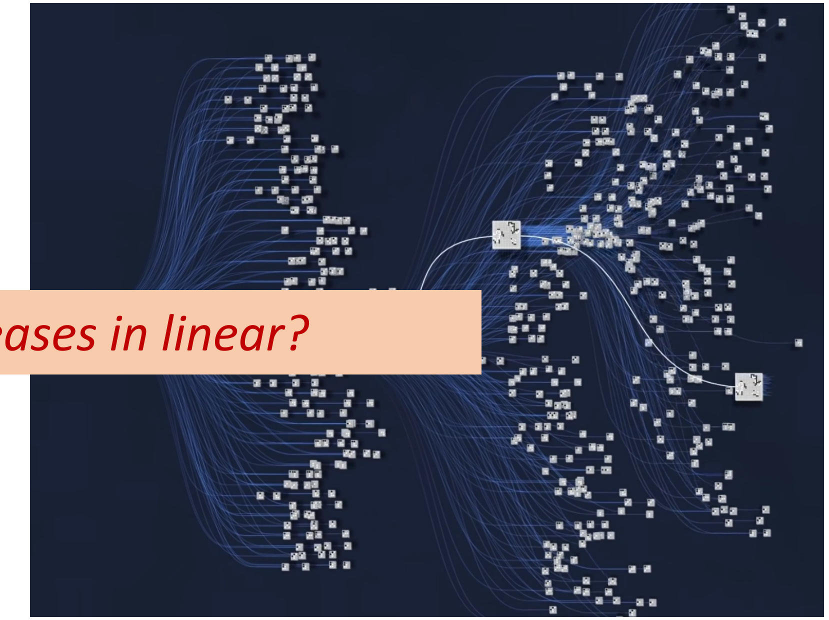
The limitation of Reinforcement Learning

— For generating compositional object, the search space increases in exponential

For every current state s_i, we have (n-1) states for search choice



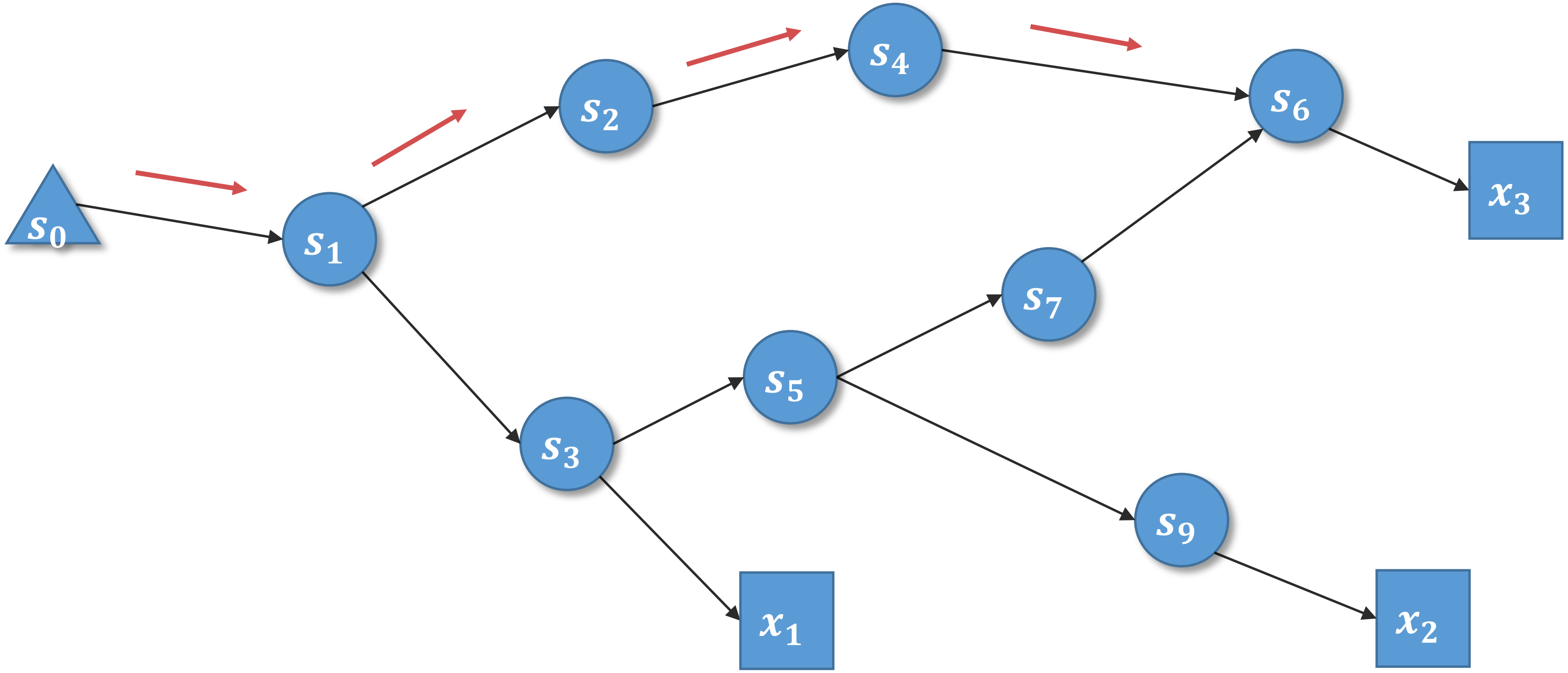
Question 2 : Is it possible that the search space increases in linear?



Go moves with exponential choices

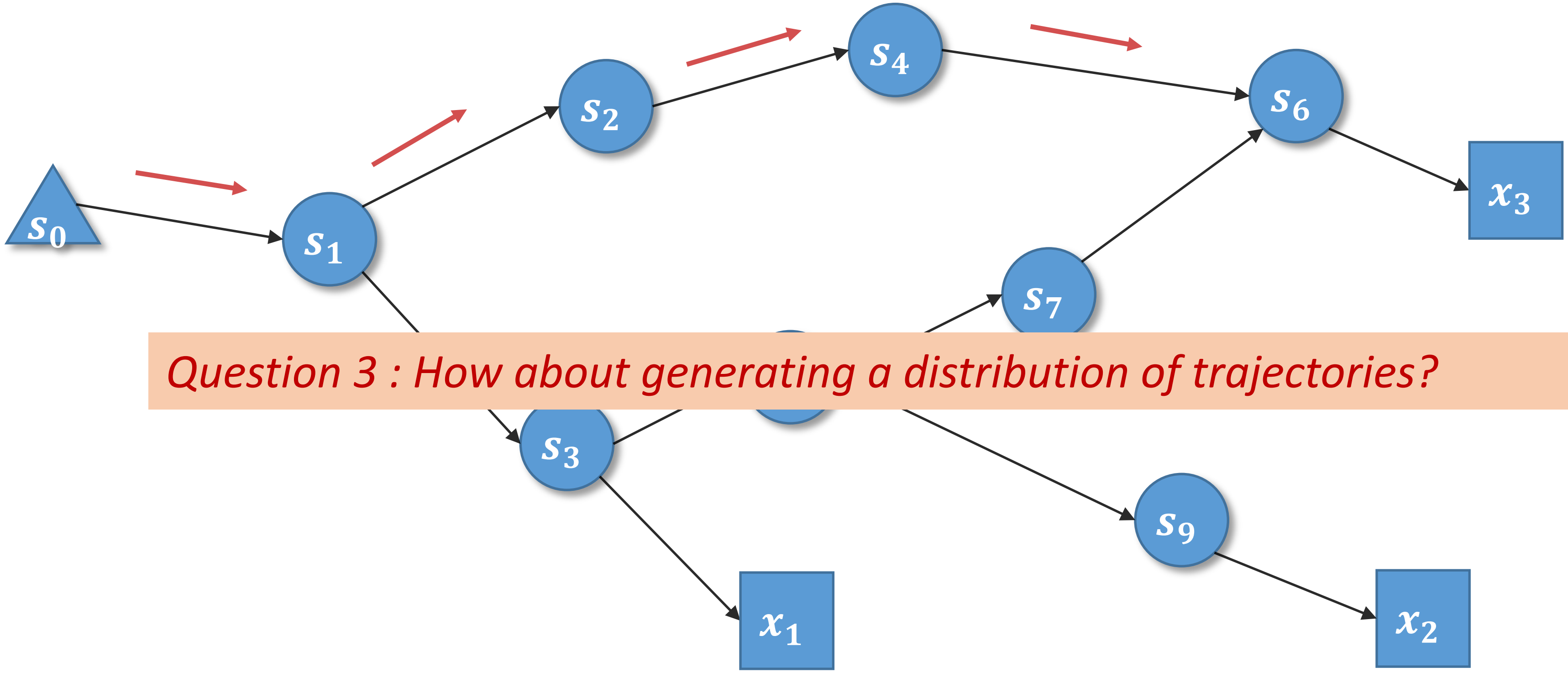
The limitation of Reinforcement Learning

— For an episodic RL setting , the agent could obtain the reward until the terminal state. The maximization of expected return in RL generates the single highest-reward sequence of actions.



The limitation of Reinforcement Learning

— For an episodic RL setting , the agent could obtain the reward until the terminal state. The maximization of expected return in RL generates the single highest-reward sequence of actions.



Question 3 : How about generating a distribution of trajectories?

Generative Flow Networks (GFlowNet)

In 17 Nov 2021. A 70-Page paper from Yoshua Bengio team: *GFlowNet Foundations,* proposed a low-network-based generative method that can turn a given positive reward into a generative policy that samples with a probability proportional to the return.

GFlowNet Foundations

Yoshua Bengio[1,2,5], Tristan Deleu[1,2], Edward Hu[6,1],
Salem Lahlou[1,2], Mo Tiwari[4], and Emmanuel Bengio[1,3]

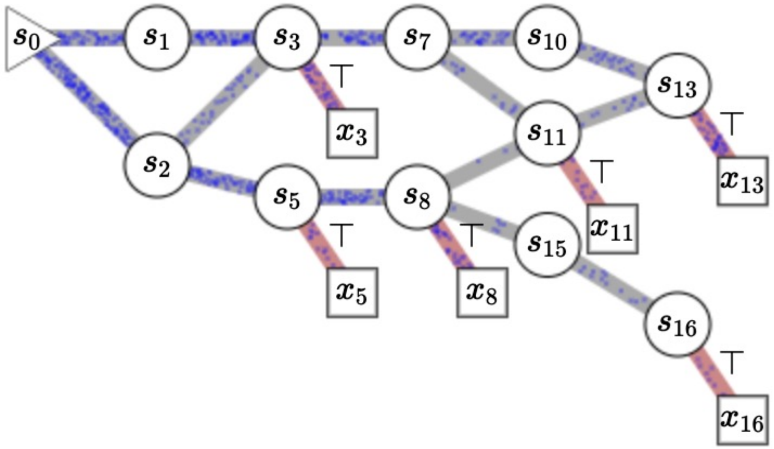
[1]Mila
[2]University of Montreal
[3]McGill University
[4]Stanford University
[5]CIFAR,IVADO
[6]Microsoft Azure AI

Paper Link : https://arxiv.org/abs/2111.09266

GFlowNet Basic

- GFlowNet: Learning a Flow

Definition 18. *A* **GFlowNet** *is a pair* $(\hat{F}(s), \hat{P}_F(s_{t+1}|s_t))$ *where* $\hat{F}(s)$ *is a state flow function and* $\hat{P}_F(s_{t+1}|s_t)$ *is a transition distribution from which one can draw trajectories* τ *by iteratively sampling each state given the previous one, starting at initial state* s_0 *and then with* $s_{t+1} \sim \hat{P}_F(s_{t+1}|s_t)$ *for* $t = 0, 1, \ldots$ *until sink state* $s_{n+1} = s_f$ *is reached for some* n*.*



GFlowNet constructs a Flow Network based on Directed Acyclic Graph (DAG) with sources and sinks , and edges carrying some amount of flow between them through intermediate nodes -- think of pipes of water.

GFlowNet Basic

- How to train GFlowNet ? A TD-like Objective Function

Inflow of a state **Outflow / Reward of a state**

$$\mathcal{L}_{\theta,\epsilon}(\tau) = \sum_{s' \in \tau \neq s_0} \left(\log \left[\epsilon + \sum_{s,a:T(s,a)=s'} \exp F_\theta^{\log}(s,a) \right] - \log \left[\epsilon + R(s') + \sum_{a' \in \mathcal{A}(s')} \exp F_\theta^{\log}(s',a') \right] \right)^2 .$$

- Satisfying the flow equation yields what we want

Let $F(s,a) = f(s,s')$ be the flow between s and s', where $T(s,a) = s'$, i.e. s' is the (deterministic) state transitioned to from state s and action a. Let

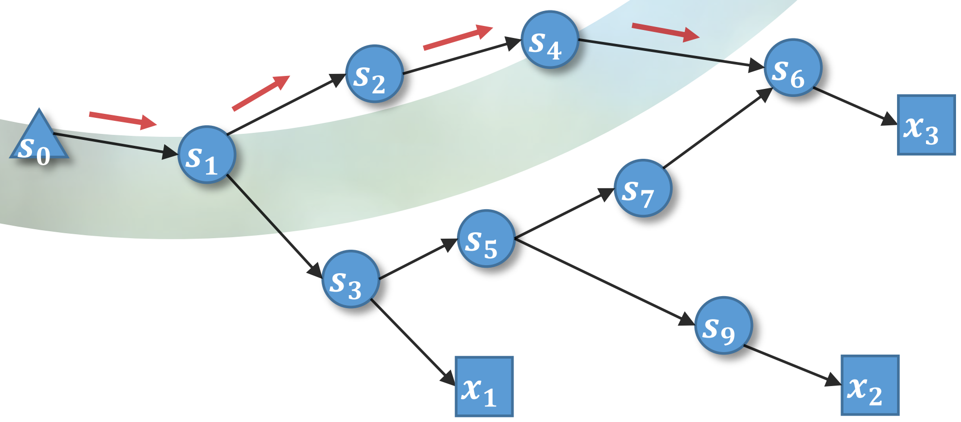
$$\pi(a|s) = \frac{F(s,a)}{\sum_{a'} F(s,a')}$$

then following policy π, starting from s_0, leads to terminal state x with probability $R(x)$ (see the paper for proofs and more rigorous explanations).

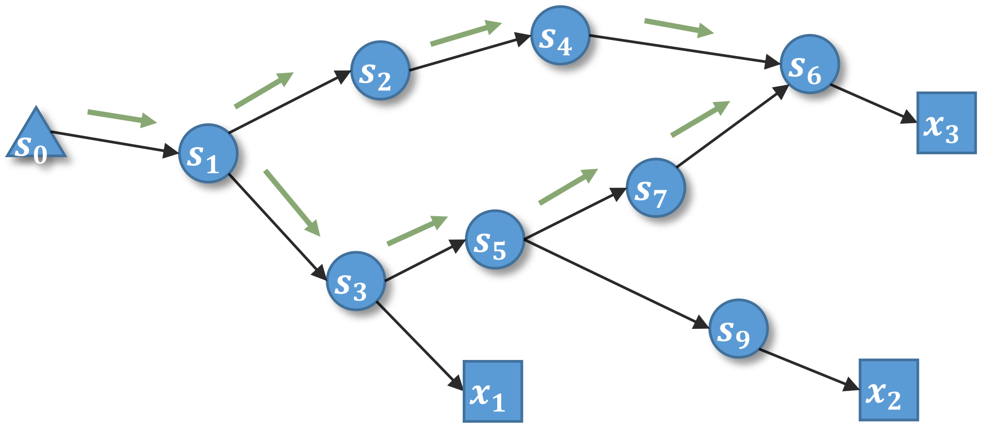
GFlowNet vs Reinforcement Learning

Method	Objective Function	Exploration Strategy	Search Space
Reinforcement Leaning	Maximization of reward	A single highest-reward solution	Exponential
GFlowNet	Flow Matching	A distribution of high rewards solution	Linear

RL



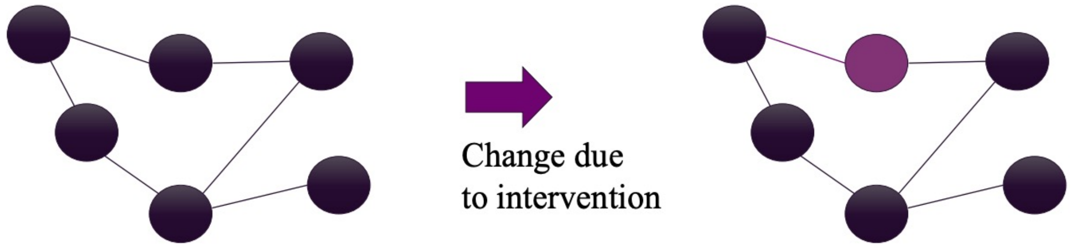
GFlowNet



We are interested in the "Why" , "How" and "So what "of modeling

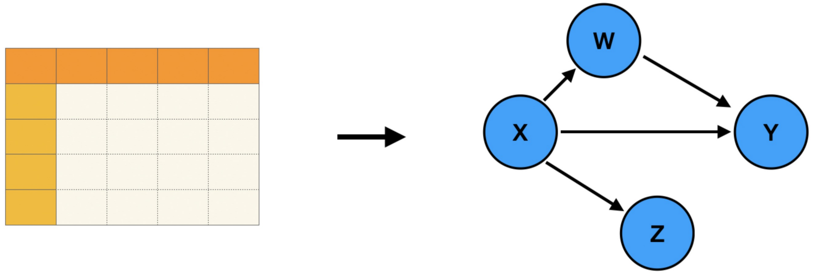
The interpretability, generalization and over-reliance on data of deep learning are currently recognized challenges.

- **Understand why** by uncovering the generating mechanisms behind the data
- **Understand how** by intervention and counterfactual
- **Understand so-what** by extending this study to structure learning



Causal Discovery

- <u>Objective:</u> It aims to infer causal structure from data. In other words, given a dataset, derive a causal model that describes it.



	X	W	Y	Z
X	0	1	1	1
W	0	0	1	0
Y	0	0	0	0
Z	0	0	0	0

Big picture goal of causal discovery

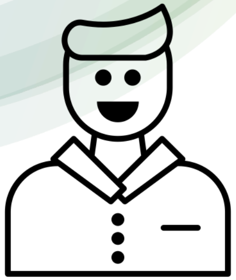
- <u>Current Solution:</u> Apply RL methods (eg. Actor-Critic algorithm), or generative models such as GAN , VAE to generate DAGs and score them based on metrics (BIC , ELBO ,etc)

- <u>Challenges:</u> Could not handle large-scale sets (more than 50 nodes) ; Could not attain the optimal solution (exponential growth)

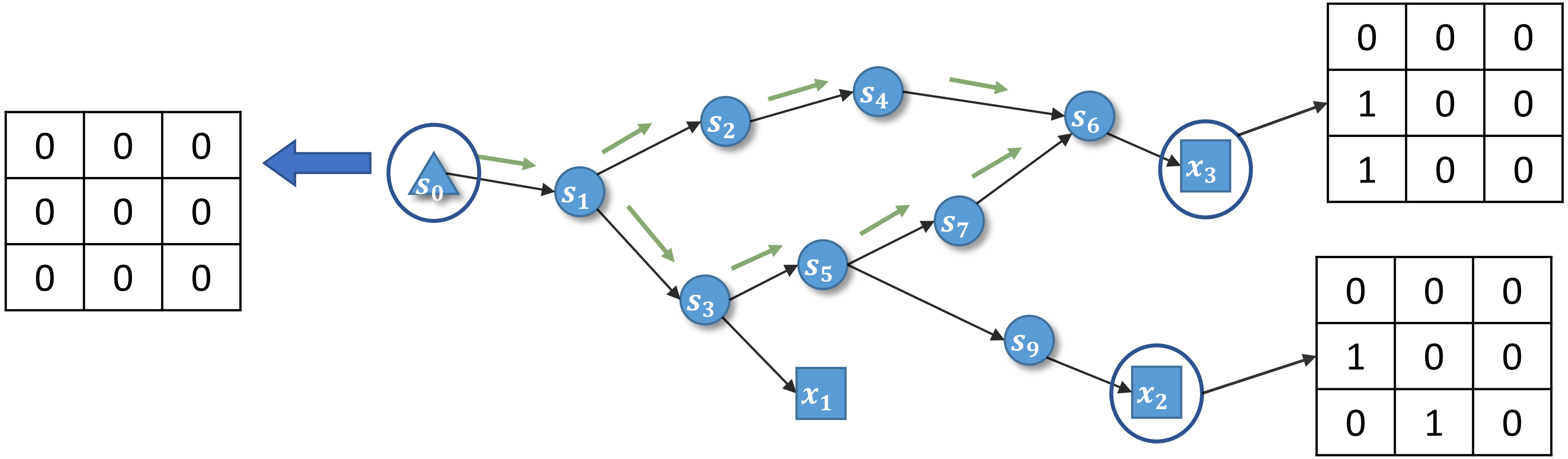
$$h(\mathbf{A}) = tr(e^{\mathbf{A} \odot \mathbf{A}}) - d = 0 \longleftarrow$$

Constraint function h for enforcing acyclicity. The matrix exponential requires O (d^3) computations

GFlowNet for Causal Discovery (GFN-Causal)



- ✓ GFlowNet constructs Flow Model based on DAG, which meets the acyclic constraint in Causal Discovery .
- ✓ For a large-scale experiment (eg. over 100 nodes/variables), it is difficult to find the highest-reward optimal structure. In addition, we could not guarantee that generated structure is ”indeed” the best one. It is better to find various good solutions.



Research Progress and Plan

What I have done :

- ✓ Designed the *(State , Action)* pairs in GFlowNet for Causal Discovery research
- ✓ Finished 1st version code. A toy example with 10 nodes, which could successfully convert a dataset into a causal matrix

Future…

- ■ Explore a better score metric in the GFN-Causal model.
- ■ Compare more SOTA baselines (NOTEARS , RL-Causal , etc.)
- ■ Try the large-scale experiments
- ■ Complete the mathematical analysis part