LAB MEETING: A Connection Between Generative Adversarial Networks, Inverse Reinforcement Learning and Energy-Based models

Suwon Suh

POSTECH MLG

Feb, 13, 2017

Goal

- Understanding Basic Models
  1) Generative Adversarial Networks (GAN)
  2) Energy Based Model (EBM)
  3) Inverse Reinforcement Learning (IRL)
- Relationship among Three models
  1) Equivalence between Guided Cost Learning and GAN
- New algorithm for EBM training with GAN
  1) New type of discriminator with model distribution (EBM) and sampling distribution
  2) We can get efficient sampler as a result!

GAN []

A generative model in adversarial setting

- Generative model with Discriminator:

$$\min_G \max_D V(G, D) = E_{x\sim P}[\log D(x)] + E_{z\sim Unif}[\log(1 - D(G(z)))] \,,$$

rewriting it as:

$$\min_G \max_D V(G, D) = E_{x\sim P}[\log D(x)] + E_{x\sim Q}[\log(1 - D(x))] \,,$$

P: Data distribution, Q: Distribution of the generator.

- Optimal discriminator D^* fixing G

$$D^* = \frac{P(x)}{P(x) + Q(x)} \tag{1}$$

A Variant of GAN minimizing $KL[Q||P]$

- The loss function for a discriminator

$$Loss(D) = E_{x\sim P}[-\log D(x)] + E_{x\sim Q}[-\log(1 - D(x))]$$

- The original loss function for a generator []

$$Loss^{org}(G) = E_{x\sim G}[\log(1 - D(x))] \ ,$$

$\log(1 - D(x)) \approx \log(1)$ when it starts to learn slowly because gradient of $\frac{d\log(x)}{dx}|_{x=1}$ is not steep, which brings an alternative loss

$$Loss^{alter}(G) = -E_{x\sim G}[\log(D(x))] \ ,$$

- We can use both []:

$$L_{gen}(G) = Loss^{org}(G) + Loss^{alter}(G) = E_{x\sim G}[\log \frac{(1 - D(x))}{D(x)}]$$

A Variant of GAN minimizing $KL[Q||P]$

- Huszar says "it minimizes $KL[Q||P]$ when D is near D^*" []:

$$E_{x\sim G}[\log \frac{(1-D(x))}{D(x)}] \approx E_{x\sim G}[\log \frac{(1-D^*(x))}{D^*(x)}]$$

$$= E_{x\sim Q}[\log \frac{Q(x)}{P(x)}] = KL[Q||P]$$

by invoking Eq. 1.

Energy Based Models (EBMs)

- Every configuration $x \in R^D$ has a corresponding energy $E_\theta(x)$.
- By normalizing them, we can define probability density function (pdf), $p_\theta(x) = \frac{\exp(-E_\theta(x))}{Z}$, where $Z(\theta) = \int \exp(-E_\theta(x'))dx'$.
- How to learn parameters θ?
  $\log p_\theta(x) = -E_\theta(x) - \log(Z(\theta))$
- Too many configuration, we need to estimate $Z(\theta)$ with samples with Markov chain Monte Carlo (MCMC)
  1) Constrative Divergence with only one K-step sample from a MCMC chain.
  2) Persistent CD maintains multiple chains to sample from the model in the learning process using Stochastic Gradient Descent (SGD).

Inverse Reinforcement Learning

Inverse Reinforcement Learning (IRL)

Given states X, actions U, dynamics $P(x_{t+1}|x_t, u_t)$ and discount factor γ in $MDP(X, U, P, c_\theta, \gamma)$ and demonstrations of experts, we need to find cost or negative reward c_θ .

- Maximum entropy inverse reinforcement learning (MaxEnt IRL) models demonstration with Boltzmann distribution

$$p_\theta(\tau) = \frac{exp(-c_\theta(\tau))}{Z} ,$$

  $\tau = \{x_1, u_1, \cdots, x_T, u_T\}$ is a trajectory $c_\theta(\tau) = \sum_t c_\theta(x_t, u_t)$
- Guided cost learning (CGL), where partition function Z is approximated by importance sampling

$$L_{cost}(\theta) = E_{\tau\sim P}[-\log p_\theta(\tau)] = E_{\tau\sim P}[c_\theta(\tau)] + \log Z$$

$$= E_{\tau\sim P}[c_\theta(\tau)] + \log(E_{\tau\sim q}[\frac{exp(-c_\tau(\tau))}{q(\tau)}])$$

Inverse Reinforcement Learning

CGL needs to match sampling distribution $q(\tau)$ to model distribution $p_\theta(\tau)$

$L_{sampler}(q) = KL[q(\tau)||p_\theta(\tau)]$,

where we only choose term that related to q:

$$L_{sampler}(q) = E_{\tau\sim Q}[c_\theta(\tau)] + E_{\tau\sim Q}[\log q(\tau)] \ ,$$

modifying sampling distribution with mixture

To reduce the variance of a estimator Z using q only, $\mu = \frac{1}{2}p + \frac{1}{2}q$ is used as sampling distribution.

$$L_{cost}(\theta) = E_{\tau\sim P}[c_\theta(\tau)] + \log(E_{\tau\sim\mu}[\frac{exp(-c_\tau(\tau))}{\frac{1}{2}\tilde{p} + \frac{1}{2}q}])$$

, where $\tilde{p}$ is a rough estimate for density of demonstrations using the current model p_θ.

Model

(Idea) Explicitly modeling a discriminator D in the form of the optimal discriminator D^*

We assume p is the data distribution, $\tilde{p}_\theta$ is a model distribution parameterized θ and q is a sampling distribution;

- Before $D^* = \frac{p(\tau)}{p(\tau)+q(\tau)}$
- After $D_\theta = \frac{\tilde{p}_\theta(\tau)}{\tilde{p}_\theta(\tau)+q(\tau)}$
- Why EBM as a model distribution?
  Product of Experts (PoE) can capture modes and put less density between modes compared to Mixture of Experts (MoE) of similar capacity.

$$D_\theta = \frac{\frac{1}{Z} exp(-c_\theta(\tau))}{\frac{1}{Z} exp(-c_\theta(\tau)) + q(\tau)}$$

- we need to evaluate the sampling density function $q(\tau)$ effectively to learn: Autoregressive model, Normalized Flow and MoE.

Equivalance between GAN and CGL

- loss from a variant of GAN

$$L_{disc}(\theta) = E_{\tau\sim p}[-\log D_\theta(\tau)] + E_{\tau\sim q}[-\log(1 - D_\theta(\tau))]$$

$$= E_{\tau\sim p}[-\log \frac{\frac{1}{Z}exp(-c_\theta(\tau))}{\frac{1}{Z}exp(-c_\theta(\tau)) + q(\tau)}] + E_{\tau\sim q}[-\log \frac{q(\tau)}{\frac{1}{Z}exp(-c_\theta(\tau)) + q}$$

- loss from GCL

$$L_{cost}(\theta) = E_{\tau\sim p}[c_\theta(\tau)] + \log(E_{\tau\sim \mu}[\frac{exp(-c_\theta(\tau))}{\frac{1}{2}\tilde{p} + \frac{1}{2}q}])$$

- Equivalence:
  1) The value of Z which minimizes L_{disc} is importance sampling estimator for the partition function
  2) For this value Z, the derivative of $L_{disc}(\theta)$ with respect to θ is equal to the derivative of $L_{cost}(\theta)$
  3) the derivative of $L_{gen}(q)$ with regard to q is equal to the derivative of $L_{sampler}(q)$

Tranining EBM with GAN

Why?

As PoEs, EBMs are good at modeling complicated manifold well. However, the sampling is not independent because it uses MCMC. This method directly learns effective sampling distribution.

- update partition function with importance sampling

$$Z \Leftarrow E_{\tau\sim\mu}[\frac{exp(-c_\tau(x))}{\frac{1}{2}\tilde{p}+\frac{1}{2}q}]$$

- update model parameter with SGD

$$L_{energy}(\theta) = E_{\tau\sim p}[c_\theta(x)] + \log(E_{\tau\sim\mu}[\frac{exp(-c_\theta(x))}{\frac{1}{2}\tilde{p}+\frac{1}{2}q}])$$

- update sampling parameter with SGD

$$L_{sampler}(q) = E_{\tau\sim q}[E_\theta(x)] + E_{\tau\sim q}[\log q(x)] \ ,$$

Discussion

- Return of EBMs
  Recently, EBMs have been subsided by VAE and GAN because its sampling and hardship to get approximated log-likelihood. In this model, we can evade these problem.
- Combination of EBMs and other generative models such as Autoregressive and VAE as sampler.
- Adversarial Variational Bayes
  Minimizing KL divergence with GAN.