A Distributional Approach to Controlled Text Generation

(ICLR 2021)
(Oral)


Muhammad Khalifa


Hady Elsahar


Marc Dymetman

NAVER LABS Europe

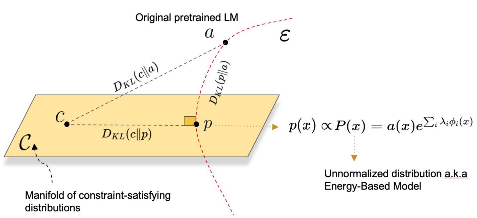
Cairo University

The first approach to controlled text generation that exploits **distributional** requirements**.**
Through minimization of KL divergence from the original pretrained LM, degeneration is avoided.

MOTIVATIONS

1. Imposing **constraints** on generations from a large pretrained LM.
2. Not only "**pointwise**" constraints on individual sequences, but also "**distributional**" constraints on statistical properties of generated sequences.
3. Avoiding **degeneration**: loss of fluency or diversity of the generated sequences.

APPROACH



- **Moment constraints** over desired generation features
- Manifold *C* of all LMs satisfying the moment constraints
- We select the LM *p* which **minimizes KL-divergence** to the original pretrained LM *a*

Step 1: From constraints to EBM

Step 2: From EBM to autoregressive policy

Desired Moment Constraints

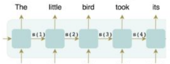
$$\mathbb{E}_{x\sim p}\phi_1(x) = \bar{\mu}_1$$
$$\mathbb{E}_{x\sim p}\phi_2(x) = \bar{\mu}_2$$
$$\dots$$
$$\mathbb{E}_{x\sim p}\phi_n(x) = \bar{\mu}_n$$

Moment Matching

$$P(x) = a(x)e^{\sum_i \lambda_i \phi_i(x)}$$

P(x) is an unnormalized form of the optimal distribution p(x) *i.e* an Energy-Based Model

Distributional Policy Gradients

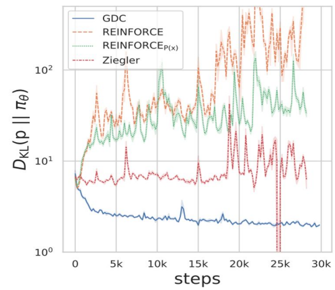


Locally normalized autoregressive policy for sampling

EXPERIMENTS

(1) Pointwise, (2) Distributional, (3) Hybrid Constraints

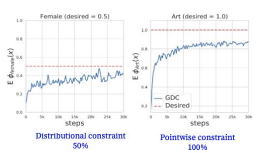
KL-Divergence from optimal policy p (↓ better)



$\mathbb{E}_{x\sim p}\phi_{she} = 0.5$

$\mathbb{E}_{x\sim p}\phi_{art} = 1.0$

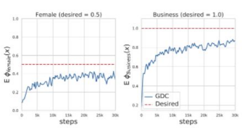
4		Desired	Before	After
	Female	50%	07.4%	36.6%
	Art	100%	11.4%	88.6%



$\mathbb{E}_{x\sim p}\phi_{she} = 0.5$

$\mathbb{E}_{x\sim p}\phi_{business} = 1.0$

5		Desired	Before	After
	Female	50%	07.4%	37.7%
	Business	100%	10.1%	82.4%



A **hybrid** experiment:

- Pointwise and distributional constraints together
- Application to **de-biasing** the pretrained LM

Paper: http://bit.ly/DIST-CTRL-ICLR2021 Code: https://github.com/naver/gdc