Generalization in Generative Adversarial Networks: A Novel Perspective from Privacy Protection

Bingzhe Wu[1], Shiwan Zhao[2], Chaochao Chen[3], Haoyang Xu,[1] Li Wang,[3] Xiaolu Zhang[3], Guangyu Sun[3], Jun Zhou[3]
[1] Peking University, [2] IBM Research, [3] Ant Financial

Introduction

The generalization ability of GAN is an interesting research topic.

Previous works are conducted on some specialized GANs:

The privacy concerns of deep learning algorithms are arisen recently

- Membership attack
- Model inversion attack
- …

We study the generalization of GANs from the view of privacy protection from both theoretical and experimental sides.

Theoretical Finding

Problem Setup:

$$\min \max E_{x\sim p_{data}}[\phi(d(x;\theta_d))] + E_{z\sim p_z}[\phi(1-d(g(z;\theta_g);\theta_d))]$$

We first analyze the information leakage of the discriminator using stability-based theory and differential privacy. We denote A as the Training algorithm for discriminator.

If A satisfies ϵ-differential privacy, then we obtain:

- (**Generalization Bound**) The generalization error of the discriminator can be bounded.
- **(Uniform Convergence)** Let $d^{(k)}\left(x;\theta_d^{(k)}\right)$ be the output of A at the k-th iteration. Then, the generalization gap of $d^{(k)}$ can be bounded.
- The privacy leakage of the generator can be obtained via composition theory of adaptive learning[1] .

Empirical Results

Empirically, we aim to validate the information leakage of various Lipschitz constrains on the discriminator.

- Results of threshold attack on LFW dataset

Strategy	F1	AUC	Gap
Original	0.565	0.729	0.581
Clip	0.486	0.501	0.113
Spectral	0.482	0.506	0.106

- Results of shadow training attack

Strategy	F1	AUC	Gap
Original	0.423	0.549	0.581
Clip	0.358	0.502	0.113
Spectral	0.347	0.497	0.106

Discussion

- Lipschitz is helpful for reducing the information leakage of GANs
- The privacy risk varies in different datasets

Dataset	F1	AUC	Std (R)
IDC	0.445	0.531	0.085
LFW	0.565	0.729	0.249

- The privacy leakage of the generator.
  - Using more advanced generalization notation[1], such as robust generalization.
  - Post-processing property

Ref

[1] Cummings et al. Adaptive learning with Robust Generalization Guarantees