Generative Modeling for Context-Aware Local Privacy
Wei-Ning Chen, Wei-Chen Chen, Dennis Rich

Problem Setup

Goal: erase sensitive info $S$ (smiling or not) from $X$ while minimizing distortion $d(X, \hat{X})$

- Privatization Channel
- $X$ → $\hat{X}$

Learning Algorithm

Objectives

\[
\min_W \mathbb{E}_{P_X, \tilde{X}} d(X, \hat{X})
\]
\[
\text{s.t. } \mathbb{P}_{\tilde{X}|S=s_1} W - \mathbb{P}_{\tilde{X}|S=s_2} W \leq \epsilon \quad \text{TV}
\]

Reformulate as GAN (dual form)

\[
\min_W \mathbb{E}_d(X, \hat{X}) + \lambda \left( \max_{f_\omega} \mathbb{E}_{P_X|s_1} f_\omega(\hat{X}) - \mathbb{E}_{P_X|s_2} f_\omega(\hat{X}) \right)
\]

Training Architecture

- Auto-encoder
- Privatizer
- CNN

- {smiling, not smiling}
- Sensitive label
- Discriminator

Loss function: pixel-wise MSE $- \lambda$ CNN’s cross-entropy

Results

- Ground Truth
- Privacy
- Utility

Conclusion and Future Works

- Proposed a data driven framework for context-aware privacy
- Achieved better privacy-utility trade-off with theoretical guarantee on privacy
- Can easily incorporate other utility measures
- What next: characterize the fundamental limits on privacy-utility trade-off curves

Related works

- Local differential privacy [1]: need data distribution, poor privacy-utility tradeoff
- Detect and perturb [2]: add random noise to sensitive parts. May not achieve minimum distortion.

Our method

- Decentralized: trust no one even data collector
- Data driven: do not need data distribution
- Optimize distortion given privacy constraint

References