

LINE、深層学習における世界トップレベルの国際学会「ICLR 2022」にて論文採択

2022.02.10 技術情報

プライバシー保護型データ合成に関する研究において既存手法と比較して優れた性能を実現したことへの高い評価

LINE株式会社（本社：東京都新宿区、代表取締役社長：出澤 剛）は、深層学習における世界トップレベルの国際学会「ICLR 2022」にて論文が採択されましたので、お知らせいたします。

「ICLR」は、The International Conference on Learning Representations (ICLR)が主催する深層学習における世界トップレベルの国際会議です。2022年は10回目の開催となり、採択された論文は2022年4月25日から29日にかけてオンラインで開催される「ICLR 2022」にて発表を行います。

■プライバシー保護型データ合成に関する研究が評価され採択

今回の「ICLR 2022」で採択された論文では、データ分析や機械学習の分野で近年注目されているプライバシー基準「Differential Privacy（差分プライバシー）^{*1}」を保証しながら生成モデル^{*2}を学習するフレームワーク「PEARL（Private Embeddings and Adversarial Reconstruction Learning）」について提案しています。従来の差分プライバシーを保証する深層学習（DP-SGD^{*3}等）では、訓練の度にデータへ繰り返しアクセスが必要な一方で、差分プライバシーの制約から訓練回数に制限があり、生成モデルを十分に訓練できませんでした。本フレームワークでは、機密データの特徴を抽出したエンベディングベクトル^{*4}を差分プライバシーが保証された形で事前に構築し、元のデータを再利用することなくエンベディングベクトルだけを利用して深層生成モデルの学習が可能です。そのため、従来手法とは異なり、訓練回数に制限がありません。本論文では、特性関数^{*5}によるエンベディングベクトルと敵対的な再重み付け^{*6}による目的関数を導入することで、本フレームワークを実現しました。実験による評価では、妥当なプライバシーレベル（プライバシーパラメータ ϵ が1以下）において他の手法よりも性能が優れていることを複数のデータセットで示しました。

LINEのプライバシー保護型データ合成に関する論文の採択は、データ工学分野のトップカンファレンス「ICDE2021」に続き2本目です。今後も引き続きデータサイエンス、AIにおけるプライバシー保護、信頼性の担保に関する研究開発を進めてまいります。

^{*1} 差分プライバシー：統計的なプライバシー基準。プライバシーパラメータ ϵ が小さいほど強いプライバシーが理論的に保証されていることを表す。

^{*2} 生成モデル：訓練データを模倣する人工データを生成する機械学習モデル。

^{*3} DP-SGD：訓練の際に勾配にノイズを加算することで、差分プライバシーを保証する技術。訓練の繰り返し回数が増えると ϵ が増大するため、訓練回数に制限が必要となる。

^{*4} エンベディングベクトル：データを構成する要素や特徴を所定の空間上で数値化して表現したベクトル。

^{*5} 特性関数：データ全体の振る舞いを表現する関数。低次元でも高い表現力を持つ。

^{*6} 敵対的な再重み付け：訓練が十分に進んでいない領域の重みを強調し、訓練を促進すること。

採択された論文

PEARL: Data Synthesis via Private Embeddings and Adversarial Reconstruction Learning

Seng Pei Liew, Tsubasa Takahashi, Michihiko Ueno (LINE AI Research)

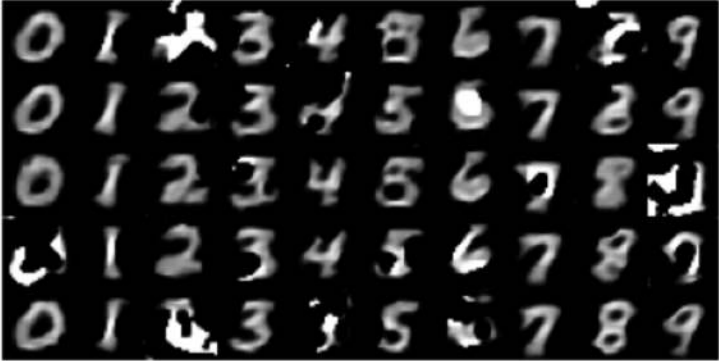
<https://openreview.net/forum?id=M6M8BEmd6dq>

● 実験による評価：妥当なプライバシーレベルにおける画像生成での提案手法と既存手法の比較

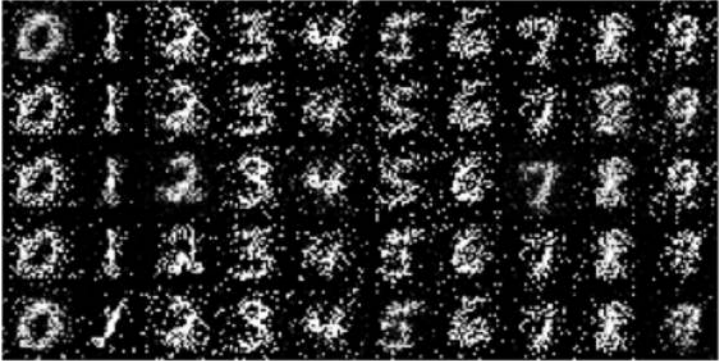
提案手法のほうが、より本物に近い画像の生成を実現しており、性能が優れていることを示しています。

数字の
画像データセット
(MNIST)

提案手法

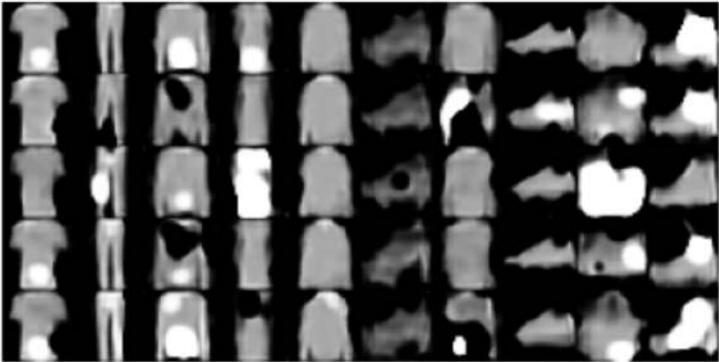


既存手法

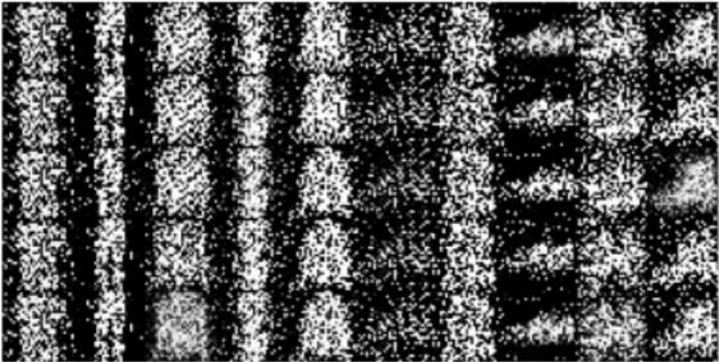


服・靴の
画像データセット
(Fashion MNIST)

提案手法

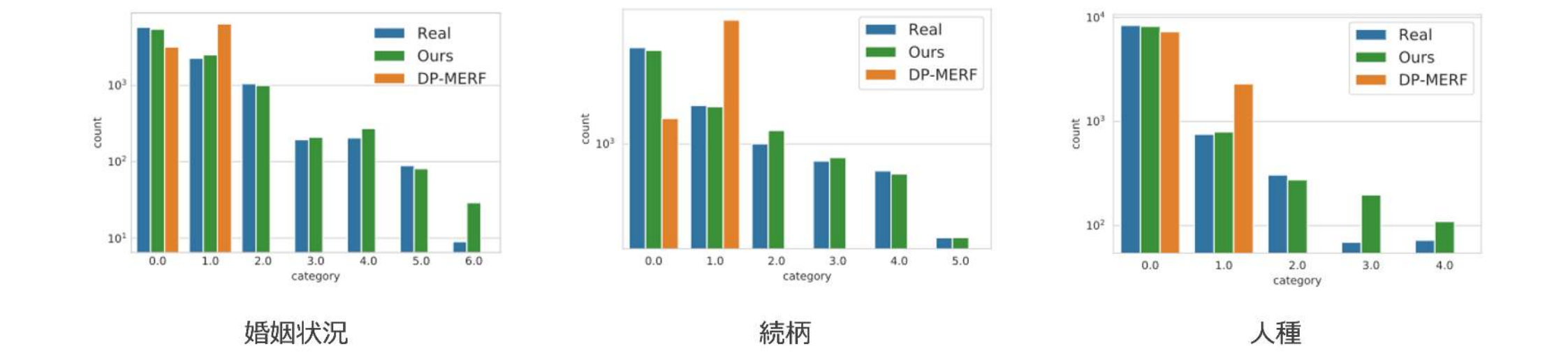


既存手法



提案手法（緑）のグラフの傾向のほうが、既存手法（オレンジ）よりも実データ（青）に近い傾向を表現しており、性能が優れていることを示しています。

成人の属性に関するデータセット (Adult)



【LINEが注力する基礎研究について】

LINEでは、AI事業を戦略事業の一つとして位置付け、AI技術の研究・開発およびAI技術を活用した事業の発展を加速させることを目的に、NAVERとの連携も行いながら、新たなAI関連サービス・新機能の創出を進めるとともに、それらを支える技術の基礎研究に注力しています。

基礎研究においては、機械学習を軸に、音声処理、言語処理、画像処理などに注力しています。2021年には、音声・音響信号処理分野の国際学会「ICASSP 2021」にて国内トップクラスの7本の論文^{*7}、コンピュータービジョン分野の国際学会「ICCV 2021」にて2本の論文^{*8}、音声処理分野の国際学会「INTERSPEECH 2021」にて国内トップクラスの6本の論文^{*9}が採択されるなど、着実な成果を収めております。

^{*7} 2021年2月26日発表プレスリリース：<https://linecorp.com/ja/pr/news/ja/2021/3639>

^{*8} 2021年7月28日発表プレスリリース：<https://linecorp.com/ja/pr/news/ja/2021/3843>

^{*9} 2021年8月30日発表プレスリリース：<https://linecorp.com/ja/pr/news/ja/2021/3891>

LINEでは、今後も事業発展やサービスの品質向上のための取り組みを積極的に行い、コミュニケーションプラットフォームとしてさらなる成長・拡大の可能性を広げてまいります。