

令和 3 年 6 月 3 日現在

機関番号：12601

研究種目：若手研究(A)

研究期間：2017～2020

課題番号：17H04693

研究課題名（和文）先端的生成モデルのための学習アルゴリズム基盤の構築

研究課題名（英文）Learning Algorithm for Advanced Generative Models

研究代表者

佐藤 一誠 (Sato, Issei)

東京大学・大学院情報理工学系研究科・准教授

研究者番号：90610155

交付決定額（研究期間全体）：（直接経費） 15,600,000 円

研究成果の概要（和文）：近年機械学習分野における生成モデルの発展は目覚ましく、現実的に存在しない人の画像や動画などが人の目には区別がつかないレベルまでになっている。しかし、そのような生成モデルの学習は、通常の識別モデルの学習とは異なり、目的関数の最適化の難しさに起因する学習の不安定性や生成物の評価の難しさに起因する定式化の非自明さなどが相まって、学習手法には多くのヒューリスティクスを要する場合が多い。本研究では、学習の不安定性の解明・緩和と既存研究ではほとんど取り扱われていなかった表形式のデータに関する生成モデルの開発と少数データ解析への応用に関して取り組みICMLなどの難関国際会議で発表した。

研究成果の学術的意義や社会的意義

生成モデルは通常の機械学習における識別モデルによる予測とはことなり、現実には存在しないデータを生成する。現在では、画像や動画などの生成において人が認識できないレベルまで到達しており、人が数時間数日かけて作成するようなものでも数秒で生成してしまう。これは人の創作活動において大きな変革をもたらす可能性を秘めている。しかし、その学習アルゴリズムは未だに解明できていないことも多く、特に学習の不安定性は問題で、学習アルゴリズムに対して多くのヒューリスティクスと試行錯誤が必要となる。本研究では、学習の安定性に関する基礎研究を行い、また生成モデルの適用範囲を表形式データに拡張し、少数データ解析へ応用した。

研究成果の概要（英文）：In recent years, the development of generative models in the field of machine learning has been remarkable, reaching a level where images and videos of people who do not exist in reality are indistinguishable to the human eye. However, learning such generative models differs from learning ordinary discriminative models in that it often requires many heuristics in the learning method due to the instability of learning caused by the difficulty of optimizing the objective function and the non-triviality of formulation caused by the difficulty of evaluating the products. In this research, we have worked on elucidating and mitigating the learning instability, developing a generative model for tabular data, which has not been dealt with in existing research, and applying it to minority data analysis.

研究分野：機械学習

キーワード：機械学習 生成モデル 学習アルゴリズム

## 1. 研究開始当初の背景

生成モデルとは、データの生成過程を統計的な定式化によって表現する数理モデルである。特に、潜在変数と呼ばれる確率変数によってデータの背後にある関係を潜在構造としてモデルに組み込み抽出することで様々なデータ解析に役立てることができる。

申請者は、およそ10年に渡り生成モデルの研究分野の最難関国際会議に恒常的に壇上し研究発表を行ってきた過程で、この10年における生成モデル研究の進歩は目覚ましいものがあるとリアルタイムに実感している。その大きな要因は、3つあると考えている。

1. データ数に劣線形の計算量の学習アルゴリズムが進展した。生成モデルを大規模データへ適用することで応用範囲が大きく広がった。
2. 深層生成モデルの学習が可能になった。深層生成モデルは多層の神経回路網によってデータの生成過程をモデル化するため、非常に複雑な構造をとり、変分推論のように効率的なアルゴリズムを導出することはできなかった。誤差逆伝搬法を用いた変分推論が提案され、近年の深層学習で用いられる自動微分と GPU の親和性から効率的に深層生成モデルが学習できるようになった。
3. 敵対的学習という生成モデルの新しい学習パラダイムが開発された。従来の生成モデルの学習は、仮定したモデル空間の中でデータの経験分布に近いモデルを探索するアルゴリズムであった。敵対的学習では、「経験分布から生成されたデータ(現実のデータ)」と「生成モデルから生成されたデータ」を識別する識別器 D (Discriminator) を用いて、識別器 D を騙すような生成モデルを構築するアルゴリズムである。敵対的学習による深層生成モデルから生成される画像は、従来手法に比べ精度が高いことが実験的に知られている。

しかしながら、上記2及び3の要因に対して、それぞれ重要な問題を残していると考えられる。本研究ではそれぞれの問題点を指摘し解決策を提示することで、生成モデル研究をより一段高いレベルに押し上げることを目的とする。さらに、より高精度な生成モデルの学習が可能になったことを用いて、希少データ解析への応用を行う。

## 2. 研究の目的

近年、変分推論と誤差逆伝搬法を組み合わせるための再パラメータ化トリック (reparameterization trick) という方法が提案され、深層生成モデルの学習が可能になった。しかし、再パラメータ化トリックが可能な潜在構造は連続変数に限られており、離散構造を扱う場合には適用できない。離散的な潜在構造は、生成モデルの表現力と解釈性という意味で重要である。連続的な構造を仮定する場合にはデータ間の関係にも連続性を仮定することになり、クラスタ構造のように生成確率が多峰になるような構造の生成モデルを学習することが難しくなる。また、潜在変数の離散性は潜在構造の解釈性の観点からも重要である。すなわち、ある潜在特徴量がどの程度の量なのかという連続的な情報ではなく、ある潜在特徴量の有無でデータが表現されるため、人が解釈しやすいものとなる(例 Google の猫)。本研究では、潜在構造が離散的であっても変分推論と誤差逆伝搬法で効率的に学習可能なアルゴリズムを開発する。

また、現在の生成モデルの研究の多くは画像形式のデータである。機械翻訳も含めて文、音声・音響といった形式の生成モデルも研究されているが、一般的な表形式のデータの生成モデルはこれまで提案されてこなかった。画像形式などマルチメディアデータとは異なり表形式のデータは実は様々な企業でデータベース化され利用されているため表形式のデータの生成モデルの適用範囲は広い。表形式データの生成モデルが構築できると、画像形式のデータで良く用いられているデータ拡張などが適用可能になる。また、データそのものの背後にある情報を抽出するのに役立つ。特に、本研究では、少数データ解析において重要なドメイン適応と呼ばれる応用を考えて表形式の生成モデルを考える。

### 3. 研究の方法

連続潜在変数では、解析的に解くことが困難な積分計算を、再パラメータ化を用いた潜在変数のサンプリング近似により誤差逆伝搬法での勾配を計算できる。しかし、離散潜在変数の場合、近似した目的関数は微分不可能となるため同じ方法が使えない。本研究では、局所期待勾配と再パラメータ化を組み合わせることでこの問題を解決した。具体的には、「局所期待値勾配を計算する際の微分と積分の順序交換をいかに正確に近似できるか」という問題に帰着し、再パラメータ化サンプリング近似を微分と積分の順序交換手法の1つとした新たな見方を与えることで改善した。この新たな視点により共通乱数法を用いることで、局所期待勾配の分散が最小になるという理論保証を与えることもできた。

次に少数データ解析への生成モデルの応用としてドメイン適応を考える。ドメイン適応とは異なるドメイン間で情報を共有することで、特定のドメインでデータが少ない場合にも効率的に学習する分野である。本研究では、従来の確率モデルによる生成モデルに対して、「生成メカニズム」と呼ばれる属性間の依存関係を考え、生成メカニズムはドメイン間で共通であるという仮定の下でドメイン適用する手法を提案した。これにより、データ数の少ないドメインにおけるデータ生成をドメイン適応させることでデータ拡張し、拡張されたデータで学習することで、適応先ドメインでの学習器の性能を向上させることができる。提案手法は、表形式のデータ全般に適用可能な一般的な手法であるため、この生成モデルを用いてデータ拡張したデータは、機械学習の任意アルゴリズムを使って学習することができたため応用範囲も広いと考えられる。

### 4. 研究成果

機械学習の最難関国際会議に複数の論文が採択されている。

生成モデルの学習に関する論文は、ICML2017に採択されている。

生成モデルを画像データから表形式のデータに応用範囲を広げ、少数データ解析への応用した研究に関する論文はICML2020に採択されている。

またその他、付随する問題として生成モデルに限らずニューラルネットワーク全般に関する学習アルゴリズムのロバスト性に関する研究を行っておりNeurIPS2018, AAAI2019に採択され、継続研究としてシミュレーションに機械学習を用いる際のロバスト性を向上させる研究がAISTATS2021に採択された。

## 5. 主な発表論文等

〔雑誌論文〕 計2件（うち査読付論文 2件 / うち国際共著 0件 / うちオープンアクセス 2件）

1. 著者名 Takeshi Teshima, Issei Sato, Masashi Sugiyama	4. 巻 -
2. 論文標題 Few-shot Domain Adaptation by Causal Mechanism Transfer	5. 発行年 2020年
3. 雑誌名 Proceedings of Thirty-seventh International Conference on Machine Learning	6. 最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 Seiya Tokui and Issei Sato	4. 巻 70
2. 論文標題 Evaluating the Variance of Likelihood-Ratio Gradient Estimators	5. 発行年 2017年
3. 雑誌名 Proceedings of the 34th International Conference on Machine Learning	6. 最初と最後の頁 3414-3423
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

〔学会発表〕 計4件（うち招待講演 0件 / うち国際学会 4件）

1. 発表者名 Takeshi Teshima
2. 発表標題 Few-shot Domain Adaptation by Causal Mechanism Transfer
3. 学会等名 International Conference on Machine Learning（国際学会）
4. 発表年 2020年

1. 発表者名 Futami, F., Cui, Z., Sato, I., & Sugiyama, M.
2. 発表標題 Bayesian posterior approximation via greedy particle optimization
3. 学会等名 Thirty-Third AAAI Conference on Artificial Intelligence (AAAI2019)（国際学会）
4. 発表年 2019年

1. 発表者名 Tsuzuku, Y., Sato, I., & Sugiyama, M.
2. 発表標題 Lipschitz-margin training: Scalable certification of perturbation invariance for deep neural networks.
3. 学会等名 Advances in Neural Information Processing Systems 31 (国際学会)
4. 発表年 2018年

1. 発表者名 Seiya Tokui
2. 発表標題 Evaluating the Variance of Likelihood-Ratio Gradient Estimators
3. 学会等名 International Conference on Machine Learning (国際学会)
4. 発表年 2017年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関