

研究終了報告書

「Energy-based Model による人と相補する生成モデルの開発」

研究期間：2021 年 10 月～2024 年 3 月

研究者：幡谷 龍一郎

1. 研究のねらい

本研究は当初、Energy-based model と呼ばれる生成モデルに着目して、人の知識を利用した制御がしやすい新たな生成モデルを開発することを目標としていた。この生成モデルは、学習にニューラルネットワークが絡んだ二重の最適化問題が含まれ、その計算が不安定になることが課題であった。しかし、本研究期間中に、拡散モデル、自己回帰モデルなどの他の生成モデルの研究は世界中で飛躍的に進み、当初目標としていたような、人が制御しやすいデータ生成は実用レベルに達し、研究者に限らない多くの人々が使うものとなっている。そのため、このように新たに発達した生成モデルが将来の機械学習に対して及ぼす影響についての研究も行った。

2. 研究成果

(1) 概要

本研究の成果として、勾配法による二重最適化の効率的な近似を行う手法を提案【主な研究成果リスト 1】、巨大生成モデルの生成データが将来のコンピュータビジョンに及ぼす影響を実験的に評価【主な研究成果リスト 2】が挙げられる。

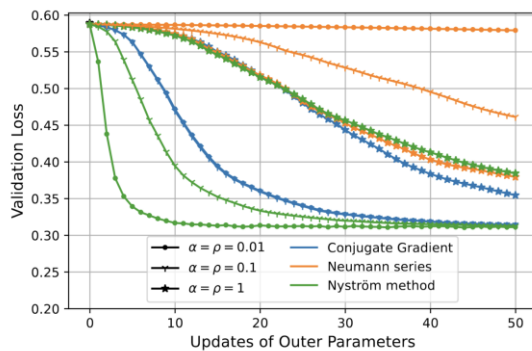
(2) 詳細

【主な研究成果リスト 1】について

本研究課題が当初対象としていた Energy-based model では、二重最適化問題（ハイパーパラメータ最適化問題）が現れるが、特にニューラルネットワークが関わる場合、その計算コストは大きく、適切な近似を用いる必要がある。そこで、本研究成果では勾配法による深層学習のハイパーパラメータ最適化の効率的な近似を行う手法を提案した。

勾配法によるハイパーパラメータ最適化を陰関数微分を利用して効率的に行うためには、ニューラルネットワークモデルのヘッセ行列の逆行列が必要となるが、深層学習などを考慮する場合、ニューラルネットワークモデルのパラメータ数が多いため、そのヘッセ行列の保持すらも難しい。既存の研究ではモデルのヘッセ行列ベクトル積と、共役勾配法などの一般的な線形方程式の逐次的数値解法を利用してヘッセ行列の逆行列ベクトル積を近似していた。一方、我々はニューラルネットワークのヘッセ行列は実質的に低ランクであるという性質に着目した。固有値が急に減衰するような低ランク行列に対して、既存法の用いていた共役勾配法などの利用は適切ではない。

そのため、我々は Nyström 法によるヘッセ行列の低ランク近似を用いた逆行列のベクトル積の推定法を提案した。本手法は高速で、ニューラルネットワークモデルの性質を利用しているため、様々なハイパーパラメータ最適化の問題において、既存法よりも安定して高い性能を発揮する。さらに設定パラメータの選択に対して鈍感であるという実用的な長所がある。左図は提案法 (Nyström method) を既存法 (Conjugate Gradient, Neumann series) と比較しており、既存法の α 、提案法の ρ という設定パラメータを変えても提案法が安定して速く最適化ができる



ていることを示している。

研究成果は機械学習分野の難関国際会議である AISTATS 2023 に採択された。また、本手法はソフトウェアライブラリとして公開した(<https://github.com/moskomule/hypergrad>)。

【主な研究成果リスト2】について

本研究期間中に起きた、提案時には想像もつかなかった事態として、生成モデルの飛躍的な発達、および一般への普及が挙げられる。特に一般への普及の効果は著しく、これまで基本的に研究者だけが行っていた生成モデルの開発に、エンジニアやデザイナーなど幅広い人材が参入したことで、本研究課題が目標としていた、人の知識を利用した制御がしやすい生成モデルが次々に生まれ、実用されている。一方、例えば一見実データと区別がつかない、高品質な画像データが可能な生成モデル(たとえば Stable Diffusion や ChatGPT)が一般に普及したことで、インターネット上には生成データが溢れるようになった。

実際の画像



生成画像



【主な研究成果リスト2】では、このような時代の流れに着目し、データセット構築時に生成データが混入することが将来のコンピュータビジョンに及ぼす影響を実験的に議論した。2022 年頃から、Stable Diffusion や MidJourney をはじめとする、テキストの指示文から高品質な画像生成を行える生成モデルが一般にも普及した。その結果、一見実際の画像と区別できないような生成画像がインターネット上に大量にアップロードされるようになった(左図)。一方、このような生成モデルをはじめとして、深層学習モデルの高い性能はインターネット上から大量のデータを収集したデータで学習していることに大きく依存している。

生成モデルは日々大量の画像を生成しているため、今後収集するデータセットには多くの生成データが含まれることとなるが、このことは将来のコンピュータビジョンにどのような影響を及ぼすのだろうか？

具体的には当時最新の画像生成モデルであった Stable Diffusion を用いて、ベンチマークデータセットを模倣した画像を 300 万枚程度生成し、この生成データが混入したベンチマークデータセットを訓練データとしてモデルを学習する。その結果、画像分類や画像説明文生成をはじめとする、さまざまなコンピュータビジョンタスクにおいて生成データの混入が性能劣化をもたらすことが分かった。さらに、生成データの混入がモデルの分布外汎化能力をも損ねることが分かっており、実用上の懸念となり得る。これまで述べたような悪影響の一因として、左図に示したように、生成データが実際の画像と比較して一部の領域に集中してしまっていること、つまり、多様性の欠如が挙げられる。生成の品質と多様性にはトレードオフの関係があるため、今後もこのような状況は改善しない可能性が大きい。本研究成果はコンピュータビジョン分野の最難関国際会議である ICCV2023 に採択された。生成したデータセットは今後の研究のために公開した(<https://github.com/moskomule/dataset-contamination>)。

3. 今後の展開

【主な研究成果リスト1】で扱ったような勾配法による二重最適化手法は、高速にハイパーパラメータ最適化を実現する方法として有望だが、一方で十分に普及しているとはいえない。今後は手法開発にとどまらず、使いやすいソフトウェアの開発を通して、技術の普及と発展を図っていきたい。

【主な研究成果リスト2】で判明した生成データの学習データへの混入による影響は今後顕在化していく。そこで、影響の分析や緩和方法の研究を今後数年は進めていく。

4. 自己評価

本研究期間中には、博士論文執筆や異動といったライフイベントや、急激な生成モデルの進化があり、当初予定していたような生成モデルの開発は、【主な研究成果リスト 1】に必要なパーツを提案するに留まった。一方で、【主な研究成果リスト 2】では、生成データによる将来機械学習への悪影響の分析という重要な問題を先取りし、世界的にも先駆的な研究ができた。

また、本研究を通して、本領域の多くのトップ研究者と親密な関係を築くことができたため、今後の研究に繋がることを期待している。

5. 主な研究成果リスト

(1) 代表的な論文(原著論文)発表

研究期間累積件数: 3 件

1. Hataya, Yamada, “Nyström Method for Accurate and Scalable Implicit Differentiation,” AISTATS 2023.
2. Hataya, Bao, Arai, “Will Large-scale Generative Models Corrupt Future Datasets?” ICCV 2023.

(2) 特許出願

研究期間全出願件数: 0 件(特許公開前のものも含む)

(3) その他の成果(主要な学会発表、受賞、著作物、プレスリリース等)

講演

“Towards accurate and scalable gradient-based hyperparameter optimization,” Seminar at Istituto Italiano di Tecnologia, May, 2023.

“Stable Gradient-based Hyperparameter Optimization,” RIKEN AIP & NCU Workshops 2023, September 2023.