公開

（事後評価・研究者用）

研　究　報　告　書

「安全なAIこそ効率的：ロバスト学習による汎化性能向上の研究」

研究期間：　平成29年10月～平成31年3月
研究者番号：　50154
研　究　者：　ホーランド　マシュー　ジェームズ

1.　研究のねらい

「人工知能」(AI)を目指す機械学習(ML)技術は、音声や物体の認識、テレビゲーム、機械翻訳、商品推薦、自動運転技術といった幅広い応用実績が世間の注目を集めている。期待が産学官の各界で高まるなか、重要度の高い応用問題で開発が難航している。例えば、個人向けの在宅医療診断、言動による感情分析、介護ロボット制御をはじめ、困難な学習課題が多く指摘されており、医療・交通・エネルギーといった領域ではすでに、現行の技術の限界が見えはじめている。また、性能が不安定で実際の働きが予想しにくいと、試行錯誤が避けられず、設計者の監視も不可欠なため、開発の進歩が遅い。この現状を踏まえて、次世代の学習課題が克服できなければ、人工知能技術の更なる発展や社会基盤技術としての普及は見込めない。

この問題の要因を俯瞰すれば、まずデータの性質の変化に注目すべきであろう。インターネットや産業界の各種センサー技術の発展により、膨大で複雑なビッグデータが台頭し、丹念に精査することも整理することも現実的でなく、その作業自体も自動化の対象となる現状である。そのプロセスにおいて、一つの段階でも観測データと学習方法の相性が悪ければ、それによる誤差があとの作業に伝播され、学習の最終結果まで尾を引いてしまう。数理モデルが現実をうまく表現できていないことも一因であるが、それよりも根源的な問題として提案者が着目するのは、パラメータの自動設定を担う学習アルゴリズムの適応能力の欠如である。データサンプルのばらつきに揺さぶられない学習方法、高い安全性と自律性を持つ学習システムといった未来の AI 技術に向けて、事前情報無しで多種多様なデータに適応する能力が必要条件である。

本研究では、ML 技術のコア部分を見直し、安定性を最優先にした制御方法を掲げる。既存の学習機は、センサー等による観測データに敏感に反応するため、極端な働きを防ぐことが難しい。試行錯誤が避けられないのはこのためである。提案者は、様々な刺激に対して頑健な学習機を設計すれば、開発時間の短縮と安全の確保を一度に実現できると考える。本研究は設計方法の確立と性能の実証的評価を最大の目標とし、世界の AI 開発を加速させる基礎技術を目指す。

2.　研究成果

(1) 概要

上記の「本研究の狙い」で述べたように、本研究の主な目的は、端的にいえば、専門知識と経験則を要しない汎用的学習アルゴリズムの開発と性能検証である。具体的には、以下の条件を満たす学習機の設計法を目指してきた。

(a)　弱い仮定でも成り立つ統計的性能保証
(b)　安易な実装

(c) 限りなく小さな計算オーバーヘッド

学習アルゴリズムやモデルの微調整による試行錯誤を減らして、明確な信頼性保証つきの AI 基盤技術を刷新することを究極の目標として掲げてきた。この ACT-I の研究期間中、理論的な土台を踏まえて実用性の高い学習アルゴリズムを新たに提案し、またその深層学習への実装枠組みを新たに開発した。この手法を発表した論文は機械学習のトップカンファレンスの一つである AISTATS (2019)に採録決定され、データに対して特に何も仮定しない（＝事前知識を必要としない）状況下では従来の手法を大きく凌駕する理論的な根拠と計算効率のバランスを実現した学習アルゴリズムとして評価された。根幹にある新規性の高い学習則に加えて、近代の AI ブームを支えてきた深層学習への拡張フレームワークもツールとして提供できるところまでたどり着いたことから、先述の目標に向けて確かなる進歩であるといえる。

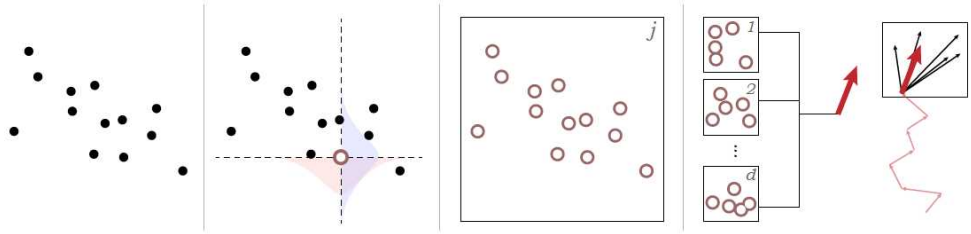
（2）詳細

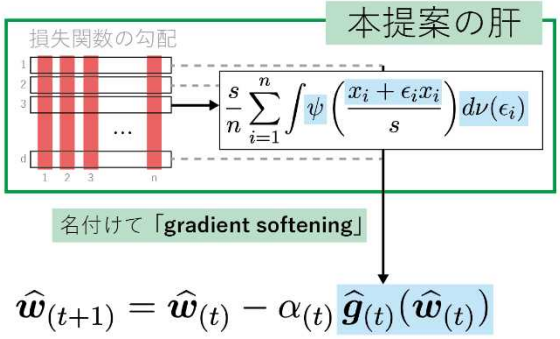
ステージ 1:学習アルゴリズムの開発

（応答方式の設計、フィードバック方式の設計）

この研究ステージでの成果の要点は下記の通りである。

- **データ分布に頑健な勾配推定方法**
  ノイズをほぼゼロタイムで考慮する最急降下法やそれに近いアルゴリズムも文献上なく，性能解析法も新しい．
- **低コストで高度なロバスト化を図る算法**
  ロバスト化勾配は陽に表せるため，計算量が少なく，並列化も簡単なため，次元ごとの処理でも速い．



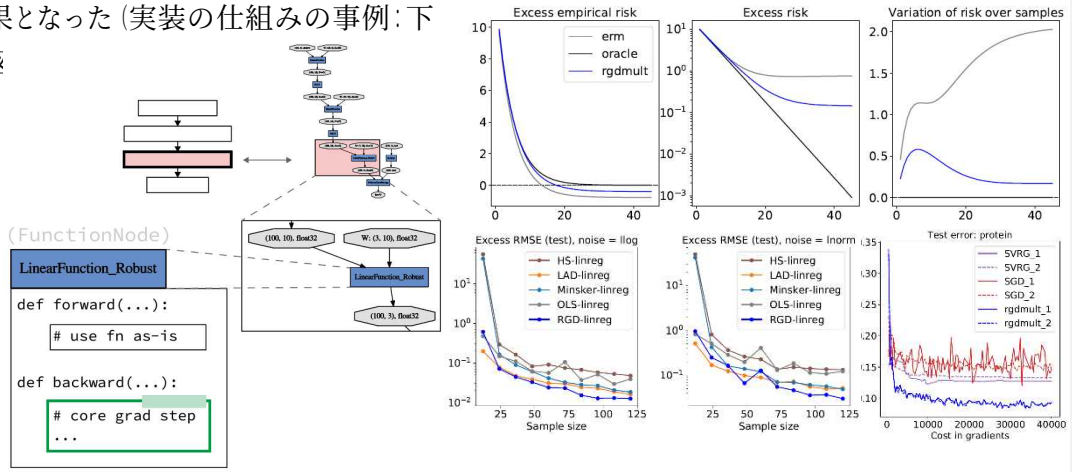


アルゴリズムの中枢をなす更新式を上図と右図のように図示した。基本的な考え方は、ノイズが統計的推定をどのように乱しうるか、それを事前に考慮し、学習前に推定量に組み込むことである。これにより、学習時には従来の標本平均とほとんど変わらない程度の計算量で、大幅な頑健性向上を実現することが可能である。換言すれば、従来の勾配ベクトル推定に、自ら設計したノイズをかける影響をあらかじめ解析的に表現できる工夫が最大のポイントとなっている。この技法の開拓と勾配降下法への適用が本研究の大きな技術的･方法論的な貢献である（原著1）。

ステージ 2:学習機の実証的性能検証
(数値シミュレーション、ビッグデータ応用、スモールデータ応用)
この研究ステージの成果の要点は下記の通りである。

- **深層学習向けの応用として, 既存 API を生かしたデモ**
  完全な Chainer 上の実装例, 数値実験等のコードを公開.
- **入念な実験的検証**
  性能保証の裏付けとして, 人工+実データで頑健性を確認.

先述の新しい技法では、損失値は使わず、勾配情報のみで構成されるため、従来の最急降下を多少書き換える程度で簡単に実装できる。また、逆伝播法をはじめとして、勾配情報を中心にパラメータ推定を行う深層学習の種々の手法での応用も和合性が高く、本研究では Chainer という著名な深層学習 API を用いて、完結した実装例とその数値的な性能検証のデモンストレーションをインターネット上で公開し、深層学習の新たな方向性を示唆する結果となった(実装の仕組みの事例:下図



さらには、提案学習アルゴリズムを対象に、緻密に設計された数値実験を行い、学習効率(=計算資源 vs. 汎化誤差)を検証した。機械学習の業界標準となるベンチマークデータに加えて、実に多種多様な学習問題を数値シミュレーションにより体系的に構築し、学習課題の条件の変化が提案手法の学習効率にもたらす影響を明らかにした(上図、右)。この成果の意義については、次のような新しい知見が得られた。一つ目に、理論上の性能保証は学習アルゴリズムにとってもっとも不都合のデータを想定しているため、担保されている性能よりも実際の性能が良い場合もあり、従来の手法との優劣の差が実際に出ない可能性が原理的にはあるが、数値実験の結果、理論的に「優位であるはず」だという条件はもれなく優位であったことから、本研究で指摘している従来手法の脆弱性は無視できないものであることがわかった。二つ目に、理論的な性能保証は確率モデルの仮定をいくつか利用しているが、現実世界ではこの仮定が成立しないことは当然考えられる。本研究の数値実験では、仮定の緩和が提案手法の性能に及ぼす影響を調べた結果、重要な仮定を相当弱めても、提案手法の安定性も相対的優位性も変わらなかった。具体的には、真の確率分布の分散につい

ての事前情報がある前提での性能保証が、一切この事前情報を持たずにデータから推定した場合でも、事前情報を用いた最良設定の場合と比較して多少の性能低下こそあれ、従来の学習則よりもはるかに広い確率分布のクラスでの高い汎化能力が確認できたことから、提案手法の実際の性能は確率モデルの仮定に対して頑健であることがわかった。

3. 今後の展開

研究開始時点から、「学習機に学習を任せる」ことを可能にすることの重要性を主張してきた。また、その技術的なカギを握るのは、ロバスト性の高い統計的推定方法にほかならないと提唱し、独自の理論研究を土台に、新たな学習アルゴリズムの枠組みを構築してきた。先述の通りに研究成果はあったが、その成果を踏まえて新しい課題も多く見えてきた。代表例として、産業界に浸透しつつある深層学習では無数のパラメータを有するグラフが使われているが、すべてのパラメータの推定を学習開始時点から終了時点まで均一にロバスト化するという方法は、必ずしも最適とはいえない。ロバスト化は統計的推定の精度を上げ、学習の安定化にも寄与するが、計算コストがかかる上に、いささかバイアスを推定結果に加えてしまう。学習しようとしているモデルの構造と学習用のデータの特性を総合的に考慮し、学習過程においてどのタイミングで、モデルのどの部分を対象にロバスト化を実行すべきか、計算資源と標本外の性能を明示的に扱った新しい方法論が求められる。

4. 自己評価

全体的には、本研究が順調に進んだと考えている。研究開始時点では「高速かつ十分にロバスト」といえる手法は文献にはなく、新しい手法を独自の切り口で開発することが本研究の最大の目標の一つで、それを提案の「soft gradient」による降下法で実現できたことは評価できると考える。新規性は高く、トップレベルの国際会議に出した初版でも査読者から高い評価を受け、手法の方向性の確かな手ごたえを得た。また、ニューラルネットワーク向けの実装方法やそれを踏まえた数値実験の設計では技術的な苦難はたくさんあったが、最終的には Chainer を母体とした実装例が一般向けにもわかりやすく、多くの利用者がプロトタイプづくりに使える形が仕上がったことも満足している。反省点としては、GPU サーバを納入するにあたって、学内の事務手続きが想定していたよりも相当長く、GPU を生かした本格的な数値実験の実施が大幅に遅れてしまった。その決裁・納入を待つ間は小規模の数値実験を現有物で行うことができたので、研究自体は 1.5 年を通して絶え間なく続けることができたが、高額の設備導入の時間的コストを今後の研究計画により正確に反映させていこうと考えている。

5. 主な研究成果リスト

(1) 論文 (原著論文) 発表

1. Holland, Matthew J. Robust descent using smoothed multiplicative noise. Proceedings of Machine Learning Research (AISTATS2019, to appear). 2019.
2. Holland, Matthew J. Classification using margin pursuit. Proceedings of Machine Learning Research (AISTATS2019, to appear). 2019.

(2) 特許出願

該当なし。

(2) 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)

Holland, Matthew J. Leveraging Uncertainty to Robustify Deep Learning Algorithms. 28th Annual Conference of the Japanese Neural Network Society (JNNS2018). ポスター発表. 2018.

Holland, Matthew J. Fast robustification of gradient descent using multiplicative noise. 21st Information-Based Induction Sciences Workshop (IBIS2018). ポスター発表. 2018.