2023 年度年次報告書
信頼される AI の基盤技術
2021 年度採択研究代表者

ホーランド マシュー ジェームズ

大阪大学 産業科学研究所
助教

学習過程における価値観の多様化と性能保証の両立

研究成果の概要

2 年半にわたる研究活動の進展と課題について報告する。

本研究の軸となる「汎化指標」（今は「学習指標」＝「learning criteria」と呼んでいる）の性質、既存概念と対比しての新規性、着目したい範囲などがまとまり、新しい学習アルゴリズムを設計する土台が整った。この土台の大半は前の年度には見えていたが、2023 年度のサーベイ論文の完成版と採録 [1]をもって、私の世界観と設計法の基盤が確立したと言える。元々は JAIR 向けのジャーナル論文として提出して採録されたが、JAIR と連携する国際会議 AAAI 2024 のジャーナルトラック主催者から誘いを受けて、oral session でこの論文に関する講演 [2]をした。

本研究は単なる指標の設計ではなく、指標と学習アルゴリズムの関係性に重みを置いて、多種多様のニーズに応えるためのアルゴリズムづくりを追求している。私の提案指標クラスに駆動される確率勾配法というもっともオーソドックスな応用を見据えた学習アルゴリズムの性質を調べた第一号の論文は 2023 年度中の AISTATS 2023 で発表した [3]が、これを皮切りに、私のさきがけ研究でのアルゴリズム開発は大きく前進した。先述の AISTATS 2023 論文では、先行研究の代表的なリスク関数族である OCE や DRO を念頭に、その「非単調版」という位置づけを狙って目的関数のクラスを絞っているが、OCE 型の「閾値」を学習中に最適化するのではなく、学習前に固定することで、大きく異なる挙動をもたらすことができる。簡潔にいうと、任意の損失関数を所与として、学習中の出来栄えを「何らかのリスクを最小化せよ」ではなく「損失分布を特定の位置に集中せよ」という指令を出すことができる。これは期待損失を特定の位置に持っていこうとするいわゆる Flooding 法 (Ishida et al., ICML 2020)の滑らかな進化版と言えて、低コストで高い安定性、暗示的な正則化、何より平均損失の優れた汎化性能をもたらす効果があることがわかった。本提案を紹介する論文の初稿 [4]ができて、現在投稿中である。この延長線上の実験とアルゴリズムの開発も進んでおり、執筆中の論文がほかにもあるので、2024 年度には単なる汎化性能の向上だけではない価値観を捉えるためのアルゴリズムの成果を発表することができる見込みである。

最後に、本研究では「学習アルゴリズムを操作するユーザーの意思決定の支援」という観点も重要視しているが、「何が大事か」というのはデータやモデルの性質だけでは決まらず、ユーザーの意思に依存する部分が必ず存在するため「自動化」は難しく、アルゴリズム側としてどこまで人の意思決定に介入すれば良いかというのが私にとって課題であった。そこで一旦、自動化ではなく、「様々なゴールに到達するための地図」というイメージを持って、学習時の目的関数としては使えないが評価時に算出できる種々の指標をゴールとしていかに学習アルゴリズムを造るか、いわゆる surrogate loss を念頭にした汎化指標の柔軟化に力を入れることにした。損失の期待値という大前提であれば特に分類界隈の理論と実用的な手法はほぼ確立しているが、「期待値ではなく、分布の集中を促す指標」という価値観に変わった途端、従来の surrogate 理論の知見が通じなくなることがわかった。同時に、たとえば「DRO や CVaR を狙う」と思って surrogate loss を DRO や CVaR の変換に渡しても、分類の評価によく使うゼロイチ損失では結局 DRO や CVaR の最小化は期待値の最小化と同等であり、「設計と意図が噛み合わない」典型例と言える。これを「意図しない criterion collapse」と名付け、本研究の指標を取り入れることによって回避できることを示している。この現象は 2022 年の領域会議の時点である程度はわかっていたので軽く発表時に紹介したが、実際の論文化は 2023 年度の終盤に入ってからようやく着手し、完成した初稿 [5]を ICML 2024 に

投稿し、すでに採択通知を受けている。この論文では collapse の回避に関する知見は二値分類に限定しているが、発想自体は多クラスへと拡張できる可能性が高く、次年度に取り組む有望な研究の一つでもある。

【代表的な原著論文情報】

1) A Survey of Learning Criteria Going Beyond the Usual Risk. Matthew J. Holland and Kazuki Tanabe. Journal of Artificial Intelligence Research, 78:781-821, 2023.
2) A Survey of Learning Criteria Going Beyond the Usual Risk. Matthew J. Holland and Kazuki Tanabe. AAAI 2024 (Journal track), Vancouver, Canada.
3) Flexible risk design using bi-directional dispersion. Matthew J. Holland. AISTATS 2023. Proceedings of Machine Learning Research 206:1586-1623, 2023.
4) Implicit regularization via soft ascent-descent. Matthew J. Holland and Kosuke Nakatani. Preprint (arXiv:2310.10006), 2023.
5) Criterion collapse and loss distribution control. Matthew J. Holland. ICML 2024 (accepted, to appear).