信頼される AI の基盤技術
2021 年度採択研究者

2021 年度
年次報告書

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

§1. 研究成果の概要

初年度の半年間の主な成果として、本提案の中軸をなす「汎化指標」の新提案が大きく前進した。まずは統計的機械学習やその周辺分野を対象とした入念な学術文献の調査を行い、期待損失以外の汎化指標の特徴と偏りをより的確に把握できるようになった [1)]。損失分布が非対称の場合、そのばらつきの「向き」によって既存の汎化指標の反応が大きく変わる。ある方向にはきわめて敏感であるのに対して、その逆方向のばらつきはほとんど捉えられず、CVaR をはじめとする OCE リスク関数族や DRO リスクなど、いずれも実質的には「期待値＋右方向のばらつき」という表現しかできない。これを補完する方策として、初年度では、損失分布の分位値の凸計画としての特徴づけを緩和し、ばらつきを測る際の中心点とスケーリング、外れ値への感度などを自由にコントロールする汎化指標クラスを提案し、その基本的な性質およびそれを各種の学習アルゴリズムに導入した際の性能を入念に解析し、理論と実験それぞれの初期成果が出始めている段階である。重要なポイントとして、

- 「学習アルゴリズムの新奇なる挙動をもたらす能力」と「汎化指標そのものの統計的な解釈しやすさ」の両面から既存の汎化指標の盲点を補完できる
- 学習前の汎化指標の調節によって、非常に複雑で不確実な学習過程を経てもなお、テストデータにおける損失分布がある程度「予想通りの性質」を有する状況を突き止めている

という 2 点を挙げる。初期成果をまとめた論文はすでに完成しており [2)]、その初版は 3 月に arXiv に掲載するとともに、ソフトウエアのリポジトリも GitHub 上で公開している [3)]。要点をまとめた要約版は機械学習の国際会議に投稿中である。

当然ながら、本提案は汎化指標単独では成り立たないが、適応性とロバスト性を兼ね備えた学習アルゴリズムの開発と透明な汎化指標デザインを結びつける方法論の構築へ向けて、初年度の取り組みによって堅固な土台を築くことができた。

【代表的な原著論文情報】

1) Designing off-sample performance metrics. Matthew J. Holland. Preprint (arXiv:2110.04996), 2021.

2) Risk regularization through bidirectional dispersion. Matthew J. Holland. Preprint (arXiv:2203.14434), 2022.

3) bdd: bidirectional dispersion and risk function design. https://github.com/feedbackward/bdd