Press Release

2024 年 4 月 23 日
東京工業大学

効果的な A/B テストのための
オフライン評価性能検証指標の新規提案

－学習・評価・検証を行うオープンソースソフトウェアを無償公開－

【要点】

○データを用いて意思決定アルゴリズムを評価するオフライン評価における新たな検証指標を提案。

○従来の性能検証指標では、方策を選択する際のリスクが評価できていないことを解明。

○提案指標の実装の公開に加え、オフラインでの方策学習や評価、さらには評価手法の性能検証までを一貫したプラットフォーム上で実装するためのオープンソースソフトウェア「SCOPE-RL」を無償で公開。

【概要】

東京工業大学 工学院 経営工学系 中田和秀教授、小林健助教、同 清原明加大学生（研究当時）、岸本廉大学生、川上孝介大学院生（兼株式会社博報堂テクノロジーズ勤務）、コーネル大学の齋藤優太大学院生（兼半熟仮想株式会社所属）らの研究チームは、データを用いて意思決定アルゴリズムの性能を評価する**オフライン評価**（用語 1）の研究において、新たな性能検証指標を開発した。既存の性能検証指標が正確性のみに注目していたがために、実運用時のリスクを適切に評価できていないことを発見し、新たにリスクとリターンのトレードオフを評価する性能検証指標を提案した。また、実運用に即した手順でオフラインでの方策学習や評価、さらには評価手法の性能検証までを一貫したプラットフォーム上で実装するためのオープンソースソフトウェア「SCOPE-RL」を GitHub 上で無償公開した。

「SCOPE-RL」では、提案した性能検証指標を容易に用いることができるだけでなく、オフラインの方策学習から評価まで一貫して行える実装パイプラインを提供している。これにより、オフライン評価のより実務に即した運用や性能検証が可能になると期待される。

本研究成果は、2024 年 5 月 7 日から 5 月 11 日に行われる国際会議 International Conference on Learning Representations（ICLR）に 2024 年 1 月 16 日付で採択された。同国際会議 ICLR は ICML や NeurIPS と並び機械学習分野での世界最高峰の国際会議のひとつとして認知され、本年度は投稿論文 7262 本のうち約 31%の論文が採択されている。

●背景

広告の入札額調整や日次予算運用をはじめとしたマーケティング業務や、バイナリデータを元にした医療での治療方針決定など、連続的な意思決定が顧客の満足度や安全性に寄与する場面が実用上では数多く存在する。こうした意思決定システムにおいて、過去の運用データを用いて、新たに運用する候補方策を評価し選択することは、サービスの品質の維持・向上において必要不可欠である。実際に、図 1 に示すように、まずはデータを用いてオフライン評価により数多くの候補方策の中から少数の有望な方策を選定し、次にそれらの選択された方策の性能 **A/B テスト**（用語 2）においてオンライン検証し、最後に運用方策をサービスに展開する流れが多くのサービスで採られている。

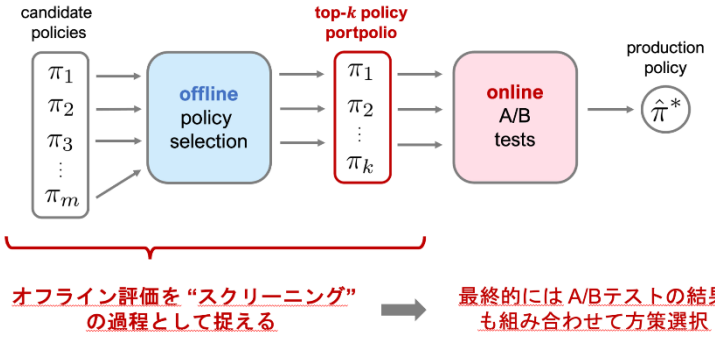

図 1 実運用における意思決定方策（policy）のオフライン評価から方策選択までの流れ

A/B テストでは実環境での方策の試運用やユーザーとの関わりが発生するため、事前のオフライン評価において信頼性の高いスクリーニングを行うことは非常に重要である。これまでに正確なオフライン評価を行うための研究が多く行われてきたが、従来の研究では正確性に注目するばかり、A/B テスト時に性能の悪い方策が採られるリスクが見逃されていた。そこで、本研究では方策選択のリスクに初めて注目し、A/B テスト時のリスクとリターンのトレードオフを評価するための性能検証指標を新たに開発した。

さらに、オフラインでの方策の学習から評価までを一貫して実装できる研究・実運用のための実装パイプラインが存在しなかった現状を踏まえ、オフライン方策学習・評価のためのオープンソースソフトウェア「SCOPE-RL」を無償公開した。

●研究成果

本研究では、図 1 の左側に示す、複数（m 個）の候補方策集合の中から上位 k（< m）個の方策をオフライン評価に基づいて選択するオフライン方策選択において、オフライン評価手法の性能検証を行う際の指標について再考した。まず、従来の「正確さ※1」に基づく性能検証指標では、（1）オフライン評価の際の過大評価と過小評価、（2）保守的な方策選択とハイリスク・ハイリターンである方策の選択、の区別が全くできず、性能の悪い方策を選択してしまうリスク※2 の比較ができていなかったことを初めて指摘した。そして、方策選択におけるリスクとリターンのトレードオフを定量的に評価するため、「SharpeRatio」という新たな性能検証指標を提案した。提案指標である「SharpeRatio」は経済学分野で株投資におけるポートフォリオのリスク評価に使われる「Sharpe ratio（シ

ャープレシオ)」から着想を得ており、発生する性能のばらつき（リスク）に対して得られる性能の上昇（リターン）がどの程度大きいかを比率として評価するものである。具体的には、すでに実運用されている方策に対して、選択された k 個の方策による A/B テストに基づき選ばれる次期の運用方策がどの程度性能を向上させるかをリターンとして見ており、これによりオフライン評価に基づき方策を刷新する際の利益を測る。また、リスクとしては、オフライン評価によって選ばれた k 個の方策の性能のばらつきをみることで、A/B テスト時に性能の悪い方策をユーザーに提示してしまう可能性の高さを検証している。最後に、既存のオフライン評価手法・指標とのベンチマーク実験を行い、提案指標は既存指標に比べて、(1) A/B テストで用いることのできる方策の数（k）に応じたリスクとリターンのトレードオフを評価することができること、(2) 既存手法が見逃していたリスクを適切に考慮してオフライン評価手法の性能検証が行うことができることを確認した。

また、今回独自に開発・公開したオープンソースソフトウェア「SCOPE-RL」により、オフラインでの方策学習・評価からオフライン評価手法の性能検証までが、一貫したプラットフォーム上で実装可能となる。これにより、より多くのエンジニアがオフライン強化学習の手順を容易かつ正確に実装できるようになり、オフライン強化学習の実応用の敷居が大幅に下がることが期待される。また本ソフトウェアには、オフライン強化学習やそのオフライン評価に関する最先端のアルゴリズムが多数実装されており、それらを一から実装することなく実務で利用することが可能になる。さらに、「SCOPE-RL」はオフライン強化学習に関する学習・評価のためのアルゴリズム自体の有効性を比較検証する機能を兼ね備えており、オフライン強化学習の新たな手法を開発する学術研究を行う際にも非常に有用である。

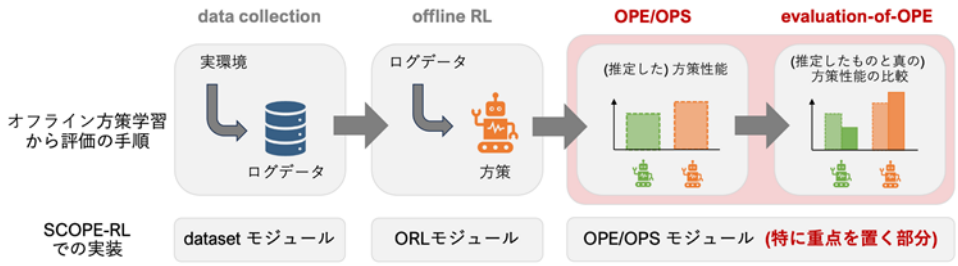

図 2「SCOPE-RL」を利用したオフライン方策学習・評価の実装手順。方策の学習から評価、さらにはオフライン評価手法の性能検証までを一貫したプラットフォーム上で実装できる。ORL（offline reinforcement learning）はオフラインでの方策学習、OPE/OPS はオフライン評価とそれに基づく方策選択（off-policy evaluation/selection）を指す。evaluation-of-OPE が今回主に議論しているオフライン評価手法の性能検証部分である。

※1　従来の「正確さ」を比較する性能検証手法には、(1) 方策の性能推定自体の正確さ、(2) 最上位方策の選択の正確さ、(3) 方策の並び替えの正確さ、の三つがあるが、いずれも「リスク」を考慮した性能評価は行えていなかった。

※2　k 個の方策の中に性能が低い方策を含んでしまうことを指す。特に、k 個の方策の中での性能のばらつきを見ることで、上記のリスクを評価している。

●社会的インパクト、今後の展開

今回の性能検証指標の開発およびオープンソースソフトウェアの公開により、新たな方策を A/B テストで評価および選択する際に、売上へ与える悪影響や消費者にとって好ましくない商品推薦や広告が届くリスクを軽減できるようになる。また、蓄積データをもとにして判断できるようになるため、専門家が長年の経験に基づいて判断していた意思決定の良し悪しを定量化し、再現性ある形で最適化できる可能性がある。さらに、オフライン評価の枠組みは、医療における治療選択やロボティクス、自動運転等に広く応用可能であり、これらの広範な技術分野に貢献する可能性も持つ。

【用語説明】

（1） **オフライン評価**：過去に集めたデータを用いて意思決定方策の性能評価（性能推定）を行うこと。

（2） **A/B テスト**：意思決定方策をオンライン環境に一定期間展開し、その方策の性能を評価すること。オフライン評価に比べ、実環境上で実験できる点でより正確な評価を行えるが、実環境上で実験を行うため顧客満足度を毀損してしまうリスクもある。

【関連リンク】

「SCOPE-RL」ドキュメント：https://scope-rl.readthedocs.io/en/latest/
「SCOPE-RL」公開リンク：https://github.com/hakuhodo-technologies/scope-rl

【論文情報】

会議名：International Conference on Learning Representations (ICLR), 2024
発表タイトル：Towards Assessing and Benchmarking Risk-Return Tradeoff of Off-Policy Evaluation
発表者・著者：Haruka Kiyohara, Ren Kishimoto, Kosuke Kawakami, Ken Kobayashi, Kazuhide Nakata, Yuta Saito
発表日時：2024 年 5 月 8 日(水) 16:30（現地時間、日本時間 23:30）開始

【問い合わせ先】

東京工業大学 工学院 経営工学系　教授
中田和秀
Email: nakata.k.ac@m.titech.ac.jp
TEL: 03-5734-3321　FAX: 03-5734-2947

【取材申し込み先】

東京工業大学 総務部 広報課
Email: media@jim.titech.ac.jp
TEL: 03-5734-2975　FAX: 03-5734-3661