第三回報告書：研究紹介と国際学会参加レポート

清原 明加（Haruka Kiyohara） / Cornell University, Computer Science

こんにちは。清原明加（きよはら・はるか）です。2023 年 8 月からコーネル大学コンピュータサイエンス（CS）学科の Ph.D. 課程にて、意思決定の最適化とその評価に関する研究をしています。今回はコーネルでの研究と、この夏にちょっとしたラッシュを迎えている (?) 学会参加の様子を書きたいと思います。

1. コーネルでの研究

Ph.D. 1 年目は授業を取ることが多いですが、今学期はあまり気になる授業がなかったので研究だけをすることにしました。[1] 今行っている研究は二つあり、それぞれ指導教官二人に一つづつ研究を見てもらっている形です。以下では、それぞれの研究について書ける範囲で紹介したいと思います。

一つ目の研究では、学部時代に取り組んだ「オフ方策評価（off-policy evaluation; OPE）」の応用として、ログデータを使って大規模言語モデル（large language model; LLM）による文章生成を最適化しようという内容を扱っています。少し背景を説明すると、昨今では ChatGPT などの LLM が多くの人にとって無料または低価格で使えるようになり、例えば映画のプロモーションに使う文章をユーザーの好みに合わせて生成したり、与えられた材料から好みのレシピを提案するようなアプリケーションでの利用が期待されています。こうしたアプリケーションにおいて自然と蓄積されるクリックなどのユーザーの報酬データを使って文章の生成を最適化するというもので、特に独自の LLM を持たない third-party の会社でも文章生成を最適化できるよう LLM の入力となる ”prompt”の最適化を行うのが目標です。[2] こうしたログデータを使った意思決定の最適化は「オフ方策評価」という研究のフレームワークを使えば議論できるのですが、従来のオフ方策手法だと意思決定の対象である ”prompt” のみに注目して、より情報量が多くユーザーに直接影響を及ぼす ”生成された文章” の特徴を全く利用することができませんでした。そこで、我々の手法では ”生成された文章” の情報を使ってより効率的に ”prompt” の最適化を行おう、というのを提案しています。[3] 納得のいく手法が作れるまで少し行ったり来たりしたプロジェクトではあったのですが、最終的には性能的にも良く、実用的にもかなり使いやすい手法を作ることができたと思うので、多くの人に使ってもらえるよう、公開に向けた大詰めをしっかりやっていきたいです。

二つ目の研究は、秋学期の「制御理論（control theory; CT）」の授業で取り組んだプロジェクトから派生した研究テーマに取り組んでいます。こちらはまだ詳細は書けませんが、ざっくりいうと、推薦システムで使われる意思決定アルゴリズムが推薦システムに与える長期的な影響を、”dynamics” と ”fairness” の観点から分析し、（近視眼的ではなく）長期的な効用を最大化する状態に落ち着ける方策を見つけ出す研究をしています。これまでの推薦システムでの意思決定最適化が基本的には今ある環境や集められた

[1] 論文を読んでディスカッションする「セミナー」には参加していたのですが、これは一般的に想起される「授業」とはちょっと違うので除外しました。

[2] 例えば、”WALL-E”という映画の説明文を生成する際に、ユーザーの好みに応じて ”sci-fi” や ”romance”, ”robots” といったキーワードとなる ”prompt” をデータを使って見つけようというタスクです。

[3] ワークショップ論文に書いている範囲で共有しています。技術的な詳細は次回の報告書に書きたいと思います。

データに対して受動的（reactive）に最適化しているのに対し、もっと積極的（proactive）に最適化を行おうという研究なのですが、この研究にはまだ seminal work と呼べる研究が存在しないくらいに新しいトピックです。そのため、随分と探索のしがいがある、未知の、手応えのあるトピックではあるのですが、その分良い分析をすれば推薦システムの見逃されていた問題に光を当て、新たな見方を提示できる論文になるはずなので、急がず堅実に研究を進めていきたいなと思っています。

さて、これら二つの研究は一見するとそれぞれ全く違うことに取り組んでいるように見えるかもしれませんが、私としては「意思決定システムを社会で実運用する際の、(1) 見逃されていた問題点を発見し、(2) それをデータを用いて解決する」という大きな研究目標の範疇で少し探索的に、新しいトピックに挑戦してみたという認識をしています。まだ 1 年目ということもあり Ph.D. thesis の内容を具体的にイメージしている訳ではありませんが、こうして長期的な研究目標を見失わない範囲でトピックを色々と探索して、自分でも想像できなかったような興味の広がりに出会い、良い意味で想像できない論文を書けたらいいなと思っています。そのために、まずは今の研究をしっかりと形にし、これまでの研究経験の活用と新たなテーマの探索のバランスをうまく取りながら研究に取り組んでいきたいです。

2. 論文採択と学会参加

上記の Ph.D. での研究の他に、春学期は嬉しいことに学部で取り組んだ論文が複数採択され、学期の終了とともに学会へ参加する運びとなりました。

まず最初に向かったのはオーストリアのウィーンで、機械学習分野で（ICML や NeurIPS と並び）最重要国際会議の一つである ICLR にて以下の論文を発表しました。

- **Haruka Kiyohara**, Ren Kishimoto, Kosuke Kawakami, Ken Kobayashi, Kazuhide Nakata, Yuta Saito. Towards Assessing and Benchmarking Risk-Return Tradeoff of Off-Policy Evaluation. *International Conference on Learning Representations (ICLR)*, 2024.

この研究は第一回報告書で紹介したライブラリ[4]に関わる姉妹研究で、「オフ方策評価」に使われる推定量の新たな評価指標を提案しています。もう少し具体的に書くと、オフ方策評価は与えられた方策の性能をログデータから推定しますが、しばしばオフ方策評価は多くの候補方策からオンライン A/B テストにデプロイする方策をオフラインでスクリーニングするために用いられます。この時に、「オフ方策推定量」の比較を、オンライン A/B テストにおける選択された上位 k 個の方策の性能をもとに行いたいというモチベーションが実用上発生しますが、従来の評価指標では「オンライン A/B テスト時に性能の悪い方策をデプロイしてしまうリスク」を全く考慮できていませんでした。そこで、リスクとリターンのトレードオフを基にオフ方策推定量を比較できる新たな評価指標を提案しました。提案指標は金融分野でよく使われるリスク・リターンの評価指標から着想を得ており、リスクを検証したいという実応用でのニーズに応えられるものになっています。また、この評価指標をオフ方策評価の実験に簡単に用いら

[4]SCOPE-RL: https://github.com/hakuhodo-technologies/scope-rl

れるよう前述のライブラリ上で公開しています。公開したライブラリについては以下の preprint に別途まとめているので、こちらもぜひ見てください（第一回報告書にも内容を詳しく記述しています）。

- **Haruka Kiyohara**, Ren Kishimoto, Kosuke Kawakami, Ken Kobayashi, Kazuhide Nakata, Yuta Saito. SCOPE-RL: A Python Library for Offline Reinforcement Learning and Off-Policy Evaluation. *arXiv preprint*, 2023.

次に、シンガポールに移動し、ウェブ・データマイニング分野で（KDD や WSDM と並び）最重要会議の一つである WWW で以下の論文を発表しました。

- **Haruka Kiyohara**, Masahiro Nomura, Yuta Saito. Off-Policy Evaluation of Slate Bandit Policies via Optimizing Abstraction. *The Web Conference (WWW)*, 2024.

この研究は「slate」と呼ばれる複数の意思決定を同時に最適化する際の「オフ方策評価」について研究しています。例えばオンライン広告やファッションアイテムの推薦においては、キャッチコピーと用いる画像、トップスとボトムスなど、複数の意思決定の組み合わせを最適化した結果、広告や全身コーディネート全体に対して単一の報酬（クリックなど）が得られるようになっています。複数のアイテムの組み合わせが発生する分、ナイーブなオフライン評価手法ではデータの少なさが問題になってしまうのですが、「slate」間の類似性をデータをもとに学習し、低次元潜在表現を使ってオフ方策評価することで、データ効率を改善しより正確なオフ方策評価を可能にしています。

ICLR と WWW の二つの研究を比べると、前者が少し遊び心のある探索的な研究になっていて、後者が理論に基づいたより堅実な研究という認識で、二つの異なる雰囲気の論文を同時期に発表できたのはとても面白かったなと感じています。学会もオフラインでポスター発表したのは初めてだったのですが、多くの人に興味を持ってもらえてとても楽しめました。近しい興味の方々との研究の雑談などモチベーションが上がる要素が多々あったので、これからもコンスタントに学会に参加できるように取り組んでいきたいなと思います。

2. 1 年目の総括と印象的だったこと

今学期は学部時代に仕込んでいた研究が順調に採択されたため、結果を見ればある程度実りがあったと思います。一方で、やはり Ph.D. での研究を早く出したいという気持ちもあったり、初めて学部生のメンターをしたもののなかなか研究のペースが掴めなかったりして、途中少し焦ってしまったこともありました。そんな中、過度なプレッシャーをかけず、ただ前向きに良い研究になるようサポートをしてくれる指導教官の偉大さを実感した貴重な機会でもありました。結果的に良い感じに開き直ることができ、焦らなくなってきたタイミングくらいから研究が徐々に良い方向に行き始めた実感があったので、自分でも心理的余裕を持つことを心がけていきたいです。末筆になりますが、いつも留学先での研究を温かく見守りご支援くださる船井情報科学振興財団の皆様に、深く御礼申し上げます。