

ARG 第21回WI2研究会 学生参加報告

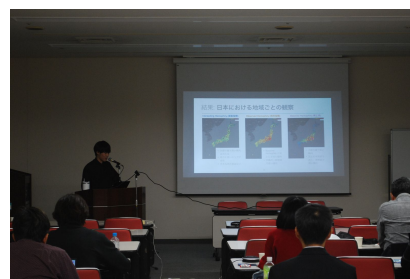
高柳 慶太郎
京都大学

概要

ARG 第21回Webインテリジェンスとインタラクション研究会では、情報推薦・支援、LLMの応用技術、ソーシャルネットワークやクラウドソーシング、情報抽出・レシピに関する25件の研究報告が行われた。2025年11月2日にオンライン、4日と5日に徳島市のアスティとくしまにて開催された。

著者の発表

著者は、4日にSNS・クラウドソーシングというセッションで「ソーシャルネットワークにおける人口の偏りを考慮した地理的選好の測定」というタイトルで発表した。この発表では、X/Twitterの相互フォローネットワークを対象に、人々が地理的にどの程度「近い相手」を選好してつながっているのかを測定する指標を、人口分布による偏りを補正した形で定義した。従来指標では都市部にユーザが多いほど偶然でも近距離のつながりが増えてしまうが、提案手法ではランダム接続の期待値を用いることで、この人口密度由来の影響を取り除き、純粋な地理的選好を評価できる点が特徴である。10カ国のデータを対象に測定したところ、すべての国で地理的に近い相手を選ぶ傾向は確認されたものの、人口規模との関係は国によって大きく異なり、都市部ほど選好が強くなるという従来の主張は必ずしも普遍的ではないことが示唆された。



著者の発表の様子

発表では、指標の仕組みや補正の必要性を丁寧に説明した分、結果の提示に割ける時間がやや短く感じられたものの、会場の理解度は高く、質疑でも前提を踏まえた議論をいただくことができた。

質疑ではまず、「Twitterのフォロー関係は必ずしも典型的なソーシャルネットワークではなく、情報ネットワークとしての側面もあるのでは」という指摘があった。引用やリプライなど別の関係を基準とした場合、距離依存性が変わる可能性についての問いであり、著者からは他種類のネットワークを用いた分析も今後の検討対象であると回答した。また、過去の研究でも定義した指標に基づいた測定結果に意味づけを与えることが難しいという経験が共有され、解釈の重要性が指摘された。

また、「人口密度による補正を行った指標にはどのような意味が残るのか」という質問もあった。著者からは、都市と地方では「近所」の距離スケール自体が異なるため、補正なし

では都市部の値が常に高く見える構造的問題が生じること、補正指標はその差異を排除したうえで地理的選好を公平に比較するためのものであると説明した。都市社会学には大都市ほど関係が地理的に分散するという仮説もあり、その検証に役立つことも付記したい。

なお、本発表に対して学生奨励賞をいただいた。研究内容への関心を多くの参加者から寄せていただけたことは大きな励みとなった。

聴講した発表

LLM の社会的バイアス低減に向けたモデルマージ手法の有効性検証

白藤大幹, 斉藤辰彦(三菱電機), 木村泰知(小樽商科大学)

この研究は、事前学習データが内包する偏見やステレオタイプにより、LLM の出力に社会的バイアスが現れる問題に対し、モデルマージ手法を用いてバイアスを抑制する取り組みである。再学習を伴う既存手法とは異なり、モデル同士のパラメータを加減算するだけでバイアスを低減できる点が特徴である。



白藤氏の発表の様子

研究では、GPT 系・LLaMA 系・Qwen 系を含む 13 種類の LLM と、7 種類の異なるマージ手法を組み合わせ、社会的バイアスの低減効果を体系的に比較している。評価には BBQ と HONEST の 2 種類のデータセットを用い、さらに SuperGLUE を通して下流タスクの性能への影響も分析した。

実験結果として、社会的バイアスを低減しようとする一部のタスクで性能が低下するというトレードオフが確認された。特に読解力や常識推論、因果推論を要するタスクでは性能劣化が顕著であった。一方で、Linear、SLERP、Nearswap といったマージ手法は比較的一貫して安定した性能を示すと報告された。

質疑では、「どの社会的バイアスを基準としているのか」「日本語モデルの場合はどう扱うべきか」など、バイアス定義そのものに関する問いが寄せられた。発表者は、本研究が用いたデータセットのバイアス定義に依拠している点、およびバイアス低減と性能維持の両立にはまだ課題が残る点を説明していた。また、包括的に比較したことで、これまで十分に把握されていなかった性能劣化の傾向を明確にできたと述べていた。

複数のモデルと手法を統一条件で評価した点に実験研究としての価値があり、単一モデルのみでは見えなかったバイアス低減の限界やトレードオフが丁寧に示されていた。特に興味深いのは、社会的バイアスの性質について深く考えさせられる点である。バイアスの定義は、コンセンサスに基づき決定されるため、社会集団が異なれば、異なるバイアスの定義が適用される可能性がある。この点を踏まえ、今後の研究として、バイアスの定義が変化した場合に、本研究で示された評価結果や限界、トレードオフの関係性が変化するのか検証するとさらに興味深いと感じた。

おわりに

今回の研究会では3日間にわたり推薦システム、Webからの情報抽出、LLMなど様々な分野の優れた研究発表を聞き、自分の知見を広げることができた。発表者同士の質問や交流も頻繁に交わされており、小規模な研究会ならではの密な関係という利点を実感した。外部に向けて自分の研究を発表することは、これまで気づかなかった観点や今後の方向性を考える契機となり、研究を進めるうえでの大きな励みとなった。



懇親会の様子