

第 21 回 ARG WI2 研究会 学生報告

2025 年 11 月 2 日（オンライン）、4 日、5 日（現地）

会場：アスティとくしま 1 階 第 2 会議室

同志社大学 文化情報学部 森岡葵子

1 はじめに

2025 年 11 月 2 日（日）、4 日（火）、5 日（水）の 3 日間、アスティとくしま 1 階第 2 会議室にて、第 21 回 ARG WI2（Web インテリジェンスとインタラクション）研究会が開催された。著者は、自身の研究に対する率直なフィードバックを得ること、他分野の知見を取り込み、研究に応用可能な示唆を得ることを目的に参加した。会場は比較的小柄なコンパクトで登壇者と聴講者の距離が近く、各発表後には活発な質疑が続いた。休憩時間やセッション間でも、教員と学生が立場を越えて議論を交わす温かな雰囲気があり、初参加者にも開かれた研究会であった。

プログラムは 3 日間にわたり、7 件のショート発表と 17 件のロング発表が、計 7 セッションに分けて行われた。各セッションでは、LLM や情報推薦をはじめ、多様なトピックが扱われ、分野横断的な視点から知見を得られる構成であった。本報告では、まず著者が行った発表、次いで著者が興味を持った発表をいくつか紹介する。

2 著者の発表

大規模言語モデル（LLM）は QA タスクや要約タスクにおいて入力されるコンテキスト内に含まれる外部知識に基づいた回答が求められる場面が多いが、モデルが事前学習によって獲得する内部知識と外部知識が矛盾する場合には外部知識に沿わない出力が生じやすいことが課題となっている。この問題の解決には文脈に応じて外部知識を優先できるようにトークンの生成確率を動的に調整する手法が有効であるとされている。

先行研究ではどの程度調整するかを外部知識を入力した際に変化した確率により決定するが、外部知識が事前学習で学習した知識と矛盾する場合に



図1：著者の発表の様子

は外部知識を入力しても内部知識を優先するため生成確率が十分に变化せず、調整が不十分となる。

そこで、著者は外部知識に近いトークンへの確率拡散を考慮するため、外部知識とトークン埋め込みとの類似度から補助分布を構成し、元の分布へ混合する枠組みを提案した。

本発表は著者にとって初めての対面での発表であり、緊張の中での登壇となったが、幸いにも多くの先生方から貴重なコメントや質問をいただくことができた（図 1）。

質疑においては、まず提案手法の肝となるパラメタ λ の決定方法や、実際の生成例について質問があった。これに対し、著者は一部の生成例を示しつつ、現状はグリッドサーチでの探索に留まっており、どのように最適値を決定すべきかは今後の課題であるという認識を共有した。

また、外部知識の信頼性をどのように扱っているかという質問もあった。これについては、本手法が、自治体の Web サイト内チャットボットのように、システム側が提供するコンテキストが正であると仮定し、それに基づく回答が必要な場面を想定しているため、外部知識自体の真偽判定は本研究のス

コープ外である旨を回答した。

さらに、手法の適用範囲や評価指標に関しても多角的な議論が展開された。具体的には、要約タスクや意見の対立を含む対話文において知識の衝突がどのように生じるか、JSD (Jensen-Shannon Divergence) などの既存指標と提案手法による検知に相関があるかといった点である。加えて、因果関係を含む知識において、生成文から LLM がどの知識を用いているかまで判定できるかという発展的な問いも投げかけられた。

これらの質疑を通じ、多くの新たな知見を得ることができ、今後の課題を改めて認識できた。

3 興味を持った発表

本研究会では、LLM や情報推薦をはじめとして、多様な分野にわたる発表が行われた。本節では、その中でも著者が興味を持った発表について述べる。

白藤氏 (三菱電機株式会社) らによる「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」は、モデルマージによる LLM の社会的バイアス低減手法を検証した。著者は、バイアスを含むデータで学習したモデルと元のモデルとの差分からバイアス成分をベクトルとして抽出するという点に非常に興味を持った。これは、特定の特徴を分離して挙動を制御できるという点で著者の研究と共通しており、バイアス除去に限らず応用範囲が広いと感じた。自身の研究においてもモデルの挙動制御は重要な課題であり、パラメタの差分を手掛かりに特性を誘導するという本発表の着想は、著者の研究における手法に新たな視点を与えるものであった。

また、大西氏 (北海道大学) らによる「正確性と多様性を両立する傾向スコア導入線形オートエンコーダ」では、推薦における正確性と多様性のトレードオフに対し、傾向スコアを導入した線形オートエンコーダにより多様性を保ちつつ推薦の正確性を保持する手法が提案された。本発表では特に、ハイパーパラメタ β の設定に関する議論が印象深かった。これは、自身の研究においてもパラメタ調整が手法の肝となっており、その決定プロセスに課題を感じていたためである。具体的には、傾向スコアにおけるシグモイド関数の傾き β の決定基準について疑問を持っていたが、太田先生 (岡山大学) の



図2：懇親会の様子

質疑に対する回答により、グリッドサーチによって精度と多様性のバランスを考慮して決定されたものであることが確認できた。さらに、吉田先生 (筑波大学) からは、時間の経過に伴う分布変化 (アイテムの人気推移など) に応じた、 β の動的な設計の必要性が提起され、分布動態を把握しパラメタ選択を妥当化することの重要性が示された。この議論は、入力コンテキストに応じた動的なパラメタ調整を課題とする自身の研究における課題の重要性を再認識する上で、極めて有意義な知見となった。

4 おわりに

今回の研究会では3日間にわたり、多くの発表を聴講した。各発表では、著者の研究においても考慮すべき点が多く議論され、多くの新たな知見を得ることができた。今後は、質疑で浮き彫りとなったパラメタ決定や JSD の値の分析に関する課題を踏まえ、動的なパラメタ設計と多角的な評価実験の設定により提案手法の実用性を高めていきたい。

また、研究会以外にも参加者同士の交流の機会が設けられており、図2は4日の夜に行われた懇親会の様子である。今回の懇親会は徳島駅付近の居酒屋で行われ、新鮮な天然地魚の刺身や阿波尾鶏などの料理を肴に、学生や先生方間で活発な意見交換が行われ、大いに盛り上がった。

末筆ながら、著者が本研究会において行った発表は優秀研究賞の栄誉を賜った。ご審査ならびにご助言を賜った皆様に深く感謝申し上げますとともに、今回の受賞を励みに、今後の研究の発展と成果の発信に一層努める所存である。