

## 大規模言語モデルを用いた検索結果の クリックシミュレーション

本藤 祐大

現代の検索エンジンは、商品購入における価格やレビュー比較、旅行計画における宿泊・交通手段選定、医療分野における症状・治療法の検索など、多様なニーズに対応する包括的プラットフォームへと進化している。一方、ユーザの増加と多様化に伴い、求められる情報は膨大かつ多岐にわたり、ユーザによって異なる検索行動が報告されている。こうした幅広い背景や個別のニーズに即した結果を迅速かつ的確に提示するには、ユーザごとに最適化されたアプローチが必要である。既存の検索システムは多くの場合、キーワード形式のクエリを頼りに結果を予測しなければならないため、異なるコンテキストを十分に考慮しきれない。たとえば「航空旅行情報」で関東のユーザと関西のユーザが同じ検索を行っても、それぞれに合った空港や交通手段情報を優先的に表示できず、検索体験を損ねる可能性がある。これを解決する手段として、検索クエリや行動履歴、ユーザ特性を組み合わせたパーソナライゼーションが重要視されており、近年は LLM を用いたユーザ行動シミュレーションの研究も進んでいる。しかし、実際のユーザ特性に基づく十分な検証はまだ不足しているのが現状である。そこで本研究では、LLM を用いた検索結果のクリックシミュレーションを、実際のユーザ特性を用いて行い、その精度を実際の検索ログに基づいて検証する方法を提案する。具体的には、実在する検索ユーザの特性データをプロンプトとして入力し、検索ログをシミュレーションの正解データとして活用することで、モデルのシミュレーション性能を評価することを目的とした。実在するユーザの特性と検索ログのペアを含む公開データセットが確認できなかったため、我々はクラウドソーシングを活用して新たにデータを収集した。TREC データセットをもとに構築した検索タスクに参加してもらい、我々が開発した検索システムで実際の検索行動ログを取得し、同じ参加者にアンケートを実施してユーザ特性を収集し、検索ログとのペアを作成した。ユーザ実験で収集したデータを用いて、LLM に「実験協力者の特性情報」「検索クエリ」「検索意図を示すディスクリプション」「検索結果一覧」を入力し、どの検索結果をクリックするかをシミュレーションさせた。シミュレートされたクリックログと実際のクリックログについて、Precision や Recall などを用いて複数の LLM モデル間で性能を比較した。さらに、検索専門性や心理学的属性など、どのユーザ特性がシミュレーション精度に影響するかについても分析した。LLM によるクリックシミュレーションにおいて、ユーザ特性を入力した場合に、ユーザ特性を入力しない場合と比較して、Precision が高く、また、どの LLM においても、ユーザ特性を考慮した場合、ランダムクリックモデルと比較して Precision が高いことがわかった。このことから、LLM によるクリックシミュレーションにおいてユーザ特性の重要性が示唆された。

(指導教員 加藤 誠)