KERAG_R: Knowledge-Enhanced Retrieval-Augmented Generation for Recommendation

---

Author: Zeyuan Meng

DOI: 10.48550/arXiv.2507.05863

Link: arXiv

Main Idea/Research Question: This paper addresses the limitations of Large Language Models (LLMs) in recommender systems, specifically their lack of domain-specific knowledge, which can lead to inaccuracies and "hallucinations". The authors propose a novel model, KERAG_R, to address this by leveraging a Graph Retrieval-Augmented Generation (GraphRAG) component to integrate external knowledge from a knowledge graph (KG) into the LLM's instructions. This allows the LLM to better utilize recommendation signals from both text-based user interactions and the knowledge graph.

Dataset(s) Used: The paper conducted experiments on three public datasets. (MovieLens-1M, MovieLens-10m, Amazon-book).

Model(s) Used: The proposed KERAG_R model leverages an LLM (e.g., Llama 3.1) and a GraphRAG component. The GraphRAG component uses a pre-trained Graph Attention Network (GAT) for triple retrieval.

Procedure/Configuration or Test: The model architecture has three parts: (a) Graph Retrieval-Augmented Generation (GraphRAG), (b) Knowledge-enhanced prompt construction, and (c) Knowledge-enhanced instruction tuning. The GAT is pre-trained to select the most relevant triples for target users.

Particular Techniques/Tricks Applied:

- Using a GraphRAG component to retrieve external knowledge from a KG to enhance the LLM's reasoning.
- Pre-training a Graph Attention Network (GAT) to select the most relevant knowledge graph triples.
- A knowledge-enhanced instruction tuning approach to incorporate relational knowledge during the LLM's tuning stage.

Performance Metrics and Results: The KERAG_R model significantly outperforms ten existing state-of-the-art recommendation methods. For example, it outperformed the strongest baseline, RecRanker, by up to 14.89% on the Amazon-Book dataset. The study also found that using GraphRAG is more effective than just using additional KG triples and that relational KG triple representations outperform natural KG sentence representations in prompts.

Key Learning/Finding: Integrating structured knowledge from a knowledge graph via a GraphRAG component is an effective way to address the lack of domain-specific knowledge in LLMs, leading to significantly improved performance in recommendation tasks.

Large Language Models as Conversational Movie Recommenders: A User Study

---

Author: Ruixuan Sun

DOI: 10.48550/arXiv.2404.19093

Link: arXiv

Main Idea/Research Question: The paper explores how users experience and evaluate LLM-based recommenders compared to traditional systems. It focuses on three main questions: how users perceive LLM recommenders versus classic ones, how prompts, scenarios, and users' prior movie-watching behavior affect recommendation quality, and what interaction strategies help users get better results. Findings show that while LLMs provide strong explainability and engaging conversations, they struggle with personalization, novelty, diversity, and trust. User history had a greater impact on perceptions than prompting style, and providing context, preferences, or examples proved crucial for obtaining more satisfying recommendations.

Dataset(s) Used: The study used data from a partnered movie recommendation platform (referred to as MovieRec). From its database, the researchers identified 3,031 qualified users who had logged in at least 12 times and rated more than 20 movies in 2023. Out of these, 449 users enrolled, 178 completed the questionnaire, and the final analysis was based on 160 unique users who completed all evaluation questions. User data included historical movie ratings and genres, which were used to generate personalized prompts (zero-shot, one-shot, few-shot) and to categorize movies by popularity tiers (high, medium, low).

Model(s) Used: The study used the pre-trained Llama-2-7b-Chat model, released by Meta in 2023, as the conversational engine for generating movie recommendations.

Particular Techniques/Tricks Applied:

- Prompting strategies: Three levels of personalization were tested — zero-shot (only top genres), one-shot (one liked, disliked, and recommended movie), and few-shot (four liked, disliked, and recommended movies) — to see if more context improved results.
- Scenario-based evaluation: Users interacted with the chatbot in three scenarios — Birthday (ask-for-others), Long Trip (personalized for self), and Niche (unpopular movies) — to capture different recommendation contexts.
- Popularity tiers: Movies were categorized as high, medium, or low popularity based on their rating frequency, allowing analysis of whether LLMs were better at recommending niche content.

Performance Metrics and Results: The study compared a Large Language Model-based recommender (LLMRec) to a classic recommender (MovieRec) and found a nuanced performance difference, where each model excelled in different areas.

- LLMRec was significantly preferred for its ability to provide explainable recommendations across all scenarios, with preference rates ranging from 45.4% to 51.0%.
- MovieRec was generally preferred for these metrics, with preference rates as high as 71.9% (for diversity in the Birthday scenario) and 66.9% (for novelty in the Niche scenario).
- MovieRec was preferred for personalization in all scenarios, with preference rates between 56.9% and 66.2%.
- The study also found that the number of movies a user has watched played a more significant role in perceived recommendation quality than different personalized prompting techniques.

Additionally, LLMs showed a greater ability to recommend lesser-known or niche movies.

Key Learning/Finding: The study found that different personalized prompting techniques did not significantly affect user-perceived recommendation quality. Instead, the number of movies a user has watched and the context provided by the user in the conversation were more crucial for getting high-quality recommendations. Users who gave minimal context or repeated queries had lower satisfaction. LLMs demonstrated a greater ability to recommend niche or lesser-known movies, which is a unique strength compared to traditional recommenders.