Nan Jiang - Self Statement

I completed my undergraduate studies at Beijing University of Posts and Telecommunications, achieving excellent results in the graduate entrance examination, which allowed me to continue my studies there as a graduate student. During my master's program, I diligently completed the project tasks assigned by my research group while actively participating in competitions. I proactively reach out to Dr. Haitao Yuan, a postdoctoral researcher in Professor Cong Gao's group at Nanyang Technological University, and collaborated with him on spatial-temporal big data research. Over the past two years of research, my first-author paper was accepted by the CCF-A conference ICDE 2024, and I co-authored a paper accepted by the DASFAA conference. Additionally, several other papers are currently under review.

Research Interests

My primary research during my master's study focused on the field of spatial-temporal big data, specifically on recommendation, prediction, and matching algorithms for trajectory data. Some detailed research topics include: (1) Developing an enhanced next POI (Point of Interest) prediction algorithm using remote sensing data, by introducing regional remote sensing embeddings to transform the classical POI prediction task into a two-step region and semantic prediction algorithm; (2) Implementing an online vehicle-road matching algorithm using reinforcement learning (Q-learning); (3) Designing a spatial mapping method for trajectory matching tasks that can bypass the limitations of the triangle inequality...(Please refer to my CV for more details)

My previous work has mainly focused on deep learning for spatial-temporal data mining in intelligent city scenarios, but **I am also eager to explore other areas** of Deep Learning or Data Mining during my Ph.D. studies. I am actively exploring new research directions, currently working on a research that combines reinforcement learning with LLMs (Large Language Models) to construct recommendation prompt agents. I am still drafting my research proposal for Ph.D application, but I can provide a brief version. (Detailed on the second page)

Through these years of research experience, I have gained well understanding of Data Mining and Deep Learning. I am particularly interested in innovatively applying deep learning models to real-world scenarios, which I believe is what makes "Data Mining with Deep Learning" so appealing for me. Nowadays, with the development of AI large models and the massive increase in user data, Data Science with AI will certainly be a major direction for the future development of computer science. Therefore, I am very eager to explore this field in depth, hoping to have the opportunity to engage further discussions with the more talented faculty and students.

Research Proposal (Brief Version)

Background

Nowadays, LLMs have demonstrated immense potential across numerous fields. Many researches are now exploring the application of LLMs in recommendation system. However, the nature of LLMs, being primarily language-based, means that their outputs are significantly influenced by the input prompt text. Existing methods attempt to mitigate this issue using techniques like Retrieval-Augmented Generation (RAG) or callable Agents, which provide additional information resources through rule-based methods. However, these approaches cannot supervise the effectiveness of the prompt resources, leading to redundant and inefficient inputs and ineffective model judgments.

Introduction

Many studies have demonstrated that LLMs are capable for recommendation tasks. However, due to the limited domain-specific knowledge stored in general LLMs, models that are not fine-tuned on specific datasets cannot produce predictions that align with specialized semantics, unless using RAG structure for information augmentation. Thereby, we summarize the **CHALLENGES** faced by large models in recommendation tasks into the following three points: 1. The RAG architecture used by LLMs only provides rule-based methods for filtering prompt information and cannot filter out potential redundant information, leading to less optimal output results. 2. The use of RAG is limited to structured databases and cannot utilize unstructured base, such as knowledge graphs. 3. It is unable to provide timely personalized recommendation responses based on online user feedback. (Since this functionality requires collecting data and fine-tuning the large model itself.)

To address these challenges, we propose a Prompt information filtering RAG model based on Knowledge Graphs and reinforcement learning, namely RGPA (Reinforcement Graph Prompt Augmentation). This model can more accurately extract and provide prompt information suitable for users from the knowledge graph through online human feedback, thereby enhancing the personalized recommendation capability of large models.

Methodology

1. The Retrieval module first extracts information nodes covered by the user's visiting history from the global knowledge base to construct a knowledge sub-graph. This sub-graph contains all relationships related to the user's visiting history.

2. The Reinforcement Learning module uses the Q-learning method to build a graph node filtering network based on the predicted results and changes in the user's visiting states. This model can filter out the most beneficial node content for the user's recommendations, thereby assisting the LLMs in producing better recommendation results.

3. The Semantic Augmentation module constructs a well designed graph mapping method to

transform the filtered node knowledge into prompt content that assists the LLM in making recommendations.