SSE 2024 Data Analysis and Mining

Workshop Details

Workshop Topics

1) **Large Language Model.** This topic focuses on the latest developments and practical applications of large language models (LLMs). Participants will delve into core technological advancements, innovative use cases, and future trends in LLMs. Topics encompass but are not limited to model architecture improvements, self-supervised learning and fine-tuning strategies, multilingual and cross-cultural understanding, ethical and privacy concerns, and practical applications of LLMs in text generation, question answering systems, document summarization, conversational interfaces, among others. Attendees are encouraged to share their research findings, experimental experiences, and profound reflections on the technological boundaries and societal impacts of LLMs.
2) **Human Energy Expenditure (Calories) Estimation.** Human energy expenditure (EE) refers to the amount of energy an individual uses to maintain essential body functions (respiration, circulation, digestion) and as a result of physical activity. Knowledge regarding the expended energy or calories could help people (e.g., athletes, obese, diabetic) to plan their physical activity for leading a healthier lifestyle. Additionally, it could be used to enable nutrition coaching for weight management purposes. Find such datasets and related methods/ algorithms or applications.
3) **Chinese Food Image Processing (Data and Algorithms).** Image processing techniques play an increasingly important role in food recognition and analysis, especially in the recognition of Chinese food. Please investigate the current datasets, commonly used algorithms, and their application scenarios in the field of Chinese food image processing, analyze the strengths and weaknesses of the current techniques, and propose possible directions for improvement.
4) **Knowledge Graph Based on News Data.** Knowledge graph can effectively organize and manage a large amount of information and provide users with rich knowledge services. In this assignment, students need to use news data as an information source to explore the method of constructing a knowledge graph, including the steps of data collection, entity recognition, relationship extraction, and graph construction. Through case studies, the potential of knowledge graph application in news data management and knowledge discovery is demonstrated.

In short, it is a brainstorming session, so feel free to share your ideas. You are encouraged to experiment with the data and present the results as a demo, which will lead to higher grades.

Note that the number of students in each group is strictly limited to four.

There are two ways to finish your assignment:

1. submit your project document without presentation. You are request to submit all documents about your project, such as source code, README file, report file with results and analysis and relevant documents. If you submit documentation, no in-class presentation is required. Note that you need to list the *student IDs* and *Chinese names* of all group members on "members.txt" in main directory, and IDs should be in ascending order.

2. submit your slides and present them in class. You are request to submit your slides for presentation on time. In addition to submitting your slides, you will need to present them in class. Please list the *student IDs* and *Chinese names* of all group members on the first page of your slides, which should be in ascending order.

Please choose one of the above two ways to complete this workshop.

Submission (both for 1 and 2)

Who is responsible for submitting: the member with the smallest student ID.

Submission date: 2024-04-21.

Please submit your files to Canvas.

Presentation (only for 2)

Presentation date: 2024-04-24

Presentation Time per Group: 5 minutes for presentation and 2 minutes for raising questions.