KOREA ADVANCED INSTITUTE OF SCIENCE AND TECHNOLOGY (KAIST)

Social network-aware ubiquitous computing - CS612/DS662

Team Project Announcement

September 10, 2024

1 Topic

"Spatiotemporal learning based on a long-term AirBnb dataset."

2 Tasks

1. Team-Up
    - Team 1: Hongju Lee, Nhat Quang Tau, and Pablo Espinosa Compos
    - Team 2: Junho Yoon, Jinghan Cheng, Victorin Turnel, and Minwoo Song
    - Team 3: Junho Lee, Yuheng Wu, and Jonghyuk Yun
2. Idea development/Proposal
    - Performing data analysis and visualisation on our AirBnB dataset.
        - Perform exploration techniques to find correlations, trends, etc within the dataset.
        - Visualise the exploration results.
    - Propose a research topic based on your exploratory results.
        - Write a 2-page proposal using the feedback and submit at least 2 days before the proposal presentation.
        - Present your findings and proposal in class to receive feedback.
        - Sample topics:
            * AirBnB data collected from different regions/areas show different trends. Look for the main (common) driving factors and use them to for future predictions.
            * Use *Butler's Tourism Area Life Cycle Model* to define the development stages and make future predictions.
            * Using spatiotemporal analysis, explore and make predictions about changes in different areas (e.g., economy, safety, tourism, population).
            * Based on AirBnB reviews, propose different changes to business time, location, price, housing scheme to optimize the business model.
            * Based on spatiotemoral correlation between Airbnb and Tourism, leverage Deep Learning models (e.g., large language model, graph model) to write introduction, advertisement for tourism.
3. Methodology Development and Term-paper writing
    - After finalising your research topics, each team proceeds to develop a methodology, which includes, but is not limited to, data preprocessing (e.g., clean-up and normalisation) methods, DL model(s), training process, experiments, etc.
    - Present your methodology and eventually, results.
    - Write a short 6-page term paper.

3 Important Dates (tentative)

- Idea proposal: 09/30
    - Notice: A two-page proposal must be submitted at least 2 days (by Sat 09/28) before the proposal presentation.
- 1st Design presentation (Algorithm): 10/28
- 2nd Design presentation (Algorithm details and experiment design): 11/25
- Final presentation (Experimental results and analysis): 12/16
    - Each team has until 12/20 to finalise and submit their term paper.

4 Dataset, Models, and Testbed

4.1 Dataset

Airbnb listing records

- Crawled dataset by AirDNA
  - Site: the City of Seoul, Korea
  - Period: 2014.11. 2022.07.
  - Files: Monthly Reported Listings and Listings' Property Information
    - Uploaded to KLMS.
  - Features:
    - Property type, Listing type, Location.
    - Reporting month, Revenue, Occupancy rate, Number of reservations, Reservation days, Available days, Blocked days.
    - Number of photos, Max guests, Airbnb rating, and other information.
- *Notice:*
  - 이 데이터셋은 이동만 교수님 연구실의 연구자산이기 때문에, 본 과제 이외의 데이터 활용 및 공개는 허가를 요청하여야 합니다.
  - This dataset is a research asset of Professor Dongman Lee's lab. Permission must be requested for the use and disclosure of data beyond the scope of this class.

4.2 Models

- The following model architectures are preferred:
  - Spatial Graph Convolutional Networks
  - Timeseries Large-language Models/Small Language Models
- Other model architectures are welcome as long as their novelty and efficacy are justified.

4.3 Testbed

- Each team is given a VM with 4 GPUs (80-96 GB vRAM) to perform data exploration, model training, and result analysis
- The number and type of GPUs are subject to change depending on their availability at the School of Computing.
- VM access details will be uploaded once the VMs are allocated by the SoC.