CS247 Project Overview

Spring 2021

Important Dates

April 10 (Saturday), 11:59 pm	Group Formation Deadline
April 29 (Thursday), 11:59 pm	Project Proposal Deadline
June 1 (Tuesday), 10 am - 11:50 am	Project Presentation Session 1
June 3 (Thursday), 10 am - 11:50 am	Project Presentation Session 2
June 3 (Thursday), 11:59 pm	Project Report Deadline

Grading Scheme

- Project accounts for 35% of the total score
- It consists of proposal (5%), presentation (10%), and final report (20%). Detailed grading criteria can be found in project guideline.
- Students in the same team MAY NOT get the same score

Submission

- All the submissions should be made on CCLE
- One group, one submission
- Late submission is NOT allowed
- What to submit:
  - <u>Proposal</u>:
    - A PDF that follow our formatting instruction
  - <u>Presentation</u>:
    - Your slides for the presentation
  - <u>Final report</u>:
    - A PDF that follow our formatting instruction
    - A link to access your code (please append it at the end of your abstract)
      - Use programming language/package that you are comfortable with
      - Code should be runnable, with clear readme file
  - Detailed requirements can be found in the project guideline

Group Formation

- 3~4 students per group

- Enough workload for each member, NO FREE RIDERS!

- Fill in this sign-up form before April 10 (Saturday), 11:59 pm:
    - https://docs.google.com/spreadsheets/d/1r5W4VMIzQDM4Hjwii0Q1CtkCpeXb7djZ3J-oF9natKA/edit?usp=sharing

How to Find Teammates?

- Make a Piazza post:
  - briefly introduce yourself (name, email, skills, potential project topic, etc.), then wait for others to contact you.
- Check the sign-up form:
  - email the group leader whose team has available position to see if you are a good fit.
- Join our KDD Cup Team:
  - will be introduced later
- Email TA and wait for random assignment:
  - please email TA (wyw10804@cs.ucla.edu) before April 11 (Sunday) 11:59 pm if you still can not find a group. TA would randomly team-up students who do not have a group by April 12 (Monday).

Choose A Topic

- Open ended, but should be related to CS247 contents and should be related to text, graph, or recommender system.

- Can be either a research work or an application work.

- Source of Ideas:
  - Explore the topics of data science competitions
  - Extend existing paper
  - Self-defined problems:
    - Feel free to talk with TA about your ideas

Source of Ideas 1: Data Science Competition

- Kaggle Competitions:
  - You can find various data science competitions in Kaggle: https://www.kaggle.com/competitions.

- KDD Cup: *OGB Large-Scale Challenge (OGB-LSC)*
  - Goal: Develop state-of-the-art graph ML models for large-scale datasets
  - Please refer to: https://ogb.stanford.edu/kddcup2021/ to get details.
  - For interested students: email Ziniu Hu (bull@cs.ucla.edu) before April 7 (Wednesday), 11:59 pm. In the email, please:
    - use this title: CS247 Student Interested in KDDCUP - Your Name
    - briefly introduce yourself (1-2 sentences)
    - briefly introduce your previous experience that might be helpful for this project (<=5 sentences)
    - attach your Curriculum Vitae to the email

  Ziniu will get back to you by April 8 (Thursday). We only have limited headcounts, please apply soon if your are interested.

Source of Ideas 2: Extend Existing Paper

- You can explore the latest methods for various classic graph-related tasks. You can follow the following procedures:
  - <u>Step 1</u>: find a classic task you are interested in.
    - E.g.: node classification, graph classification, link prediction, recommender systems, etc.
  - <u>Step 2</u>: find some benchmark datasets.
    - We recommend small-scale datasets, which are relatively easier to handle. We would not provide computation resources (such as GPU/CPU).
  - <u>Step 3</u>: implement one recent model published in top-notch conferences
    - Machine Learning Conferences: *NeurIPS, ICML, IJCAI, ICLR, AAAI···*
    - Data Mining Conferences: *KDD, WSDM, WWW, CIKM···*
    - Natural Language Processing Conferences: *ACL, EMNLP, NAACL···*
  - <u>Step 4</u>: tune the model, compare their performance to the other baseline models, discuss its pros and cons, and try to improve the model.
- You may find this site helpful: https://paperswithcode.com/.

Last Year's Sample Projects

- <u>Sample 1</u>: Graph Adversarial Attack and Defense
    - Inspired by last year's KDD competition, focus on attack side
    - Use Graph Convolutional Network (GCN) as target model, propose and apply some attack strategies and discuss possible defense strategies
- <u>Sample 2</u>: Long-Tail Recommendation System
    - Self-defined problem
    - Propose solutions to long-tail recommendation system problem
        - Long-tail problem: for new items in the system, they have very few interactions so it would be hard to be discovered and be recommended to users.

Last Year's Sample Projects

- <u>Sample 3</u>: Novel COVID-19 Drug Discovery Using Deep Learning And Knowledge Graph Representations for Drug-Target Interaction
  - Self-defined problem
  - Propose framework for drug discovery, the framework integrates deep learning and knowledge graph representation techniques
- <u>Sample 4</u>: Differentiable Architecture Search for Graph Neural Network
  - Self-defined problem
  - Propose an approach to automatically design the graph neural networks' architecture

What Else…

- Please carefully read the project guideline before you start.
- Feel free to come to the office hours if you want to discuss your project.
- Good luck with your project ☺