Algorithms and Applications of Data Mining

Yijun Lin

About This Course

- Lecture: Friday 7-9 PM CDT
- Mentor Session: Saturday 8-9:30 PM CDT
- Instructor: Yao-Yi Chiang (yaoyic@usc.edu)
- Mentor: Yijun Lin (yijunlin@usc.edu)
- TA: Xinyi Hu (samaritanhu@gmail.com)

Week 1-4

- Week 1:
  - Review the lecture content (MapReduce)
  - Introduction to Spark (Part I)
  - Set up the assignment environment
  - Introduce Assignment 1 and Yelp datasets

- Week 2:
  - Review the lecture content
  - Introduce Spark (Part II)
  - Continue Assignment 1 (submission)

Week 1-4 (Cont.)

- Week 3:
  - Review the lecture content (Recommendation System)
  - Building a Recommendation System (Part I)
  - Practice Collaborative Filtering algorithm with pseudocode and examples;
  - Introduce Assignment 2 and Yelp datasets

- Week 4:
  - Review the lecture content (Recommendation System)
  - Building a Recommendation System (Part II)
  - Continue Assignment 2 (submission)

Week 5-6

- Week 5:
  - Review the lecture content (Clustering)
  - (Optional) Discuss Assignment 3
  - Introduce how to do a research presentation with a demo project

- Week 6:
  - (Optional) Discuss Assignment 3
  - Introduce how to write a research paper with Overleaf

Assignments

- Assignment 1 - Basic Spark Operations
- Assignment 2 - Recommender System
- Assignment 3 (Optional) - Clustering

Config Environment

- **Python** is required for all the assignments

- Implementing with Apache Spark Framework
  - python=3.7
  - pyspark=3.0.1
  - git clone https://github.com/linyijun/data-mining-teaching-materials

- Install miniconda/anaconda
  - conda env create -f spark-env.yml python=3.7

- Install PyCharm