DATA ANALYTICS & MINING

Project 1

Lin Guosheng

Email: gslin@ntu.edu.sg

School of Computer Science and Engineering,
Nanyang Technological University

Project 1: Technical Review Report

- Group based project
  - Each group has up to 4 members
  - The assessment criteria will be the same for any group size.

- What you need to submit:
  A zip file containing the following files:
  - 1. Review report PDF
  - 2. Source code
  - 3. Contribution summary PDF
    - Describe the contribution for each member

- Where to submit
  - Please submit your zip file in NTULearn:
    the course site -> Assignments -> submission link for project 1

- Submission
  - Please name your zip package by "GroupidXX.zip", where XX is replaced with your group id.
  - One group only makes one submission.
  - If your file is too large, you can provide external links to download your file.
  - You can submit multiple times (no limit).
  - You don't need to submit your datasets.
  - You don't need to explain your code or provide inline comments.
  - There is no video component.

- Submission due date: 13 Oct 2024
  (end of Week 8)

- Grouping:
    - please edit the online form below to create your group.
    - https://docs.google.com/spreadsheets/d/1YMiw326R1CJ_k_C9Dklqqmh tLcQMhNEQ-0nQeKROs20/edit?usp=sharing

    - Each group has up to 4 members
    - You can form a group less than 4 members.
    - The first person for each group is the coordinator.

- Please complete the grouping by 26 Aug 2024 (Mon, Week 3)
    - I will release the grouping in Week 3.
      You need to verify your grouping.

    - I will create random groups for the students who are not in any groups.

Requirements for the technical review report

1. Select a topic/task from the list below:

  - Clustering
  - Link Analysis
  - Similarity Search
  - Graph neural networks or deep networks on graphs
  - Graph clustering or graph community detection
  - Recommender System (covered in the 2nd half)

  - Any sub-topics that belong to the above topics
  - Any topics/applications that use the above techniques

- 2. Select methods for review and comparison

  - Select at least 2 methods (algorithms)
    - You can choose any methods
      - They can be any methods published online, including but not limited to what you learn in the course

      E.g., K-means and K-means++, or K-mean and DBSCAN
    - You can use any public code
    - You can propose your own methods
      - e.g., extensions of existing methods
    - reviewing more than 2 methods will be a plus

Notes:

1. If you would like to work on the topics that are not covered in the first half, you need to learn the topics by yourself

- E.g., recommender system

2. If you want to explore more advanced methods, you need to search online and learn them by yourself.

1. e.g., Stanford course CS224w :
http://snap.stanford.edu/class/cs224w-2020/

3. Use at least 2 datasets

- Any datasets, including public datasets or the dataset you create.
- No requirement on the size/scale of the dataset
- Using more than 2 datasets will be a plus

4. Experimental analysis and comparison

- Should not copy results from others (including published papers)
  - You need to produce the results by yourself.
- Analyze and discuss the strengths and weaknesses of the methods
- Analyze and discuss the parameter settings
  - E.g., the number of K in K-means
- Ablation study of the key components in a method (if applicable)
  - Only applicable to the methods that have multiple components
- Discuss the key factors that affect the performance
  - E.g., K-means are sensitive to centroid initialization
- Compare the performance of the methods and provide discussions
- Illustrate and discuss the successful cases and failed cases (if applicable)
- Other possible analysis ….

5. Format
- At least 4 pages, no upper limit. (font-size: 11pt or 12pt)
- Include the following sections:
  Abstract/Introduction/Methods/Experiments/Conclusion

6. A good report will meet the following:
- Meet the requirements well.
- Well organized, and structured in a sensible way.
- Clear explanation of the topic and the methods, easy to read.
- Comprehensive and in-depth analysis.
- Insightful and inspiring discussion.
- Result visualizations.

About teamwork

Responsibilities for every team member:

1. Ensure all members work together and everyone is involved and contributes to the project.

2. Sufficient coordination and communication in the team.
   - If a team member fails to contribute due to insufficient communication and coordination, the entire team will face penalties.

3. Report an issue
   - If a team member refuses to contribute, please email the instructor at least 4 weeks before the submission deadline.

   - If you cannot report the issue 4 weeks before the deadline, it indicates that the entire team lacks communication and coordination. We will not handle any issues if they are not reported in time.

Contact:

Lin Guosheng <gslin@ntu.edu.sg>

Or send me message in MS Teams