CZ/CE 4032 DATA ANALYTICS & MINING

Project 2

Lin Guosheng

School of Computer Science and Engineering,
Nanyang Technological University

Project 2: Technical Review Report

- Group based project
  - Use the same grouping as for project 1
- What you need to submit:
  - A zip file containing your review report (in PDF) and your code.
  - 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 for downloading your file.
- Submission
  - Please submit your zip file in NTULearn:
    the course site -> Assignments -> submission link for project 2
  - 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: 22 Nov 2021

Requirements for the technical review report

- 1. Select a topic/task from the list below:
  - Recommender System
  - Similarity Search
  - Clustering
  - Graph clustering or graph community detection
  - Graph neural network
  - Any sub-topics that belong to the above topics
  - Any topics that use the above techniques

- 2. Select methods for review and comparison
  - Select at least two methods (algorithms)
    - You can choose any methods
      - They can be the ones you learn in our course

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

Notes:

If you would like to work on the topics below

- Graph clustering or graph community detection
- Graph neural network

You may refer to the slides below form the Stanford course CS224w

1. CS224w course web page:
http://snap.stanford.edu/class/cs224w-2020/

2. Graph community detection:
http://snap.stanford.edu/class/cs224w-2020/slides/13-communities.pdf

3. Graph neural network
http://snap.stanford.edu/class/cs224w-2020/slides/06-GNN1.pdf
http://snap.stanford.edu/class/cs224w-2020/slides/07-GNN2.pdf
http://snap.stanford.edu/class/cs224w-2020/slides/08-GNN-application.pdf

- 3. Select at least 1 dataset
  - Any datasets, including public datasets or the dataset you create.
  - No requirement on the size/scale of the dataset
  - Using more than one dataset will be a plus

- 4. Experimental analysis and comparison
  - Should not copy results from others (like published papers)
    - You need to generate the results.
  - 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 discussion
  - Illustrate and discuss the successful cases and failed cases (if applicable)
  - Other possible analyses ….

- 5. Format
  - At least 3 pages, no upper limit. (font-size: 11pt or 12pt)
  - Research paper format
    - Include the following sections:

    Abstract/Introduction/related work/methods/experiments/conclusion

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

Contact:

Lin Guosheng <gslin@ntu.edu.sg>

Or send me message in MS Teams