CS 6220: Course Project Description

Instructor: Yizhou Sun

9/29/2014

General Goal

- Apply data mining algorithms and techniques to real-world problem
  - Formalize the problem into a data mining task
  - Apply or propose solutions to the task
  - Evaluate different solutions you have proposed

Dataset (ArnetMiner Publication Data)

- Data:
  - Provided by ArnetMiner (ArnetMiner.org)
- #index ---- index id of this paper
  #* ---- paper title
  #@ ---- authors (separated by semicolons)
  #t ---- year
  #c ---- publication venue
  #% ---- the id of references of this paper (there are multiple lines, with each indicating a reference)
  #! ---- abstract

Problem

- Goal:
  - Citation prediction for papers. Our goal is to predict the top-10 references for a given paper published in 2013, based on the information such as abstract, authors, venue, and title of the paper.

Grading

- Group formation (1 point)
  - 3-4 people per group
  - Where to submit: Sign-up in blackboard
  - When to submit: by this Sunday (9/21/14, 11:59pm)
  - What to submit: Group name, group members, group leader
- Midterm report (4 points) (deadline: 10/19/14)
  - A first submission in Kaggle
  - A report indicating your solution plan
- Kaggle competition result, final report and code (25 points) (Deadline: 12/8/14)
  - Kaggle competition result (based on up to 3 different solutions): 15 points
  - Report and code: 10 points

Kaggle Inclass Link

- http://inclass.kaggle.com/c/nu-cs6220-14f/
  - Every team can only have one account: the same name as your group name (GroupID_GroupName)
  - You can select up to 3 versions for your final review
  - The final testing file would be a superset to the existing one

Collaboration Rules

- Every team member get the same score
  - Exception: the team has the right to claim someone as a freerider, and votes to downgrade his/her score
- In the final report, you need to include a table describing your work distribution
  - e.g.,

Task	People
1. Collecting and preprocessing data	Student A
2. Implementing Algorithm 1	Student B
3. Implementing Algorithm 2	Student C and D
4. Evaluating and comparing algorithms	Student A
5. Writing report	Student B and C

- Finally, you will be asked to submit a peer evaluation form (only be seen to the instructor and TAs)