

CS423: Data mining

Assignment 2: Recommendation System

February 20, 2019

Instruction

In this assignment, you will be implementing a content-based recommendation system for recommending movies to users. The system will learn from a dataset from MovieLens provided on the course's webpage.

- The file *movies.csv* contains information of roughly 9000 movies in the database. Each movie is described by its (multiple) genres. For example “Pacific Rim: Uprising” is categorised as Action, Fantasy and Sci-fi.
- The file *rating.csv* contains information about user's rating. The score is between 0 and 5. This will be useful for constructing a utility matrix.

For simplicity, we will take genres as movie's features. You may need to parse the data which is originally in csv format into your program before training the system. Once finish use your system to answer the following questions.

1. Which user most frequently gives out ratings?
2. Which movie has been watched (and voted) the most ?
3. What are top 3 movies does your program recommend to $userID=LastThreeDigitsOfYourStudentNo$ when Euclidean distance is used as similarity measure ?
4. What are the 3 movies that $userID=LastThreeDigitsOfYourStudentNo$ will dislike, when cosine similarity is used ?

Useful Julia packages

CSV, DataFrame

What to hand in ?

A zipfile containing

1. A report answering the questions posed above. (HW2_5xxxxxxx.pdf)
2. A source code of your recommendation system (HW2_5xxxxxxx.jl)

Email the zipfile to `jakramate.b@cmu.ac.th` using ‘[CS423-HW2]’ prefix in your email title.