note **31** views

HW4: Movie Recommender System

HW4: Movie Recommender System

Published Date:
Oct. 24, 2016, 1:45 p.m.
Deadline Date:
Nov. 14, 2016, 5 p.m.
Description:
Description:
****

This is a team assignment with maximum size of 2. Deadline is 11/14/2016 4:29 PM EST.
****

Overview and Assignment Goals:
The objective of this assignment are the following:

- Develop a Recommender System that Uses the Rating Information(Matrix) and Side-Information (i.e., Additional Content).
- Think about using classification, clustering, or any ideas you may have
- The Scoring Metric will Root Mean Squared Error where the predictions are rating in the range 1-5; As such, you may want to switch to a regression setting since the output can be real-valued.

Detailed Description:
Recommender Systems are all pervasive. The objective of this movie recommender system is to predict the 5-star rating a movie will get for a given user. You can use content features as well as the rating matrix to make your final predictions

Data Description:
As part of the training I provide you several different files all zipped together as additional_files.tar.gz (Uploaded in the training portion of this assignment)

Once you unzip this file you will find a readme.txt with a listing of the files and useful information about them (replicated below)
* train.dat
This file containts the rating of a user for a give movie.
*test.dat
This file contains user movie pairs but no rating **(Your goal is to predict these ratings for user-movie pairs)**
* movie_genres.dat

This file contains the genres of the movies.

* movie_directors.dat

This file contains the directors of the movies.

* movie_actors.dat

This file contains the main actores and actresses of the movies. A ranking is given to the actors of each movie according to the order in which they appear on the movie IMDb cast web page.
* tags.dat

This file contains the set of tags available in the dataset.
* user_taggedmovies.dat

These files contain the tag assignments of the movies provided by each particular user.

* movie_tags.dat

This file contains the tags assigned to the movies, and the number of times the tags were assigned to each movie.
test.dat: Test set consistting of user-movie pairs for which you need to produce the ratings

example_entry.dat: A sample submission with 71299 entries in the range of 1-5

Rules:

- This is an team assignment. Discussion of broad level strategies are allowed but any copying of prediction files and source codes will result in honor code violation.
- Feel free to use the programming language of your choice for this assignment.
- While you can use libraries and templates for dealing with this problem. However, you should be able to explain these methods and their choice in sufficient detail.
- You are allowed 5 submissions in a 24 hour cycle.

Deliverables:

- Valid Submissions to the Miner.vsnet.gmu.edu website
- **Blackboard Submission of Source Code and Report:**
-
  - Create a folder called HW4_LastName1_LastName2
  - Create a subfolder called src and put all the source code there.
  - Create a subfolder called Report and place a 2-Page, single-spaced report describing details regarding the steps you followed for developing the recommender system. Be sure to include the following in the report:

1. 1. Team Name(s) Registered on miner web-site.
   2. Rank & RMSE score for your submission (at the time of writing the report).
   3. Your Approach
   4. Your methodology of choosing the approach, the content features you used if any and what worked, what did not?.

   - Archive your parent folder (.zip or .tar.gz) and submit via Blackboard for HW4.

Grading:
Grading for the Assignment will be split on your implementation (50%), report (20%) and ranking results (30%).

Grading:
Grading for the Assignment will be split on your implementation (50%), report (20%) and ranking results (30%).

Files:

- *Train Data:* Download File
- *Test Data:* Download File
- *Format File:* Download File

hw4

Updated 7 days ago by Huzefa Rangwala

followup discussions *for lingering questions and comments*