Documentation

1. An overview of the function of the code

Scraper of Netflix with video name, image url, netflix url, synopsis, casts, creators, mood tag
Storage of information into csv file and reading from csv file
Web api using flask with get request including get one video info by id, get all videos info, get recommendation by video id
Basic recommender system for content based recommendation of a video
Web page content rendering using react

2. How the software is implemented

Scraper is implemented with BeautifulSoup and webdriver from selenium. It scrapes Netflix with video name, image url, netflix url, synopsis, casts, creators, mood tag and store them into python dictionary, which will be appended to a list. The list will then be stored into csv file.

Recommender system is implemented using content based filtering with cosine similarity. Features including name, synopsis, mood tag, genres, casts, creators are used to generate a combined features, and stops words are removed. The result string is then feed for CountVectorizer, which is then used to get the count matrix then cosine similarity matrix. At last, sort the result by the similarity scores and get the top 10 similar videos.

Web api is inplemented using flask. All three api need to read scv file to get data. Get recommendation by video id method also need to use recommender.py to get 10 list of recommendations by the index of video in csv file.

Web content is rendered using react framework. Three react components are implemented including List, Detail, Recommendation. List is used to display all the netflix videos from csv file. Detail is used to display the detailed information of a video. Recommendation is used to display 10 recommended videos below the details section of that video.

3. Self-evaluation

Have you completed what you have planned?

Yes, I have completed all the tasks that I have planned including scraper, recommender system, web api, web content.

Have you got the expected outcome?

Yes, the scraper can scraper about 200 videos from netflix, web api works well, recommender system can recommend similar videos in my scraped data properly. Since I only have about 200 videos which is too small as a database, the result of my recommendation is not very accurate compared with the recommendation from netflix website. However, for the videos that have many similar videos in database like animation and TV dramas, the recommendation works very good. The web content rendered using react framework works just as I expected.

4. Setup instructions of the software

Install python package

pip install pandas
pip install sklearn
pip install nltk
python -m nltk.downloader stopwords
pip install collections
pip install flask
pip install flask-cors
pip install csv
pip install bs4
pip install selenium
pip install time

Install Node.js and npm

Download Node.js by this link: https://nodejs.org/en/download/
If node and npm is not detected after downloading, try to restart your computer.

5. Usage of the software

1. If you want to create a complete new data csv file by using scraper, then follow this instruction: (Note that this will rewrite videos.csv file and scraping also takes a long time (approximately 20 minutes))

Run netflix scraper in the root directory by command:
python -m scraper.main

If you just want to test scraper, then follow these instructions:
Go to 'constant.py' in 'scraper' folder, change value of NUMBER_EACH_GENRE to 2, change value of NUMBER_SECTION to 3. Go to 'utils.py' and change value of CSV_TO_WRITE to 'test.csv'

Then run netflix scraper in the root directory by command:
python -m scraper.main

2. If you want to open the video recommender web application, then follow these instructions.

First navigate to the root directory, then run command:
python -m flask_app

Open a second terminal, navigate to the root directory, then run these commands:
cd video-recommender-app
npm install
npm start