

Amazon Recommendation System

In order to successfully run the program you need to have the following files in same directory. Link to download the data: <http://jmcauley.ucsd.edu/data/amazon/links.html>

1. project.py: This file contains the code to extract data and implement the algorithm used for the recommendation system
2. ProjectUI.ui: UI layout created in QT Creator that projects our product recommendations
3. UICreator.py: This file contains the code to integrate the ProjectUI.ui and the Project.py files
4. meta_Electronics.json.gz: This is the compressed file in loose json format that contains the transaction data we used. To download this go to the link provided above and download the metadata file for Electronics category from the last table on page. The table is named as Per-category files. Figure 1 below shows the file highlighted with yellow color. Direct link is also given below. **Save it is metadata.json.gz**
5. ratings_Electronics.csv: This file contains ratings by reviewers and can be downloaded using above link from the table named “Ratings only”. Figure 2 below shows the file highlighted with yellow color. Direct link is also given below. **Save it as ratings.csv**

Our code is tested for Electronics data. You can use our same algorithm for any category data from the same tables as for files # 4 and #5 mentioned above.

Direct Links for Electronics data

meta_Electronics.json.gz:

http://snap.stanford.edu/data/amazon/productGraph/categoryFiles/meta_Electronics.json.gz

ratings_Electronics.csv:

http://snap.stanford.edu/data/amazon/productGraph/categoryFiles/ratings_Electronics.csv

You need to run the project.py file

Note: When you run the code for a new category for the first time, make sure that the following code in lines 40-42 in project.py file is not commented. This piece of code is used to extract data from json.gz file and write it into another file “output.strict” after converting it into strict json format. Once you have the output.strict file for a category, the rest of the code uses that file only and you do not need to run the below code every time.

```
#f = open("output.strict", 'w')
#for l in parse("metadata.json.gz"):
# f.write(l + '\n')
```

Per-category files

Below are files for individual product categories, which have already had duplicate item reviews removed.

Books	reviews (22,507,155 reviews)	metadata (2,370,585 products)	image features
Electronics	reviews (7,824,482 reviews)	metadata (498,196 products)	image features
Movies and TV	reviews (4,607,047 reviews)	metadata (208,321 products)	image features
CDs and Vinyl	reviews (3,749,004 reviews)	metadata (492,799 products)	image features
Clothing, Shoes and Jewelry	reviews (5,748,920 reviews)	metadata (1,503,384 products)	image features
Home and Kitchen	reviews (4,253,926 reviews)	metadata (436,988 products)	image features
Kindle Store	reviews (3,205,467 reviews)	metadata (434,702 products)	image features
Sports and Outdoors	reviews (3,268,695 reviews)	metadata (532,197 products)	image features
Cell Phones and Accessories	reviews (3,447,249 reviews)	metadata (346,793 products)	image features
Health and Personal Care	reviews (2,982,326 reviews)	metadata (263,032 products)	image features
Toys and Games	reviews (2,252,771 reviews)	metadata (336,072 products)	image features
Video Games	reviews (1,324,753 reviews)	metadata (50,953 products)	image features
Tools and Home Improvement	reviews (1,926,047 reviews)	metadata (269,120 products)	image features
Beauty	reviews (2,023,070 reviews)	metadata (259,204 products)	image features
Apps for Android	reviews (2,638,173 reviews)	metadata (61,551 products)	image features
Office Products	reviews (1,243,186 reviews)	metadata (134,838 products)	image features
Pet Supplies	reviews (1,235,316 reviews)	metadata (110,707 products)	image features
Automotive	reviews (1,373,768 reviews)	metadata (331,090 products)	image features
Grocery and Gourmet Food	reviews (1,297,156 reviews)	metadata (171,760 products)	image features
Patio, Lawn and Garden	reviews (993,490 reviews)	metadata (109,094 products)	image features
Baby	reviews (915,446 reviews)	metadata (71,317 products)	image features
Digital Music	reviews (836,006 reviews)	metadata (279,899 products)	image features
Musical Instruments	reviews (500,176 reviews)	metadata (84,901 products)	image features
Amazon Instant Video	reviews (583,933 reviews)	metadata (30,648 products)	image features

Citation

Figure 1: metadata file

Ratings only: These datasets include no metadata or reviews, but only (user,item,rating,timestamp) tuples. Thus they are suitable for use with [mymedialite](#) (or similar) packages.

Books	5-core (8,898,041 reviews)	ratings only (22,507,155 ratings)
Electronics	5-core (1,689,188 reviews)	ratings only (7,824,482 ratings)
Movies and TV	5-core (1,697,533 reviews)	ratings only (4,607,047 ratings)
CDs and Vinyl	5-core (1,097,592 reviews)	ratings only (3,749,004 ratings)
Clothing, Shoes and Jewelry	5-core (278,677 reviews)	ratings only (5,748,920 ratings)
Home and Kitchen	5-core (551,682 reviews)	ratings only (4,253,926 ratings)
Kindle Store	5-core (982,619 reviews)	ratings only (3,205,467 ratings)
Sports and Outdoors	5-core (296,337 reviews)	ratings only (3,268,695 ratings)
Cell Phones and Accessories	5-core (194,439 reviews)	ratings only (3,447,249 ratings)
Health and Personal Care	5-core (346,355 reviews)	ratings only (2,982,326 ratings)
Toys and Games	5-core (167,597 reviews)	ratings only (2,252,771 ratings)
Video Games	5-core (231,780 reviews)	ratings only (1,324,753 ratings)
Tools and Home Improvement	5-core (134,476 reviews)	ratings only (1,926,047 ratings)
Beauty	5-core (198,502 reviews)	ratings only (2,023,070 ratings)
Apps for Android	5-core (752,937 reviews)	ratings only (2,638,172 ratings)
Office Products	5-core (53,258 reviews)	ratings only (1,243,186 ratings)
Pet Supplies	5-core (157,836 reviews)	ratings only (1,235,316 ratings)
Automotive	5-core (20,473 reviews)	ratings only (1,373,768 ratings)
Grocery and Gourmet Food	5-core (151,254 reviews)	ratings only (1,297,156 ratings)
Patio, Lawn and Garden	5-core (13,272 reviews)	ratings only (993,490 ratings)
Baby	5-core (160,792 reviews)	ratings only (915,446 ratings)
Digital Music	5-core (64,706 reviews)	ratings only (836,006 ratings)
Musical Instruments	5-core (10,261 reviews)	ratings only (500,176 ratings)
Amazon Instant Video	5-core (37,126 reviews)	ratings only (583,933 ratings)

Complete review data

Figure 2: ratings file