Evaluation Metrics for Recommender Systems

Installing and Importing Important Libraries like Surprise and recmetrics apart from Numpy and Pandas

```
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
!pip install surprise
import surprise
from surprise import SVD, Dataset, Reader
from surprise.model_selection import train_test_split
!pip install recmetrics
import recmetrics
```

```
Requirement already satisfied: keyring>=15.1 in /usr/local/lib/python3.10/dist-packages (from twine<5.0.0,>=4.0.0->recmetric
Requirement already satisfied: rfc3986>=1.4.0 in /usr/local/lib/python3.10/dist-packages (from twine<5.0.0,>=4.0.0->recmetri
Requirement already satisfied: rich>=12.0.0 in /usr/local/lib/python3.10/dist-packages (from twine<5.0.0,>=4.0.0->recmetrics
Requirement already satisfied: zipp>=0.5 in /usr/local/lib/python3.10/dist-packages (from importlib-metadata>=3.6->twine<5.0
Requirement already satisfied: parso<0.9.0,>=0.8.0 in /usr/local/lib/python3.10/dist-packages (from jedi>=0.16->ipython<8.0.
Requirement already satisfied: jaraco.classes in /usr/local/lib/python3.10/dist-packages (from keyring>=15.1->twine<5.0.0,>=
Requirement already satisfied: SecretStorage>=3.2 in /usr/local/lib/python3.10/dist-packages (from keyring>=15.1->twine<5.0.
Requirement already satisfied: jeepney>=0.4.2 in /usr/local/lib/python3.10/dist-packages (from keyring>=15.1->twine<5.0.0,>=
Requirement already satisfied: ptyprocess>=0.5 in /usr/local/lib/python3.10/dist-packages (from pexpect>4.3->ipython<8.0.0,>
Requirement already satisfied: wcwidth in /usr/local/lib/python3.10/dist-packages (from prompt-toolkit!=3.0.0,!=3.0.1,<3.1.0
Requirement already satisfied: attrs>=19.2.0 in /usr/local/lib/python3.10/dist-packages (from pytest>=4.6->pytest-cov<3.0.0,
Requirement already satisfied: iniconfig in /usr/local/lib/python3.10/dist-packages (from pytest>=4.6->pytest-cov<3.0.0,>=2.
Requirement already satisfied: pluggy<2.0,>=0.12 in /usr/local/lib/python3.10/dist-packages (from pytest>=4.6->pytest-cov<3.
Requirement already satisfied: exceptiongroup>=1.0.0rc8 in /usr/local/lib/python3.10/dist-packages (from pytest>=4.6->pytest
Requirement already satisfied: tomli>=1.0.0 in /usr/local/lib/python3.10/dist-packages (from pytest>=4.6->pytest-cov<3.0.0,>
Requirement already satisfied: bleach>=2.1.0 in /usr/local/lib/python3.10/dist-packages (from readme-renderer>=35.0->twine<5
Requirement already satisfied: docutils>=0.13.1 in /usr/local/lib/python3.10/dist-packages (from readme-renderer>=35.0->twir
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.10/dist-packages (from requests>=2.20->twine<5.0
Requirement already satisfied: charset-normalizer~=2.0.0 in /usr/local/lib/python3.10/dist-packages (from requests>=2.20->tw
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests>=2.20->twine<5.0.0,>=4
Requirement already satisfied: markdown-it-py<3.0.0,>=2.2.0 in /usr/local/lib/python3.10/dist-packages (from rich>=12.0.0->t
Requirement already satisfied: ipython-genutils in /usr/local/lib/python3.10/dist-packages (from ipykernel->jupyter<2.0.0,>=
Requirement already satisfied: jupyter-client in /usr/local/lib/python3.10/dist-packages (from ipykernel->jupyter<2.0.0,>=1.
Requirement already satisfied: tornado>=4.2 in /usr/local/lib/python3.10/dist-packages (from ipykernel->jupyter<2.0.0,>=1.0.
Requirement already satisfied: widgetsnbextension~=3.6.0 in /usr/local/lib/python3.10/dist-packages (from ipywidgets->jupyte
Requirement already satisfied: jupyterlab-widgets>=1.0.0 in /usr/local/lib/python3.10/dist-packages (from ipywidgets->jupyte
Requirement already satisfied: lxml in /usr/local/lib/python3.10/dist-packages (from nbconvert->jupyter<2.0.0,>=1.0.0->recme
Requirement already satisfied: beautifulsoup4 in /usr/local/lib/python3.10/dist-packages (from nbconvert->jupyter<2.0.0,>=1.
Requirement already satisfied: defusedxml in /usr/local/lib/python3.10/dist-packages (from nbconvert->jupyter<2.0.0,>=1.0.0-
Requirement already satisfied: entrypoints>=0.2.2 in /usr/local/lib/python3.10/dist-packages (from nbconvert->jupyter<2.0.0,
Requirement already satisfied: jinja2>=3.0 in /usr/local/lib/python3.10/dist-packages (from nbconvert->jupyter<2.0.0,>=1.0.0
Requirement already satisfied: jupyter-core>=4.7 in /usr/local/lib/python3.10/dist-packages (from nbconvert->jupyter<2.0.0,>
Requirement already satisfied: jupyterlab-pygments in /usr/local/lib/python3.10/dist-packages (from nbconvert->jupyter<2.0.0
Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.10/dist-packages (from nbconvert->jupyter<2.0.0,>=1
Requirement already satisfied: mistune<2,>=0.8.1 in /usr/local/lib/python3.10/dist-packages (from nbconvert->jupyter<2.0.0,>
Requirement already satisfied: nbclient>=0.5.0 in /usr/local/lib/python3.10/dist-packages (from nbconvert->jupyter<2.0.0,>=1
Requirement already satisfied: nbformat>=5.1 in /usr/local/lib/python3.10/dist-packages (from nbconvert->jupyter<2.0.0,>=1.0
Requirement already satisfied: pandocfilters>=1.4.1 in /usr/local/lib/python3.10/dist-packages (from nbconvert->jupyter<2.0.
Requirement already satisfied: tinycss2 in /usr/local/lib/python3.10/dist-packages (from nbconvert->jupyter<2.0.0,>=1.0.0->r
Requirement already satisfied: pyzmq>=17 in /usr/local/lib/python3.10/dist-packages (from notebook->jupyter<2.0.0,>=1.0.0->r
Requirement already satisfied: argon2-cffi in /usr/local/lib/python3.10/dist-packages (from notebook->jupyter<2.0.0,>=1.0.0-
Requirement already satisfied: nest-asyncio>=1.5 in /usr/local/lib/python3.10/dist-packages (from notebook->jupyter<2.0.0,>=
Requirement already satisfied: Send2Trash>=1.8.0 in /usr/local/lib/python3.10/dist-packages (from notebook->jupyter<2.0.0,>=
Requirement already satisfied: terminado>=0.8.3 in /usr/local/lib/python3.10/dist-packages (from notebook->jupyter<2.0.0,>=1
Requirement already satisfied: prometheus-client in /usr/local/lib/python3.10/dist-packages (from notebook->jupyter<2.0.0,>=
Requirement already satisfied: qtpy>=2.0.1 in /usr/local/lib/python3.10/dist-packages (from qtconsole->jupyter<2.0.0,>=1.0.0
Requirement already satisfied: webencodings in /usr/local/lib/python3.10/dist-packages (from bleach>=2.1.0->readme-renderer>
Requirement already satisfied: platformdirs>=2.5 in /usr/local/lib/python3.10/dist-packages (from jupyter-core>=4.7->nbconve
Requirement already satisfied: mdurl~=0.1 in /usr/local/lib/python3.10/dist-packages (from markdown-it-py<3.0.0,>=2.2.0->ric
Requirement already satisfied: fastjsonschema in /usr/local/lib/python3.10/dist-packages (from nbformat>=5.1->nbconvert->jup
Requirement already satisfied: jsonschema>=2.6 in /usr/local/lib/python3.10/dist-packages (from nbformat>=5.1->nbconvert->ju
Requirement already satisfied: cryptography>=2.0 in /usr/local/lib/python3.10/dist-packages (from SecretStorage>=3.2->keyrin
Requirement already satisfied: argon2-cffi-bindings in /usr/local/lib/python3.10/dist-packages (from argon2-cffi->notebook->
Requirement already satisfied: soupsieve>1.2 in /usr/local/lib/python3.10/dist-packages (from beautifulsoup4->nbconvert->jup
Requirement already satisfied: more-itertools in /usr/local/lib/python3.10/dist-packages (from jaraco.classes->keyring>=15.1
Requirement already satisfied: cffi>=1.12 in /usr/local/lib/python3.10/dist-packages (from cryptography>=2.0->SecretStorage>
Requirement already satisfied: pyrsistent!=0.17.0,!=0.17.1,!=0.17.2,>=0.14.0 in /usr/local/lib/python3.10/dist-packages (fro
Requirement already satisfied: pycparser in /usr/local/lib/python3.10/dist-packages (from cffi>=1.12->cryptography>=2.0->Sec
```

Importing CSV files for analysis. This data contains user's ratings of movies, as well as movie genre tag.

```
ratings = pd.read_csv('/content/ratings.csv')
ratings = ratings.query('rating >=3')
ratings.reset_index(drop=True, inplace=True)
```

```
ratings.head()
```

	userId	movieId	rating	timestamp
0	1	296	5.0	1147880044
1	1	306	3.5	1147868817
2	1	307	5.0	1147868828
3	1	665	5.0	1147878820
4	1	899	3.5	1147868510

```
#only consider ratings from users who have rated over n movies
n=1000
users = ratings["userId"].value_counts()
users = users[users>n].index.tolist()
```

```
ratings = ratings.query('userId in @users')
print(ratings.shape)
ratings.head(3)
```

```
(81874, 4)
```

	userId	movieId	rating	timestamp
19897	187	1	3.5	1277374478
19898	187	2	3.5	1277374864
19899	187	3	3.0	1277839361

```
rated_movies = ratings["movieId"].tolist()

movies = pd.read_csv('/content/movies.csv')
movies = movies.query('movieId in @rated_movies')
movies.set_index("movieId", inplace=True, drop=True)

movies = movies["genres"].str.split("|", expand=True)
movies.reset_index(inplace=True)
movies = pd.melt(movies, id_vars='movieId', value_vars=[0, 1, 2, 3, 4, 5, 6, 7, 8, 9])

movies.drop_duplicates("movieId", inplace=True)
movies.set_index('movieId', inplace=True)

movies = pd.get_dummies(movies.value)
movies.head()
```

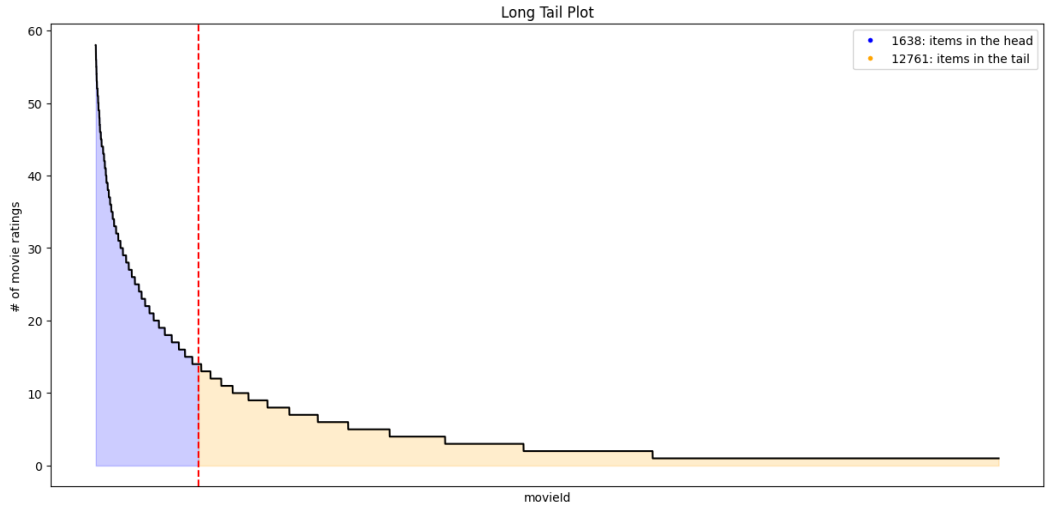
e	Animation	Children	Comedy	Crime	Documentary	Drama	Fantasy	Film-Noir	Horror	Musical	Myster
1	0	0	0	0	0	0	0	0	0	0	
1	0	0	0	0	0	0	0	0	0	0	
)	0	0	1	0	0	0	0	0	0	0	
)	0	0	1	0	0	0	0	0	0	0	
)	0	0	1	0	0	0	0	0	0	0	

Long Tail Plot

This plot is used to explore popularity patterns in user-item interaction data such as clicks, ratings, or purchases. Typically, only a small percentage of items have a high volume of interactions, and this is referred to as the “head”. Most items are in the “long tail”, but they only make up a small percentage of interactions.

```
fig = plt.figure(figsize=(15, 7))
recmetrics.long_tail_plot(df=ratings,
             item_id_column="movieId",
             interaction_type="movie ratings",
             percentage=0.5,
             x_labels=False)
```

```
/usr/local/lib/python3.10/dist-packages/seaborn/_decorators.py:36: FutureWarning: Pass the fol
  warnings.warn(
/usr/local/lib/python3.10/dist-packages/recmetrics/plots.py:60: FutureWarning: The frame.apper
  head = head.append(tail.head(1))
```




```
reader = Reader(rating_scale=(0, 5))
data = surprise.Dataset.load_from_df(ratings[['userId', 'movieId', 'rating']], reader)
```

```
trainset, testset = train_test_split(data, test_size=0.25)
```

SVD reduces the dimensionality reduction and gets us important latent features which can almost approximate all values in the utility matrix (including null values). Already available values in the utility matrix can be used to evaluate the predictions and tune the parameters/weights in the latent features (if matrix factorization used) or help us to pick the number of latent features from the SVD decomposition. Once number of latent features to be picked are decided, we can populate the null values and use these predicted ratings of a user for movie to recommend the movie with highest predicted rating to that particular user.

```
algo = SVD()
algo.fit(trainset)
```

```
<surprise.prediction_algorithms.matrix_factorization.SVD at 0x7f9c65a104c0>
```

Making Predictions based on the model created above

```
test = algo.test(testset)
test = pd.DataFrame(test)
test.drop("details", inplace=True, axis=1)
test.columns = ['userId', 'movieId', 'actual', 'cf_predictions']
test.head()
```

	userId	movieId	actual	cf_predictions
0	5114	91542	3.0	3.699022
1	2403	69604	3.5	3.666884
2	1652	99114	4.0	4.346246
3	7248	1361	4.0	3.791791

Evaluate model with MSE and RMSE

```
print("MSE: ", recmetrics.mse(test.actual, test.cf_predictions))
print("RMSE: ", recmetrics.rmse(test.actual, test.cf_predictions))
```

```
MSE:  0.30261220628201635
RMSE:  0.5501019962534369
```

Double-click (or enter) to edit

✓ 0s completed at 1:20 PM