A Movie Recommendation Service

OMSBA 5312

August 28th, 2020

Jennifer Grosz

Use Case Name: A Movie Recommendation Service

Use Case URL:
https://www.codementor.io/@jadianes/building-a-recommender-with-apache-spark-python-example-app-part1-du1083qbw

Introduction

- Built a Movie Recommendation Model by analyzing user's preferences and other similar users' preferences to provide the best possible movie recommendations
- Used data from MovieLens website
- Collaborative Filtering and matrix factorization techniques using Spark's Alternating Least Squares have been used for the implementation
- End result: a model that recommends the top 15 movies for a specific user

Data Used

ml-latest-small:

- 100,836 ratings
- 3,683 tag applications
- 9,742 movies
- 610 users

ml-latest:

- 27,753,444 ratings
- 1,108,997 tag applications
- 58,098 movies
- 283,228 users

Both datasets describes 5-star movie rating and tagging activity

Technical Details

- Platform
  - Amazon Web Services (AWS)
    - EC2 Instance
      - T2.medium
      - Ubuntu Connection
      - Jupyter Notebook
        - Python 3
- Library
  - Spark MLlib
    - Collaborative Filtering
    - Matrix Factorization
    - Alternating Least Squares (ALS)
      - Root Mean Square Error (RMSE)

Debugging Details

- Create a SparkContext configured for local mode. This code was not included in the tutorial.
  - import pyspark
  - sc = pyspark.SparkContext('local[*]')
- Update download location path from what was provided in the tutorial.
  - Tutorial code:
    - datasets_path = os.path.join('..', 'datasets')
  - Modified code:
    - datasets_path = os.path.join('/home/jovyan', 'work')

Debugging Details Cont.

- Modified the code in order to download the files.
  - Tutorial code:
    - import urllib
    - small_f = urllib.urlretrieve (small_dataset_url, small_dataset_path)
    - complete_f = urllib.urlretrieve (complete_dataset_url, complete_dataset_path)
  - Modified code:
    - import urllib.request
    - small_f = urllib.request.urlretrieve (small_dataset_url, small_dataset_path)
    - complete_f = urllib.request.urlretrieve (complete_dataset_url, complete_dataset_path)
- Added print commands for illustrative purposes after parsing the small_ratings_file (ratings.csv) and small_movies_data (movies.csv) files into new RDDs.
  - Added code:
    - *# From section rating.csv*
    - print ('There are {} recommendations in the small dataset'.format(small_ratings_data.count()))
    - *# From section movies.csv*
    - print ('There are {} movies in the small dataset'.format(small_movies_data.count()))

Debugging Cont.

- Please refer to *Appendix B #5* for details on converting the Python 2 tutorial source code to Python 3 code
- Identified missing "\" in tutorial code when training the recommender model for the complete dataset
  - Tutorial code:
    - complete_model = ALS.train(training_RDD, best_rank, seed=seed,
    - iterations=iterations, lambda_=regularization_parameter)
  - Modified code:
    - complete_model = ALS.train(training_RDD, best_rank, seed=seed, \
    - iterations=iterations, lambda_=regularization_parameter)

Results – User1

Given Movie Ratings for **User1**:

- Movie: 300 (2007) - Rating: 5
- Movie: Big Lebowski, The (1998) - Rating: 5
- Movie: Liar Liar (1997) - Rating: 5
- Movie: American History X (1998) – Rating: 4
- Movie: Departed, The (2006) – Rating: 5
- Movie: Tom Segura: Disgraceful (2018) - Rating: 4
- Movie: Lord of the Rings: The Fellowship of the Ring, The (2001) - Rating: 5
- Movie: Conjuring, The (2013) - Rating: 4
- Movie: I Am Legend (2007) - Rating: 3
- Movie: Hereditary (2018) - Rating: 4

The new model for User1 was trained in 141.37 seconds.

Movie Recommendations for User1

TOP 15 recommended movies (with more than 25 reviews):

- Planet Earth II (2016) – Rating: 5.086000501239582 – No. of Reviews: 853
- Cosmos – Rating: 5.0400235412392504 – No. of Reviews: 157
- Planet Earth (2006) – Rating: 5.038495931163268 – No. of Reviews: 1384
- Band of Brothers (2001) – Rating: 5.0358961040492645 - No. of Reviews: 984
- Cosmos: A Spacetime Odissey – Rating: 4.969682616987463 – No. of Reviews: 37
- Black Mirror: White Christmas (2014) – Rating: 4.9575931055347695 – No. of Reviews: 1074
- The Godfather Trilogy: 1972-1990 (1992) – Rating: 4.932762991405745 – No. of Reviews: 421
- Shawshank Redemption – Rating: 4.914848872737281 – No. of Reviews: 97999
- The Reichenbach Fall (2012) – Rating: 4.834817936262212 – No. of Reviews: 48
- Human Planet (2011) – Rating: 4.817085115694141 – No. of Revies: 283
- The Blue Planet (2001) – Rating: 4.811914089159956 – No. of Reviews: 421
- Usual Suspects – Rating: 4.811001968486607 – No. of Reviews: 62180
- Godfather – Rating: 4.8071098594525665 – No. of Reviews: 60904
- Dave Chappelle: Killin' Them Softly (2000) – Rating: 4.797301138772596 – No. of Reviews: 56
- Over the Garden Wall (2013) – Rating: 4.792893725555963 – No. of Reviews: 377

TOP 15 recommended movies (with more than 100 reviews):

- Planet Earth II (2016) – Rating: 5.086000501239582 – No. of Reviews: 853
- Cosmos – Rating: 5.0400235412392504 – No. of Reviews: 157
- Planet Earth (2006) – Rating: 5.038495931163268 – No. of Reviews: 1384
- Band of Brothers (2001) – Rating: 5.0358961040492645 – No. of Reviews: 984
- Black Mirror: White Christmas (2014) – Rating: 4.9575931055347695 – No. of Reviews: 1074
- The Godfather Trilogy: 1972-1990 (1992) – Rating: 4.932762991405745 – No. of Reviews: 421
- Shawshank Redemption – Rating: 4.914848872737281 – No. of Reviews: 97999
- Human Planet (2011) – Rating: 4.817085115694141 – No. of Reviews: 283
- The Blue Planet (2001) – Rating: 4.811914089159956 – No. of Reviews: 421
- Usual Suspects – Rating: 4.811001968486607 – No. of Reviews: 62180
- Godfather – Rating: 4.8071098594525665 – No. of Reviews: 60904
- Over the Garden Wall (2013) – Rating: 4.792893725555963 – No. of Reviews: 377
- Blue Planet II (2017) – Rating: 4.792759622006393 – No. of Reviews: 349
- Death Note: Desu nôto (2006–2007) – Rating: 4.784354308191439 – No. of Reviews: 110
- Black Mirror – Rating: 4.779923722988428 – No. of Reviews: 180

Results – User2

Given Movie Ratings for **User2:**

- Movie: Man in the Iron Mask, The (1998) – Rating: 4
- Movie: Last Samurai, The (2003) – Rating: 3
- Movie: Hereditary (2018) – Rating: 4
- Movie: Incredibles, The (2004) – Rating: 4 (0, 8961, 4), #)
- Movie: Wedding Crashers (2005) - 3
- Movie: Freaky Friday (2018) – Rating: 1
- Movie: Bert Kreischer: Secret Time (2018) – Rating: 5
- Movie: Mean Girls (2004) – Rating: 4
- Movie: Saw (2004) - Rating: 4
- Movie: Forgetting Sarah Marshall (2008) – Rating: 5

The new model for User2 was trained in 122.748 seconds.

Movie Recommendations for User2

TOP 15 recommended movies (with more than 25 reviews):

- Elway To Marino (2013) – Rating: 5.257835236472239 – No. of Reviews: 25
- Connections (1978) – Rating: 5.202213372934146 – No. of Reviews: 49
- Cosmos – Rating: 5.177670126907799 – No. of Reviews: 157
- Rabbit of Seville (1950) – Rating: 5.163484150616505 – No. of Reviews: 30
- Baseball (1994) – Rating: 5.159287547672342 – No. of Reviews: 42
- Last Lions – Rating: 5.12409635143586 – No. of Reviews: 38
- Harakiri (Seppuku) (1962) – Rating: 5.113893949692688 – No. of Reviews: 679
- Jim Henson's The Storyteller (1989) – Rating: 5.1133157364957675 – No. of Reviews: 36
- Who's Singin' Over There? (a.k.a. Who Sings Over There) (Ko to tamo peva) (1980) – Rating: 5.104265150473019 – No. of Reviews: 45
- Century of the Self – Rating: 5.1040310208256585 – No. of Reviews: 213
- Human Condition III – Rating: 5.098909786081757 – No. of Reviews: 91
- In the blue sea – Rating: 5.085521273384728 – No. of Reviews: 37
- Planet Earth (2006) – Rating: 5.075628923151591 – No. of Reviews: 1384
- Planet Earth II (2016) – Rating: 5.073128889749725 – No. of Reviews: 853
- World of Tomorrow Episode Two: The Burden of Other People's Thoughts (2017) – Rating: 5.072479932851294 – No. of Reviews: 39

TOP 15 recommended movies (with more than 100 reviews):

- Cosmos – Rating: 5.177670126907799 – No. of Reviews: 157
- Harakiri (Seppuku) (1962) – Rating: 5.113893949692688 – No. of Reviews: 679
- Century of the Self – Rating: 5.1040310208256585 – No. of Reviews: 213
- Planet Earth (2006) – Rating: 5.075628923151591 – No. of Reviews: 1384
- Planet Earth II (2016) – Rating: 5.073128889749725 – No. of Reviews: 853
- Death on the Staircase (Soupçons) (2004) – Rating: 5.050088533873144 – No. of Reviews: 130
- Human Condition I – Rating: 5.020291878251406 – No. of Reviews: 151
- Trou – Rating: 5.018590309877705 – No. of Reviews: 189
- The Godfather Trilogy: 1972-1990 (1992) – Rating: 5.005446975265841 – No. of Reviews: 421
- All Watched Over by Machines of Loving Grace (2011) – Rating: 4.999820045022098 – No. of Reviews: 157
- Ikiru (1952) – Rating: 4.980738267656976 – No. of Reviews: 1551
- O.J.: Made in America (2016) – Rating: 4.979816561433598 – No. of Reviews: 431
- Seven Samurai (Shichinin no samurai) (1954) – Rating: 4.961174999608279 – No. of Reviews: 14578
- Larry David: Curb Your Enthusiasm (1999) – Rating: 4.958620844702779 – No. of Reviews: 177
- Come and See (Idi i smotri) (1985) – Rating: 4.953725519063973 – No. of Reviews: 703

Predicted Rating for Quiz Show (1994)

User1:

- 0.6211926549081777

User2:

- 0.9924341686034573

Insights

- Recommendation models provide content by taking what other people recommend as well as user explicit ratings into consideration
- Collaborative Filtering and Matrix Factorization that solve many of the common scalability and scarcity problems encountered when creating recommendation engines

Insights Cont.

- Recommendation model not only provides top personalized movie recommendations, but it also predicts user rating of movies they have not seen which creates unique content-customization opportunities for users
  - Personalize user experiences by providing predicted ratings for movies they have not seen
    - Increase user engagement with website
    - Increase traffic to website
    - Increase revenue

Resources

- Dianes, Jose. “Building a Movie Recommendation Service with Apache Spark & Flask - Part 1”. *Codementor*, September 14th, 2015. https://www.codementor.io/@jadianes/building-a-recommender-with-apache-spark-python-example-app-part1-du1083qbw
- F. Maxwell Harper and Joseph A. Konstan. 2015. The MovieLens Datasets: History and Context. ACM Transactions on Interactive Intelligent Systems (TiiS) 5, 4: 19:1–19:19. https://doi.org/10.1145/2827872
- Pal, Sayan Kumar. “Coding Standards and Guidelines”. *GeeksforGeeks*, Feburary 7th, 2019. https://www.geeksforgeeks.org/coding-standards-and-guidelines/
- Ajitsaria, Abhinav. “Build a Recommendation Engine with Collaborative Filtering”. *Real Python,* 2012 – 2020. https://realpython.com/build-recommendation-engine-collaborative-filtering/
- Lou, Shuyu. “Introduction to Recommender System Approaches of Collaborative Filtering: Nearest Neighborhood and Matrix Factorization”. *Towards Data Science*, December 10th, 2018. https://towardsdatascience.com/intro-to-recommender-system-collaborative-filtering-64a238194a26