Next Movies to Enjoy

A Movie Recommender on Deep Learning

Daihong Chen

Why Movie Recommender?

- Increasing Entertainment Market
- Personalization on movie recommendation
- Better engage customers on movie/tv product

Data Source & Properties

- ★ Amazon Movies/TV reviews from UCSD
- ★ Json.gz file with 19 years data (8,765,568 reviews)
- ★ Subsample to 2018 ratings/reviews (209,060 reviews)
- ★ Scrape review webpage link for each movie from the data
- ★ Drop unrelated variables

Data_2018 Exploratory

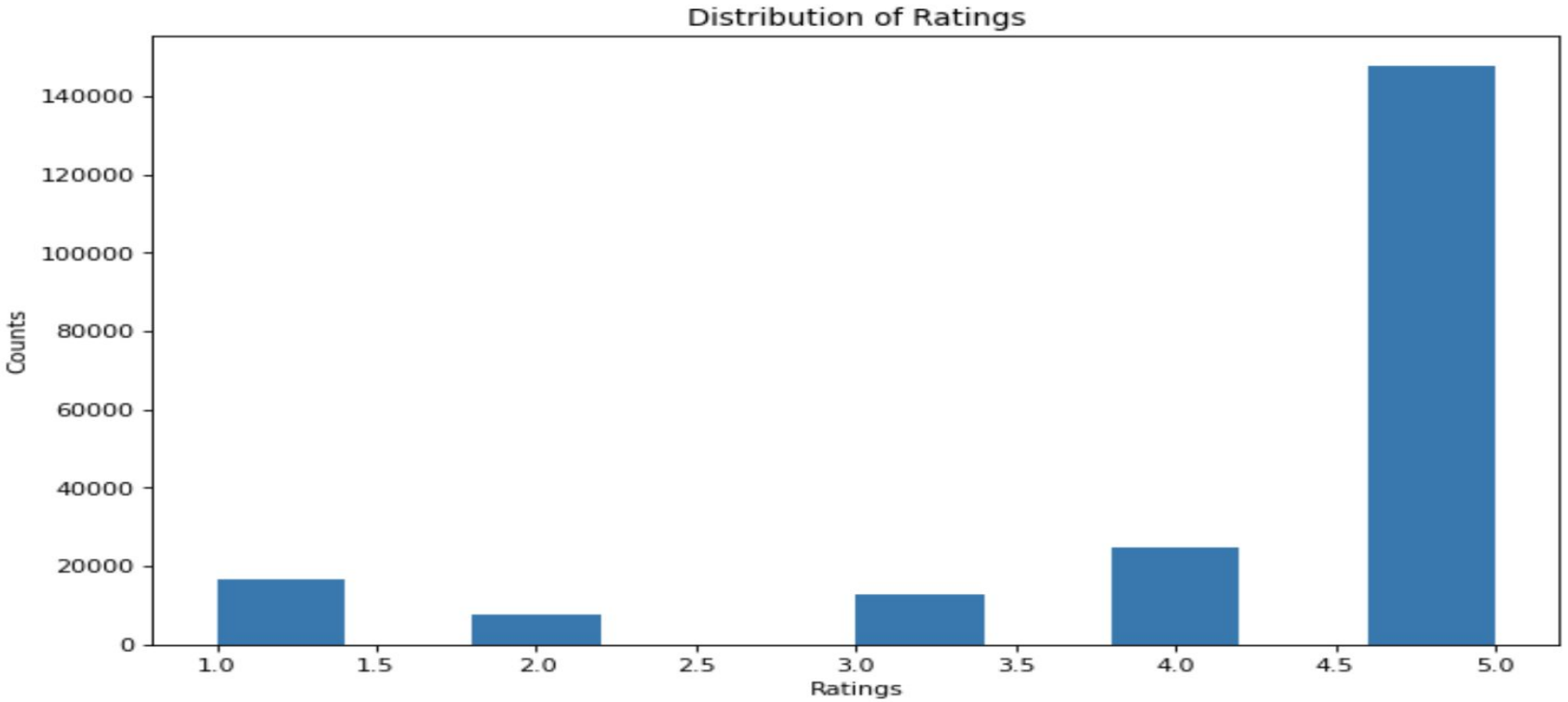
```
Count of Reviewers: 119945
Count of Products: 38846

Ratings descriptive statistics:
count    209060.000000
mean          4.336109
std           1.230864
min           1.000000
25%           4.000000
50%           5.000000
75%           5.000000
max           5.000000
Name: rating, dtype: float64
```

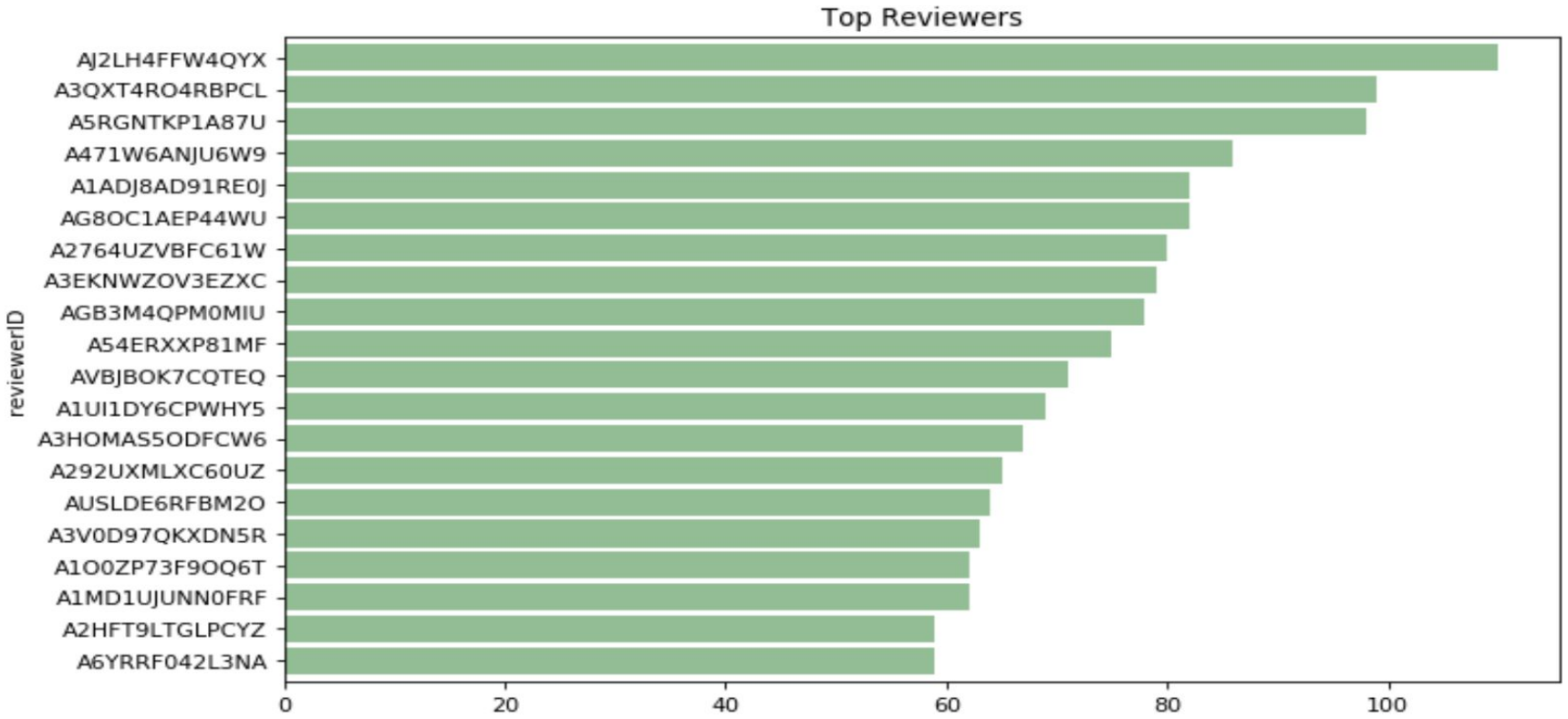
```
Reviewers by count descriptive statistics:
count    119945.000000
mean          1.742966
std           2.365081
min           1.000000
25%           1.000000
50%           1.000000
75%           2.000000
max         110.000000
Name: rating, dtype: float64
```

```
Products by count descriptive statistics:
count    38846.000000
mean         5.381378
std         17.192630
min          1.000000
25%          1.000000
50%          2.000000
75%          4.000000
max        959.000000
Name: rating, dtype: float64
```

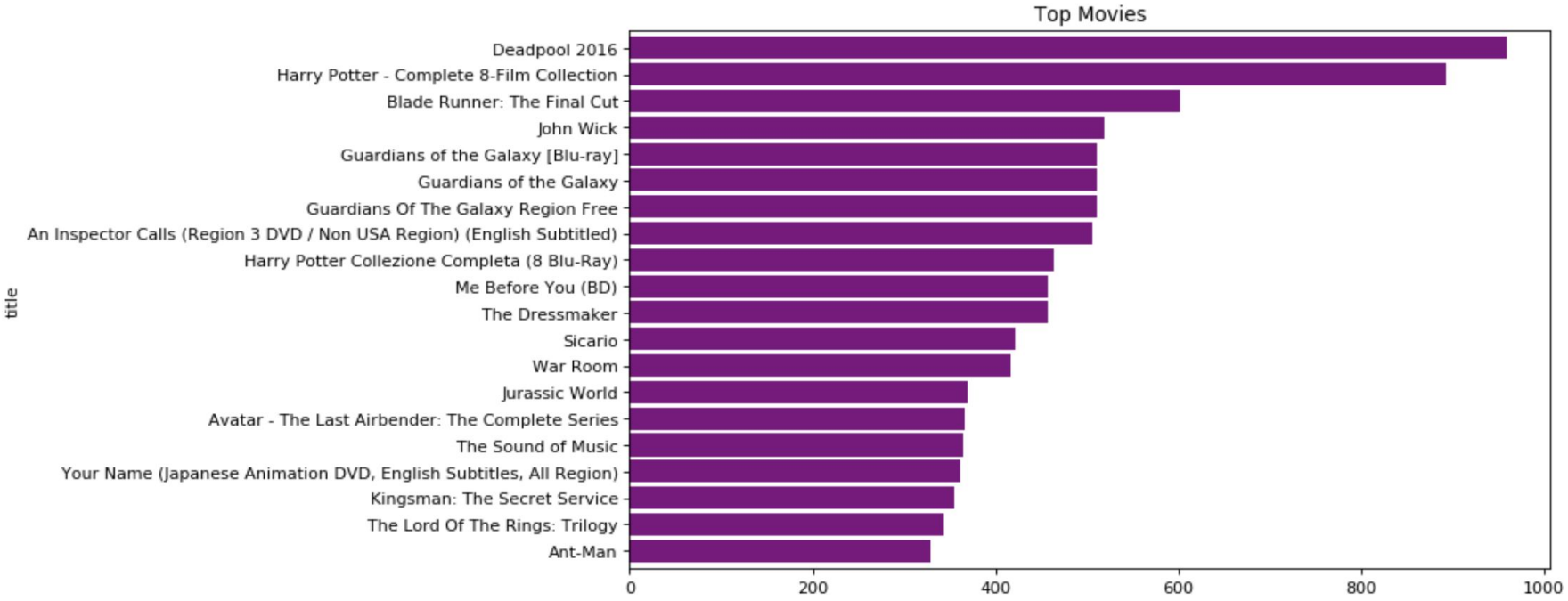
Data_2018 Exploratory



Data_2018 Exploratory



Data_2018 Exploratory



Base Model

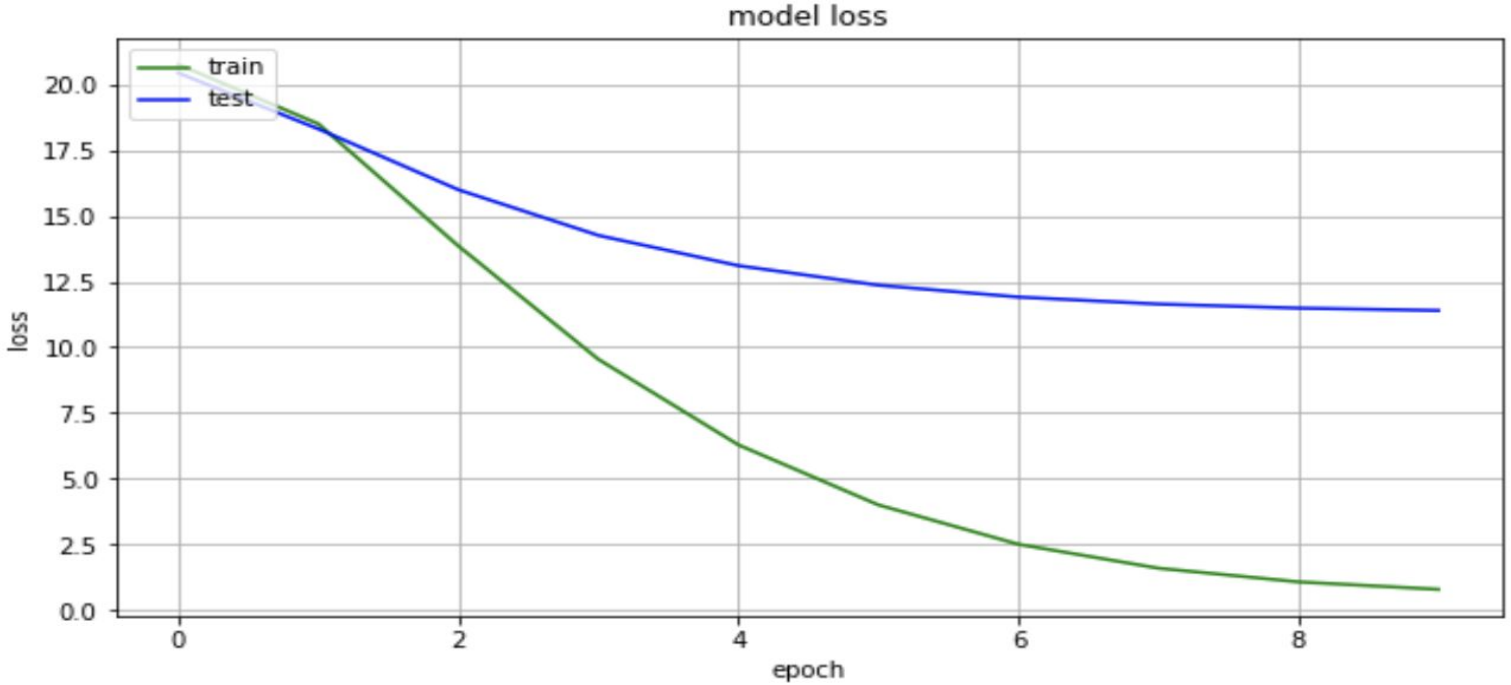
- ★ Collaborative Filtering

- ★ Keras Embedding
  - Create the reviewer embeddings and movie embeddings
    - Discomposing the utility matrix
  - Dot.product to merge two embeddings on reviewers and movies

- ★ Metrics: MAE

Base Model Evaluation

Matrics: MAE = 2.4

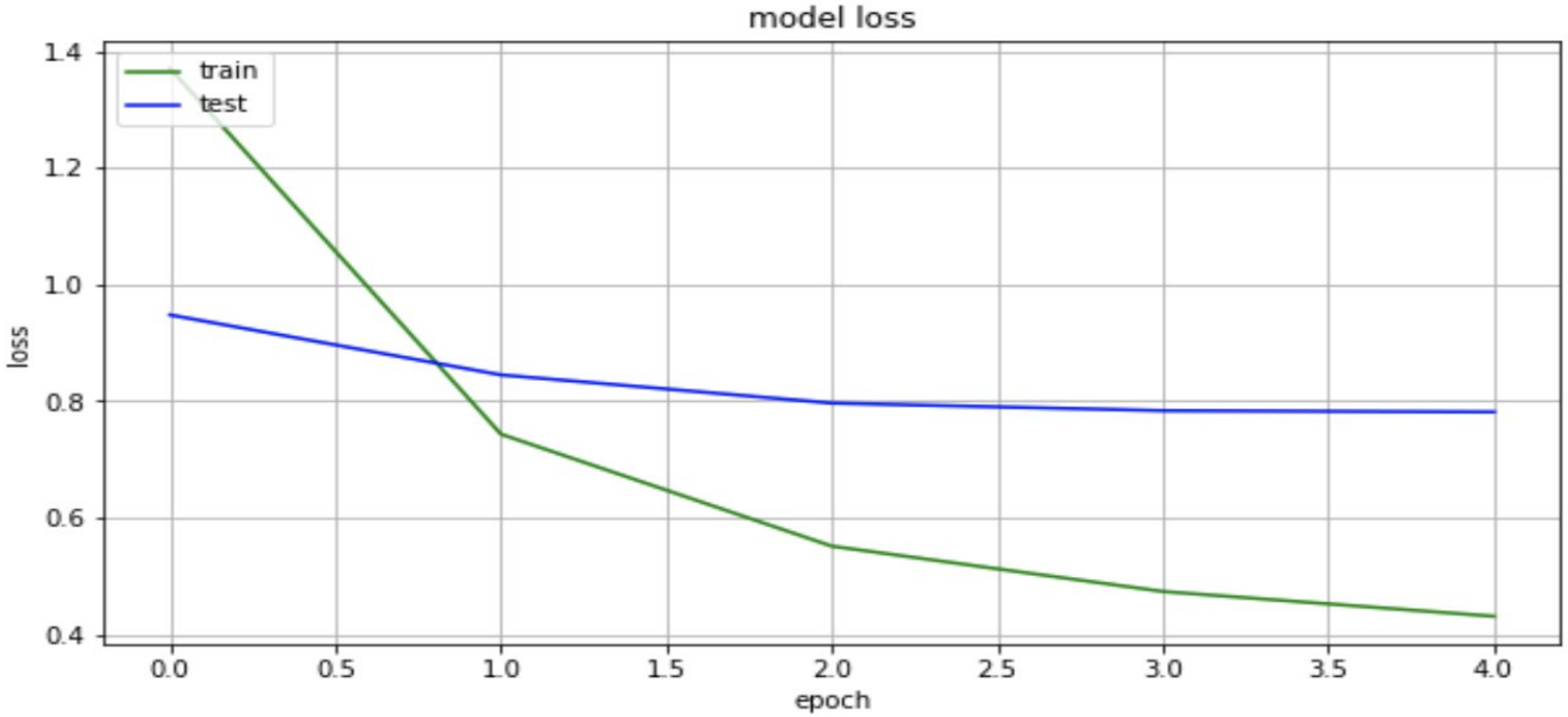


Final Model

- ★ Collaborative Filtering
- ★ Keras Embedding
  - Create the reviewer embeddings and movie embeddings
    - Discomposing the utility matrix
  - Dot.product to merge two embeddings on reviewers and movies
- ★ Metrics: MAE
- ★ Add hidden layers / dropout

Final Model Evaluation

Metrics: MAE = 0.58



Cross Validation on Final Model

Metrics: Mean Absolute Error (MAE)

StratifiedKFold, n_splits=5

1. 0.398,
2. 0.402,
3. 0.399,
4. 0.399,
5. 0.399

Average of MAE: 0.40

Standard Deviation of MAE: 0.00131

Make Recommendation

```
]: # load the recommeder
from src.recommend_movies import recommender
```

```
]: rec = recommender(1000)
```

```
/Users/daihongchen/opt/anaconda3/envs/learn-env/lib/python3.6/site-packages/tensorflow_core/python/framework/indexed_slices.py:433: UserWarning: Converting sparse IndexedSlices to a dense Tensor of unknown shape. This may consume a large amount of memory.
  "Converting sparse IndexedSlices to a dense Tensor of unknown shape. "
```

```
]: rec
```

]:

	movie	title	average_rating	category	description	price	links
19449	1162	Lean on Me	4.857143	['Movies & TV', 'Studio Specials', 'Warner Hom...	['Based on the true story of new jersey high s...	$15.00	https://www.amazon.com/product-reviews/6305133...
20855	2499	The Family Man	5.000000	['Movies & TV', 'Genre for Featured Categories...	NaN	$22.31	https://www.amazon.com/product-reviews/B000057...
34954	18583	Back to the Future 30th Anniversary Trilogy	4.586538	['Movies & TV', 'Boxed Sets', 'Comedy']	['Experience the future all over again with th...	$22.51	https://www.amazon.com/product-reviews/B011Q0F...
60138	4300	Evelyn	4.833333	['Movies & TV', 'Studio Specials', 'MGM Home E...	["From the director of Driving Miss Daisy come...	$16.67	https://www.amazon.com/product-reviews/B00008D...
85188	15127	Futurama: The Complete Series	4.383333	['Movies & TV', 'Studio Specials', '20th Centu...	['As vast and hilarious as the universe itself...	$68.45	https://www.amazon.com/product-reviews/B00F77M...