Nextale

A Book, Movie, and Video Game Recommender

Created by Will Griffin

Recommender systems: What do they do? Do they do things? Let's find out!

Why recommendations?

- To boost sales!
  - Amazon, Target
  - "You liked that? Try this too!"
- To sling that 'tent!
  - Netflix, Hulu
  - "That show is over, time to start this one!"
- To keep you engaged
  - Twitter, Facebook, News Apps
  - Keep reading, swiping, scrolling

My Recommender System

- Amazon customer reviews from 1995-2015 (20 years of data!)
  - Books
  - Movies
  - Video Games
- Shared narrative element, hence the title “Nextale”
- In Total
  - Over 7 million reviews!
  - More than 3.5 million unique customers
  - In excess of 133,000 products
- Unique Feature
  - An exclusion term: an additional search box for the user to filter something out of their results
  - Examples:
    - You want movies like “Saw” but you don’t like the “Final Destination” series
    - You want a video game like Animal Crossing but didn’t enjoy Harvest Moon

Cleaning

```
Initial size: 5.20014832
Initial shape: (5049291, 12)
Size after dropping products w/reviews < 10: 4.493609664
Null Preview: customer_id            0
review_id              0
product_id             0
product_parent         0
product_title          0
star_rating            0
helpful_votes          0
total_votes            0
verified_purchase      0
review_headline       58
review_body           26
review_date          374
dtype: int64
Null Percentage: 0.01%
Size after dropna(): 4.489704914
Size after concatenation: 4.245067022
Tokenizing, lemmatizing, and removing stopwords...hold please
Size after tokemmitization: 3.313707545
Final size: 3.313707545
Final shape: (4405432, 11)
[3]: 'File saved as movie_dvd.csv'
```

Raw file - more than 5 GB, over 5 million rows!

Minimum 10 reviews

All nulls represent less than .01% of total dataset

Combine review titles and text

Tokenize, lemmatize, and remove stop words to simplify potential future NLP AND save memory

In total, cut almost 2 GB from the file size while preserving 4.4 million rows

How does it work?

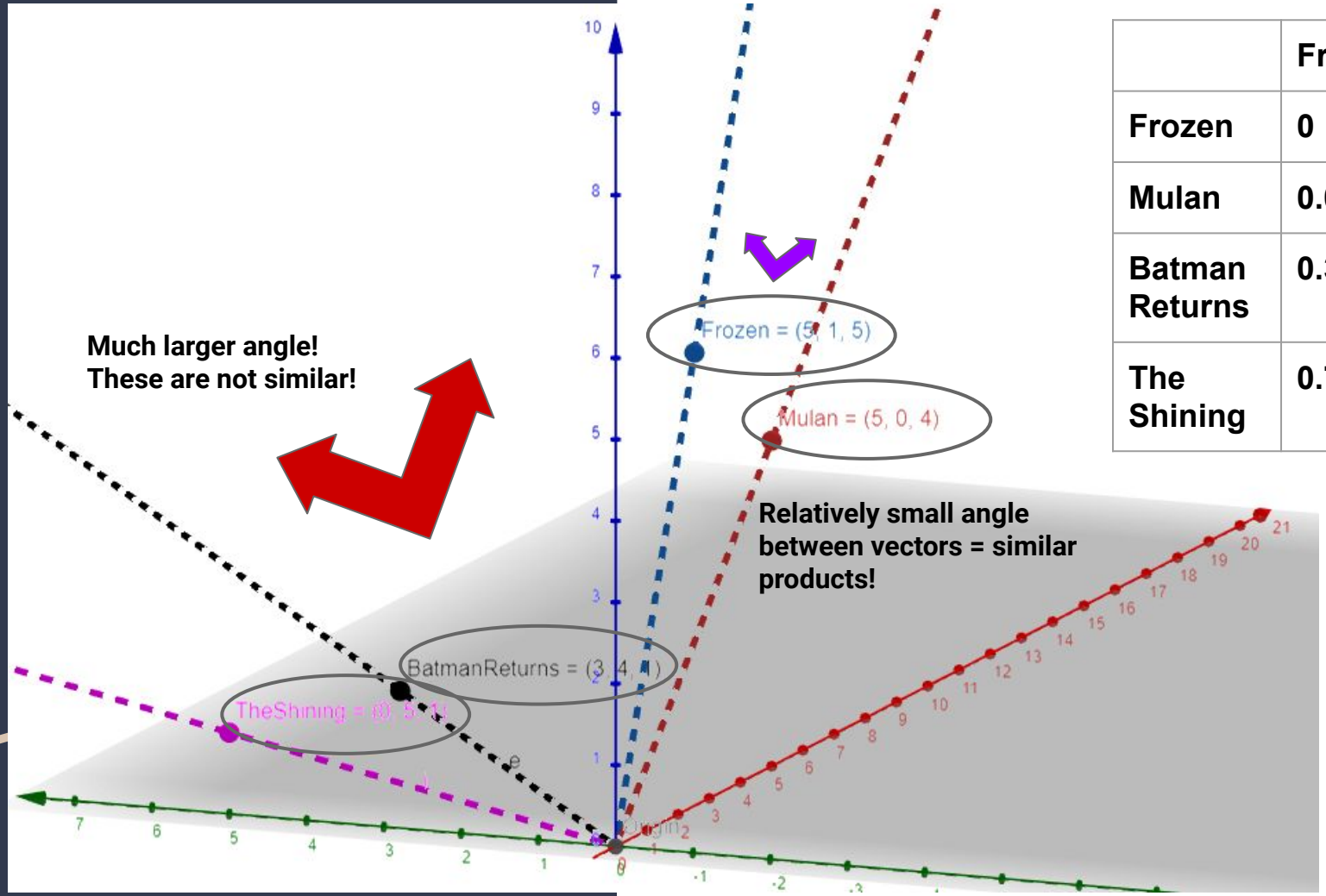
Movie	user_1	user_2	user_3
Batman Returns	3	4	1
Frozen	5	1	5
Mulan	5	0	4
The Shining	0	5	1

Movie vectors:

- Batman Returns [3, 4, 1]
- Frozen [5, 1, 5]
- Mulan [5, 0, 4]
- The Shining [0, 5, 1]

Each product in my recommender has a vector just like this… but longer!

Unfortunately, anything beyond 3-dimensions is extremely difficult to visualize!

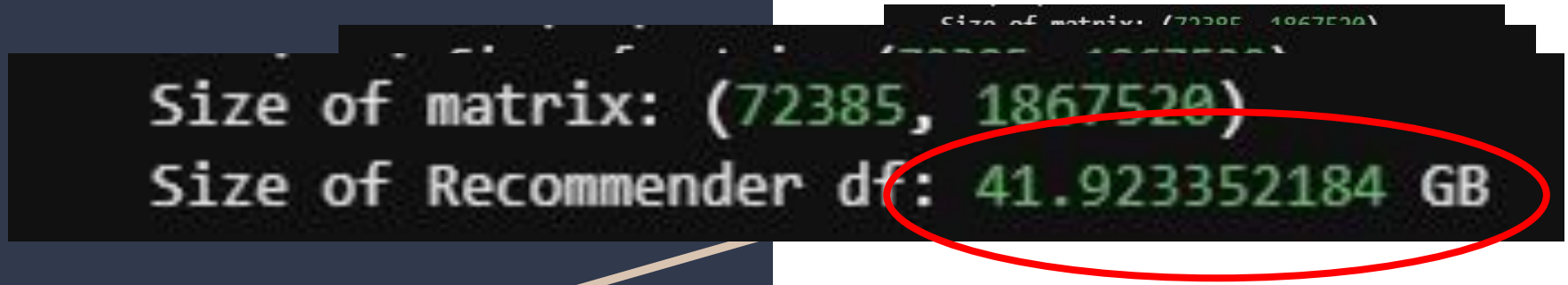


	Frozen
Frozen	0
Mulan	0.02
Batman Returns	0.34
The Shining	0.73

Building the system

Recurring Challenge - Data Compression!

- Over 72,000 movies, 1.8 million customers
  - The recommender is a dataframe of products by products, meaning 72,000 x 72,000 or...
  - 5,184,000,000 cells!




TOO MUCH
DATA
WOLO
NEW YORK
COMEDY CENTRAL

Building the system

Recurring Challenge - Data Compression!

- Over 72,000 movies, 1.8 million customers
  - The recommender is a dataframe of products by products, meaning 72,000 x 72,000 or...

```
Size of matrix: (72385, 1867520)
Size of Recommender df: 41.923352184 GB
```

- The Solution
  - A Sparse DataFrame!
    - Store all 1's as sparse
  - New size:
    - 1.8 GB !
  - Repeat for all recommenders!

Web Deployment

- Streamlit.io
  - Written exclusively in Python
  - Further compression!
    - Stored top 50 recommended items for each product in a dictionary
      - Takes less memory (the largest recommender, formerly 1.8 GB, now 22 MB !)
      - Produces results MUCH faster
  - Added Features
    - Total Number of Reviews
    - Average Star Rating
    - BONUS: NLP
      - Top 5 most frequently occurring words in reviews for that item
  - Demonstration...

Next steps:

- Expand search parameters - add keyword searches over review text and / or genre search, author, etc.
- Add more products!
- Deeper NLP
  - Refine list of stop words for each category
  - Compare review text by star rating, product, average star rating, or by grouped (similar) products
- Add more data!
  - Products that have come out since 2015
  - Reviews/Ratings that have been created since 2015 (for new or old products)

Thank You!

https://share.streamlit.io/griffinwt/nextale/main/Nextale_Streamlit.py