Building a Movie Recommender

Amelia Dobronyi

Overview

- Business problem
- Data and model
- Recommendations and areas for further analysis
- Conclusion

Business Problem

- Recommendation systems for media content are utilized by streamers to help users make decisions
- Streamers effectively collect their own data given user decisions every day

Data

- 100k IMDB dataset - movie title and genres for ~10k movies
- ~100k user ratings from a total of 610 users

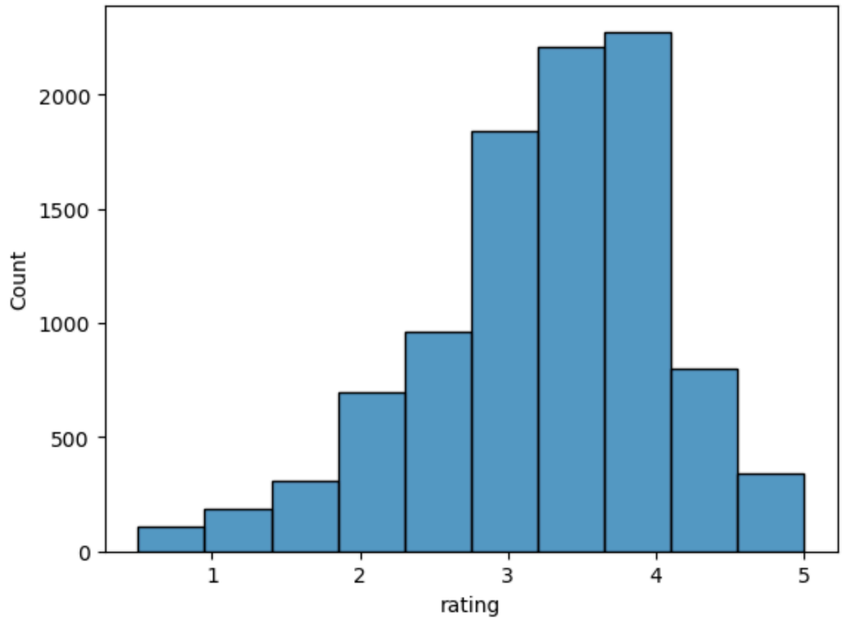

Fig 1. Average movie ratings lie between 3 and 4

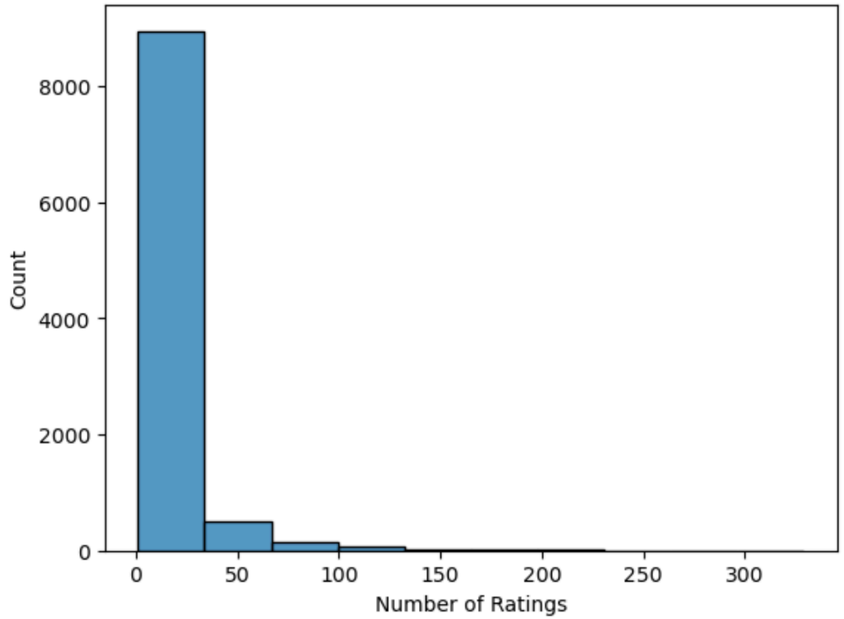

Fig 2. Most movies have a relatively small number of ratings

Recommender Approaches

- *Collaborative filtering*: recommendations solely based on user ratings
- *Content filtering*: recommendations based on movie content
- *Hybrid approach* uses both
- For this project:
  - Use collaborative filtering to produce the ratings
  - Use movie genres to further filter specific recommendations

Final Model

- Collaborative filtering using SVD model tuned on data
- No improvement from incorporating genre preferences
- Cold start problem (no existing user): give most highly rated films based on genre preferences
- Performance metrics:
  - RMSE 0.55
  - MAE 0.42

Final Model

Fig 3. Recommender output for a new user inputting genres Fantasy, Comedy, Action

```
1. 5634    Monday (2000)
Name: title, dtype: object
2. 8597    Dragonheart 2: A New Beginning (2000)
Name: title, dtype: object
3. 9394    Maximum Ride (2016)
Name: title, dtype: object
4. 899    Princess Bride, The (1987)
Name: title, dtype: object
5. 960    Evil Dead II (Dead by Dawn) (1987)
Name: title, dtype: object
```

Fig 4. Recommender output for an existing user inputting genres Fantasy, Comedy, Action

```
1. 899    Princess Bride, The (1987)
Name: title, dtype: object
2. 916    Army of Darkness (1993)
Name: title, dtype: object
3. 2798    Big Trouble in Little China (1986)
Name: title, dtype: object
4. 8881    Kung Fury (2015)
Name: title, dtype: object
5. 8357    The Lego Movie (2014)
Name: title, dtype: object
```

Future Improvements

- Incorporate hybrid content more broadly - lots of data for movies
- Try using similarities between movies rather than users - data size issue

Contact

- Email: adobronyi@mac.com
- GitHub: @adobronyi
- LinkedIn: https://www.linkedin.com/in/amelia-dobronyi-cfa-3a12b96b/