Collaborative Filtering for Movie Recommendation

Comparing Classical and Neural Algorithms

Aditya Srivastava
Harsh Gupta
Akimun Jannat Alvina
Niharika Narasimhiah Govinda

Given a large number of movies, how do you make individually targeted recommendations for users to watch?

Motivation

- Reduce decision making complexity with increasing options
- Drive user engagement and improve customer satisfaction
- Provide an edge over competitors

Movielens-100k Dataset

- Some numbers;
  - 100,000 movie ratings
  - 943 unique users
  - 1682 unique movies
- Extensively peer-reviewed
- Popular recommendation benchmark

	user	item	score	timestamp	age	gender	occupation	zip code	movie title	release date	video release date	IMDb URL	unknown
0	195	241	3	881250949	49	M	writer	55105	Kolya (1996)	24-Jan-1997	NaN	http://us.imdb.com/M/title-exact?Kolya%20(1996)	0
1	185	301	3	891717742	39	F	executive	00000	L.A. Confidential (1997)	01-Jan-1997	NaN	http://us.imdb.com/M/title-exact?L%2EA%2E+Conf...	0
2	21	376	1	878887116	25	M	writer	40206	Heavyweights (1994)	01-Jan-1994	NaN	http://us.imdb.com/M/title-exact?Heavyweights%...	0
3	243	50	2	880606923	28	M	technician	80525	Legends of the Fall (1994)	01-Jan-1994	NaN	http://us.imdb.com/M/title-exact?Legends%20of%...	0
4	165	345	1	886397596	47	M	educator	55113	Jackie Brown (1997)	01-Jan-1997	NaN	http://us.imdb.com/M/title-exact?imdb-title-11...	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...
99995	879	475	3	880175444	13	M	student	83702	First Wives Club, The (1996)	14-Sep-1996	NaN	http://us.imdb.com/M/title-exact?First%20Wives...	0
99996	715	203	5	879795543	36	F	administrator	44265	Back to the Future (1985)	01-Jan-1985	NaN	http://us.imdb.com/M/title-exact?Back%20to%20t...	0
99997	275	1089	1	874795795	21	M	student	95064	Sliver (1993)	01-Jan-1993	NaN	http://us.imdb.com/M/title-exact?Sliver%20(1993)	0
99998	12	224	2	882399156	47	M	educator	29206	101 Dalmatians (1996)	27-Nov-1996	NaN	http://us.imdb.com/M/title-exact?101%20Dalmati...	0
99999	11	202	3	879959583	28	F	other	06405	Unforgiven (1992)	01-Jan-1992	NaN	http://us.imdb.com/M/title-exact?Unforgiven%20...	0

100000 rows × 13 columns

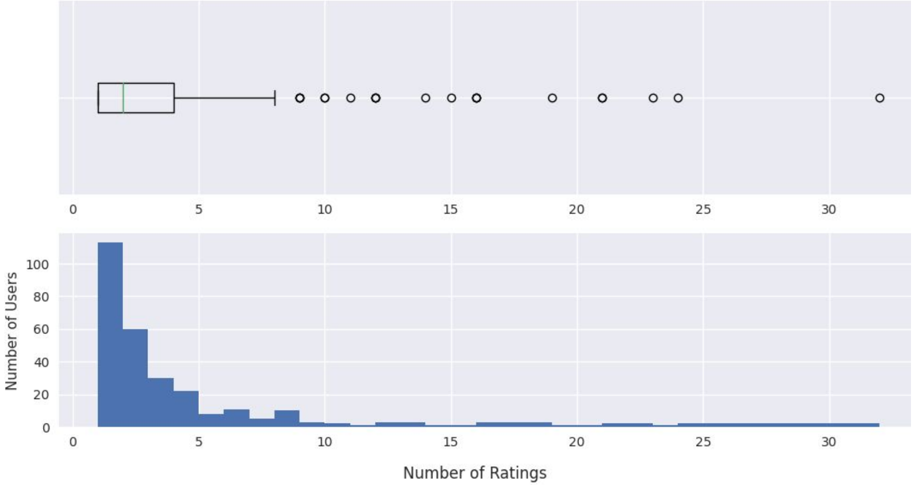
	Action	Adventure	Animation	Children's	Comedy	Crime	Documentary	Drama	Fantasy	Film-Noir	Horror	Musical	Mystery	Romance	Sci-Fi	Thriller	War	Western
0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	1	0	0	0	1	0	0	1	0	0	1	0	0
2	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	1	1
4	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
99995	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
99996	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0
99997	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
99998	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
99999	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

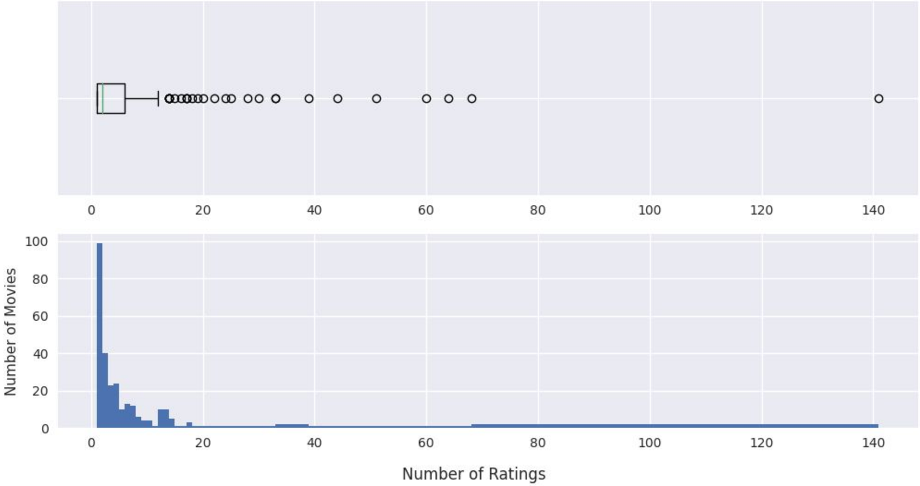
100000 rows × 18 columns

Data Preprocessing

- Converted explicit ratings (1 to 5) to implicit (1 if rated, 0 if unrated)
- Min-Max scaled numerical features to [0, 1]
- One-Hot encoded categorical features
- Three types of model inputs generated
  i. Training:
     - ***(user_id, item_id, rating)***
     - ***(user_id, pos_item_id, neg_item_id)***
  ii. Testing/Evaluation:
     - ***(user_id, pos_item_id, [neg_item_id*** $_1$***, neg_item_id*** $_2$***, … neg_item_id*** $_{100}$***])***
- Negative features sampled randomly

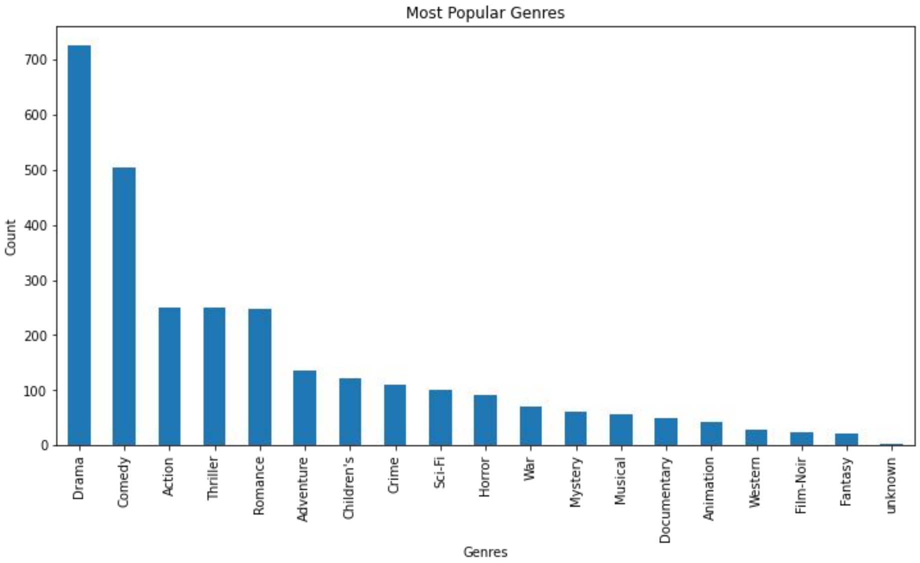
Data Analyses

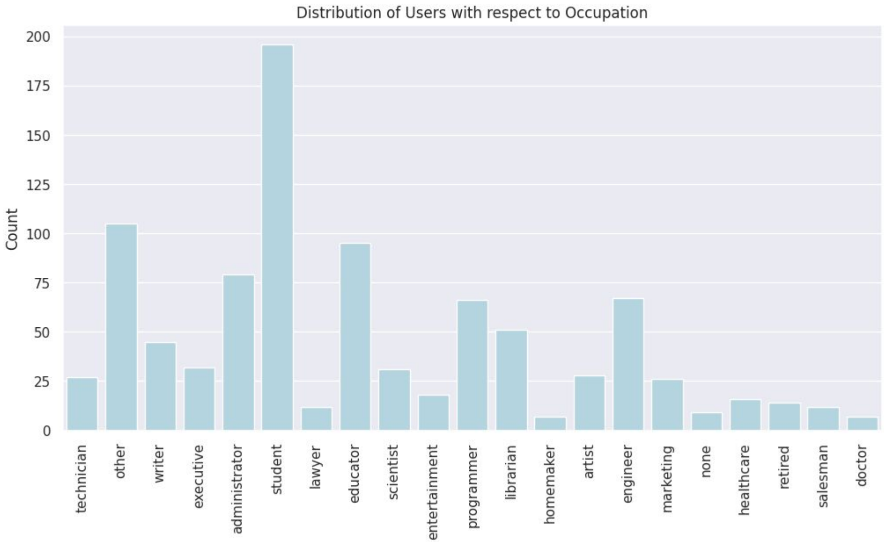




- Majority of users rate only a few movies
- Indicates a significant subset of highly active users
- Two distinct groups: casual raters and avid movie watchers
- Avid users contribute extensive ratings for hundreds of films

- Prevalence of movies with low ratings
- Dataset may feature numerous niche films
- Users tend to rate movies they are familiar with, overlooking lesser-known films





- Drama emerges as the most popular genre
- Followed by Comedy and Action in terms of frequency

- Students exhibit the highest inclination for rating movies
- Students consume more media, potentially leading to increased movie ratings
- Homemakers and doctors are less common in the dataset, as highlighted by the figure

Correlation analysis:

- **Genre and Gender Correlation**

  - Women show a preference for Romance movies, while men lean towards Sci-Fi

- **Inter-Genre Relationships**

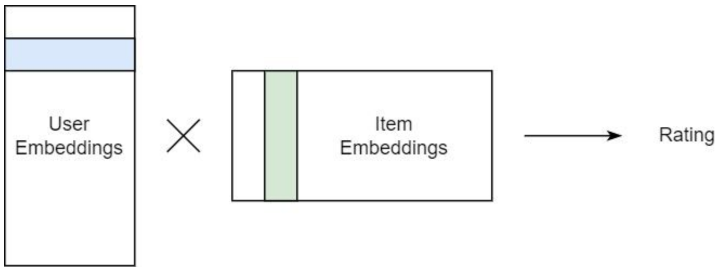
  - Certain genres frequently co-occur, e.g., Animation with Children's and Musical, and Mystery with Thriller

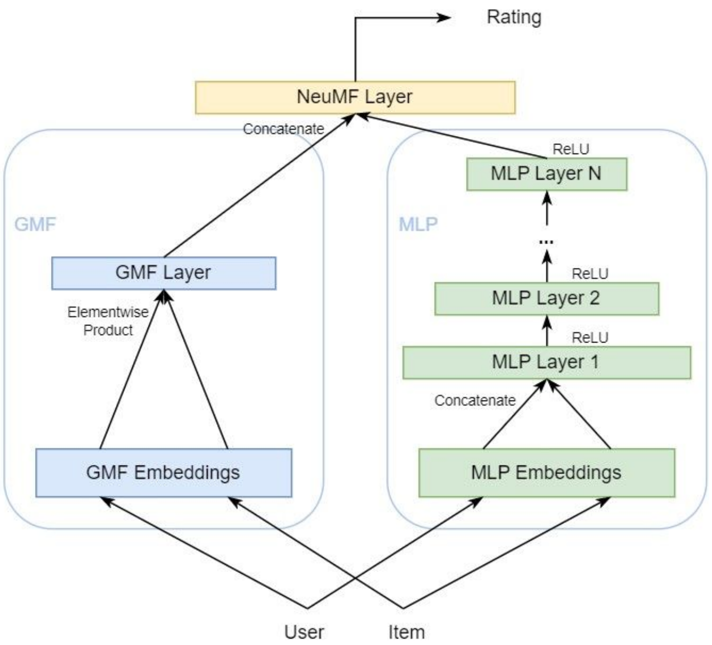
- **Gender and Occupation Bias**

  - Strong correlations suggest potential biases, with male users associated more with technical professions and female users linked to healthcare and homemaking occupations

Method

Models





Matrix Factorization (MF) **Neural Collaborative Filtering (NCF)**

Objective Functions

- Three different loss functions:
  1. **Root Mean Squared Error (RMSE)**
     - *Minimize difference between predicted rating and actual rating*
  2. **Binary Cross-Entropy (BCE)**
     - *Minimize entropy of predicted distribution vs. actual distribution*
  3. **Bayesian Personalized Ranking (BPR)**
     - *Maximize the difference in predicted ratings of positive items and negative items*

Evaluation Metrics

- Two different evaluation metrics:

  1. **Hit Ratio (HR@K)**
     - *Probability that a positive item’s predicted rating falls within the top K ratings in a set where every other item is negative*
  2. **Normalized Discounted Cumulative Gain (NDCG@K)**
     - *Same as HR however, there is higher penalization for ranking lower among the top K ratings*

Experimental Setting

- Experiments with combinations of models and objectives
- Model hyperparameters were adopted as given in source material
- Optimizers used:
  - Stochastic Gradient Descent (SGD) for RMSE and BPR
  - Adam for BCE
- Optimizer hyperparameters were grid searched
- NCF's GMF and MLP submodules were pretrained before NCF finetuning

Results

	HR			NDCG		
	@1	@5	@10	@1	@5	@10
MF-RMSE	0.31	0.65	0.80	0.31	0.49	0.54
NCF-BCE	0.23	0.62	0.77	0.30	0.47	0.52
MF-BPR	0.34	**0.73**	**0.87**	0.34	**0.55**	**0.59**
NCF-BPR	**0.38**	0.38	0.38	**0.38**	0.38	0.38

Comparing loss functions:

- RMSE and BCE models shows similar performance
- Both RMSE and BCE models are outperformed by BPR

Comparing models:

- Classical MF yields better results than NCF for almost every metric and with every loss function
- NCF-BPR may have better precision

Overall best: Matrix factorization model with BPR loss

Results

	HR			NDCG		
	@1	@5	@10	@1	@5	@10
MF-RMSE	0.31	0.65	0.80	0.31	0.49	0.54
NCF-BCE	0.23	0.62	0.77	0.30	0.47	0.52
MF-BPR	0.34	**0.73**	**0.87**	0.34	**0.55**	**0.59**
NCF-BPR	**0.38**	0.38	0.38	**0.38**	0.38	0.38

Conclusions

- Surprisingly, **classical algorithm > neural counterpart**
- Choice of loss function plays a crucial role in model effectiveness
- Indicative of traditional ML techniques still being relevant and competitive
- Factors that may have influenced performance
  - Dataset size disparity
  - Skipped content-based features
  - Model complexity and optimization
    - ***Neural models are more sensitive to tuning on smaller datasets***

Challenges

- Dealing with data bias
- Complex data preprocessing
- Cold-start problem
- Discrepancies between reference papers
- Missing reference details in reference papers
- Lack of reference code

Future Work

- **Dataset expansion:**
  - *Test models on the MovieLens 1M dataset for broader evaluation*
- **Comprehensive grid search:**
  - *Extend grid search to optimize model parameters, not just optimizers*
- **Content-based feature integration:**
  - *Explore leveraging content-based features in the dataset*
- **Content-based and hybrid model exploration:**
  - *Investigate content-based and hybrid models for more comprehensive analyses*

Thank you!