Movie Recommendation System : A Hybrid Approach

Group - 20

Under the co-guidance of

Dr. Vinay Joseph

(Assistant Professor, ECED)

Dr. Nujoom Sageer Karat

(Faculty Member, ECED)

Movie Recommendation Systems

- Recommender systems are an essential innovation by which content is curated by intelligently predicting user's interests and preferences.
- Recommender systems give us a plethora of relevant options which aid us in selecting the content to be consumed.
- A Movie Recommendation System can suggest a set of movies to users based on their preferences or the popularities of the movies.
- This saves time for making decisions while consuming content for entertainment, especially in today's digital age.

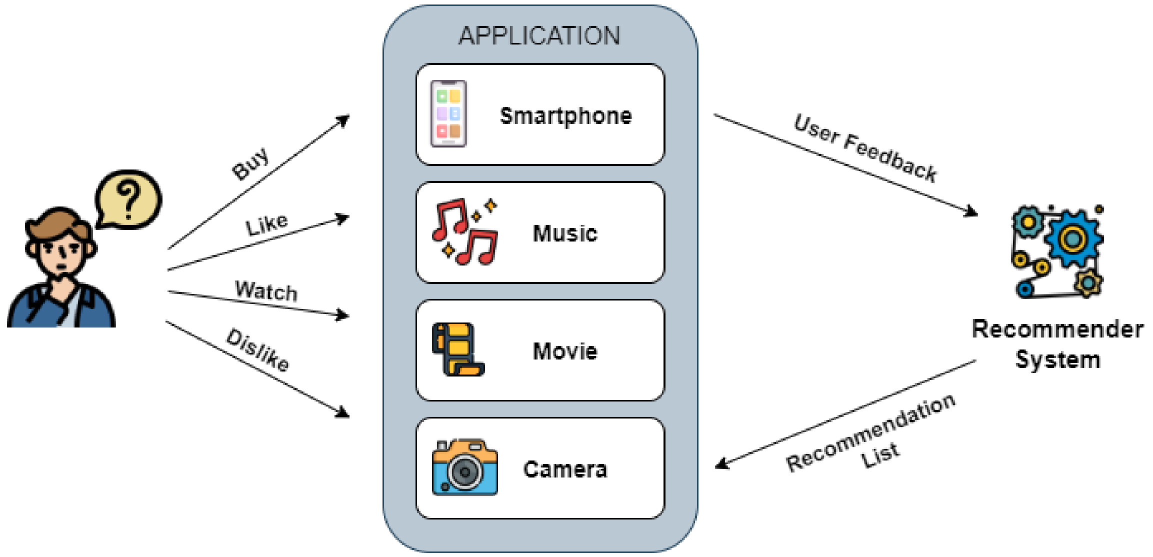


Figure-1: Application Scenario

Types of Recommender Systems

- ❖ There are broadly three classifications of recommendation systems in wide usage
    1. Content-based filtering
    2. Collaborative filtering
    3. Hybrid methods
- ❖ Hybrid approach is the integration of two or more recommendation algorithms like content-based filtering, collaborative filtering etc.

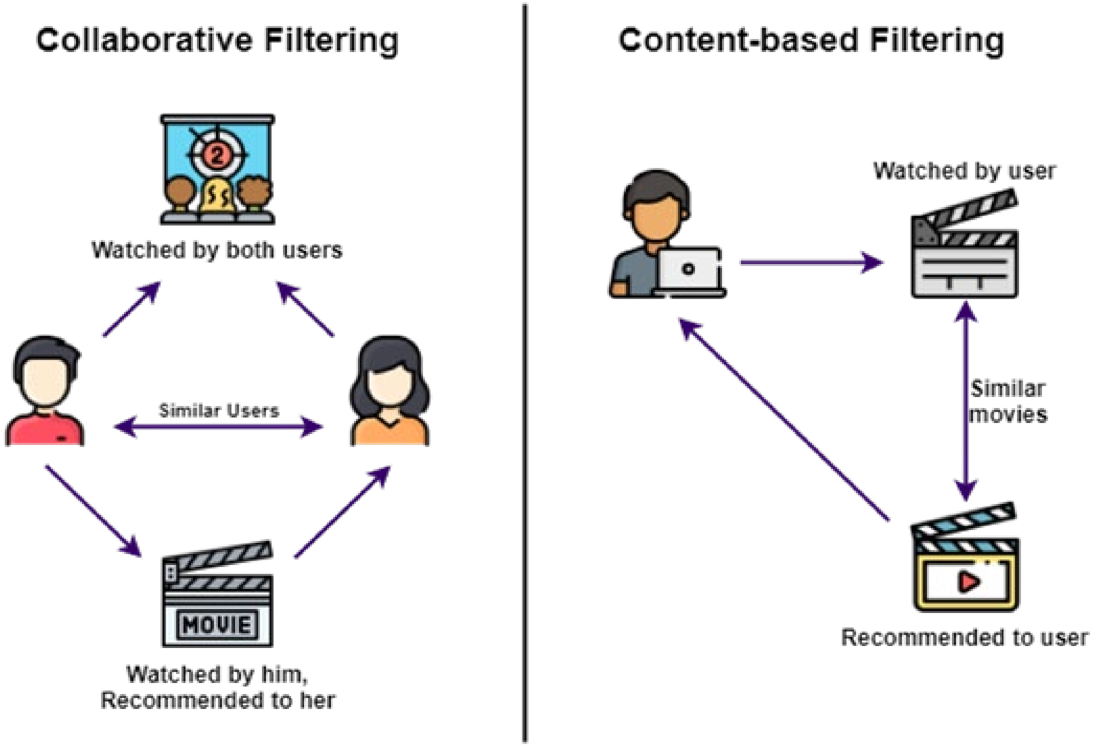


Figure-2: Types of Recommender Systems

Problem Statement

- To design and assess a novel movie recommendation system that utilizes a hybrid approach.
- To develop the hybrid approach using Matrix Factorization and BERT4Rec.
- To investigate the performance of our hybrid model using appropriate metrics.
- To tackle the disadvantages like lack of information about the domain dependencies in collaborative filtering and people’s preferences in content based systems, with a hybrid approach.

Our Approach

- Pre-processing the dataset to meet our specifications.
- Design and assess the collaborative filter recommendation system utilizing Matrix Factorization.
- Design and assess the content-based filter recommendation system using BERT4Rec.
- Design of hybrid approach.
- Evaluating the performance of our model.

DATASET : ML-1m dataset which includes 1,000,209 anonymous ratings of approximately 3,900 movies made by 6,040 Movie-Lens users was employed.

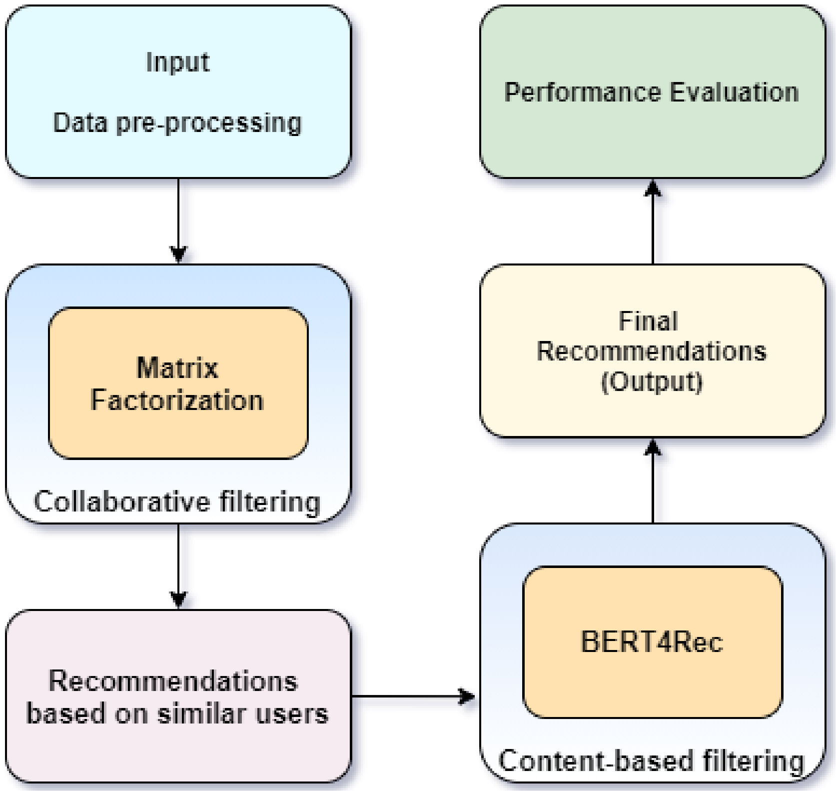


Figure-3: Abstract architecture of hybrid model

Collaborative Filtering

❖ Collaborative Filtering recommender systems create predictions based on user-item relationships.

❖ The user-item rating matrix is used in latent factor models for movie recommendations to profile both users and items using a number of latent variables.

❖ A feature in the context of movies can be a genre, target age range, or even something utterly unintelligible.

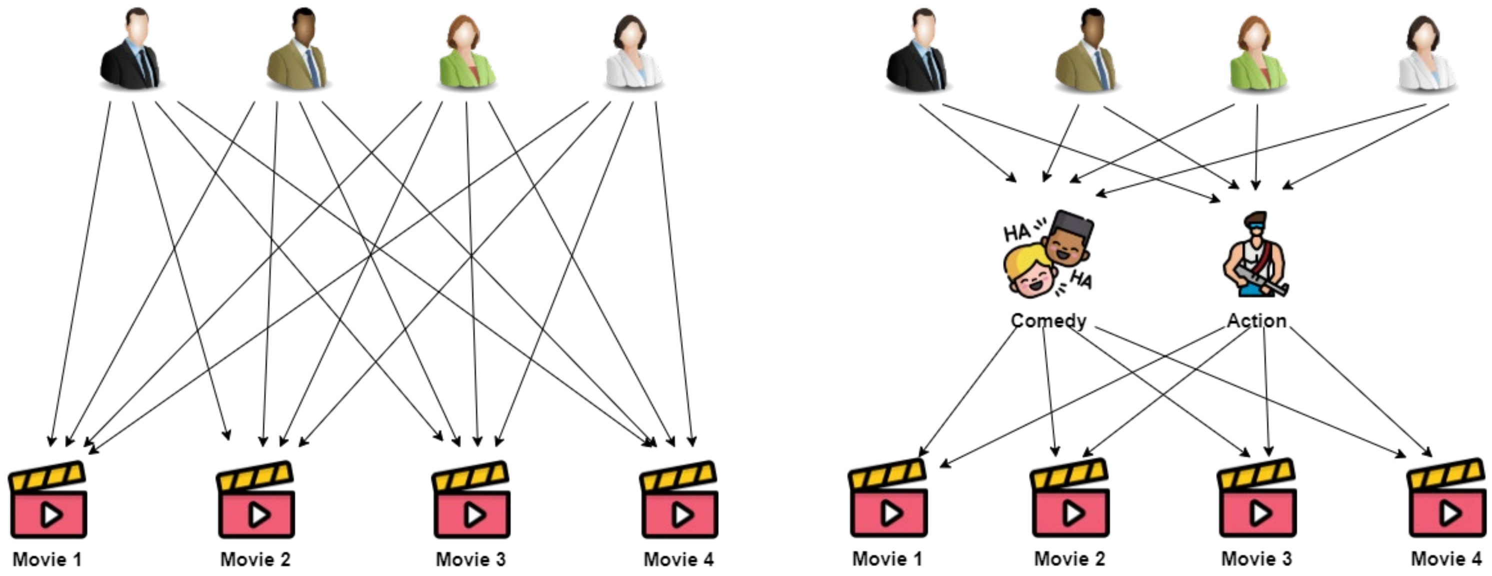
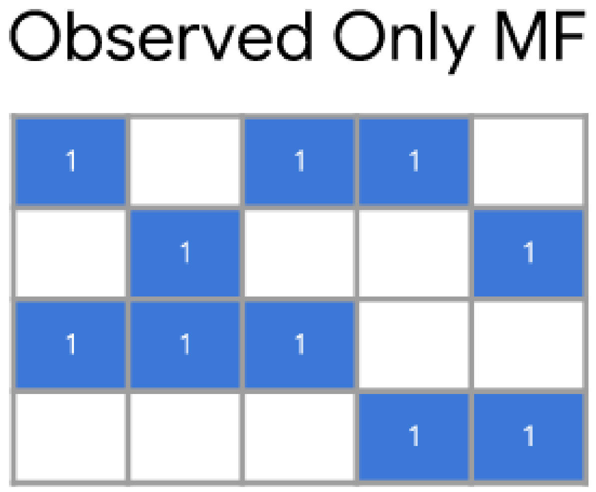


Figure-4: Collaborative Filtering

Matrix Factorization

- Matrix Factorization finds two rectangular matrices with smaller dimensions to represent a big user-item relationship matrix.
- The feedback (or rating) matrix is denoted as $A \in R^{m \times n}$, where ’m’ is the number of users (or queries) and ‘n’ is the number of items. ‘d’ denotes the number of latent factors considered.
- The model learns:
  - A user embedding matrix $U \in R^{d \times m}$, where row i is the embedding for user i
  - An item embedding matrix $V \in R^{n \times d}$, where row j is the embedding for item j
- To handle sparsity, only sum over observed pairs (i, j), that is, over non-zero values in the feedback matrix.

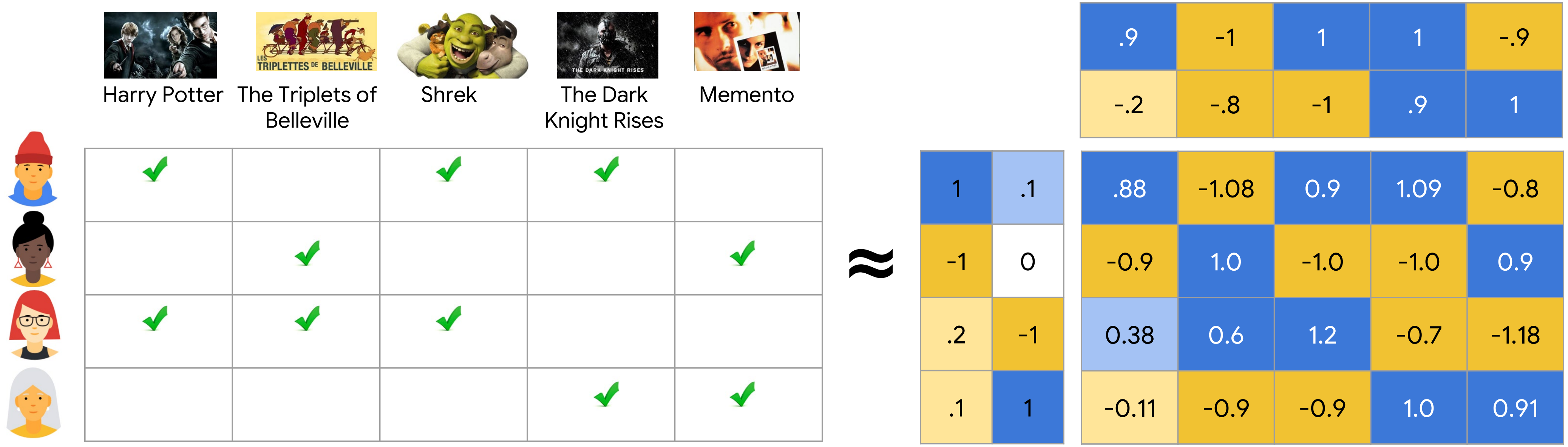


$$\Sigma_{(i,\, j) \in obs} (A_{ij} - U_i \,.\, V_j)^2$$

Figure-5: Observed Only MF

Matrix Factorization

- The embeddings are learned such that the product UV^T is a good approximation of the feedback matrix A.
- Furthermore, the embeddings can be learned automatically, without relying on hand-engineering of features.
- We recommend by performing dot product of the factor matrices to fill in the missing entries in the rating matrix.



.9	-1	1	1	-.9
-.2	-.8	-1	.9	1

1	.1
-1	0
.2	-1
.1	1

.88	-1.08	0.9	1.09	-0.8
-0.9	1.0	-1.0	-1.0	0.9
0.38	0.6	1.2	-0.7	-1.18
-0.11	-0.9	-0.9	1.0	0.91

Figure-6: Matrix Factorization

Evaluation of MF model

Mean Squared Error

❖ MSE is a statistical metric that represents the standard deviation between a set of estimated values to the actual values.

❖ In recommender systems, it has been used to measure how far a set of predictions are, from the true values.

$$L = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - \widehat{y}_i\right)^2$$

where,

N = number of data points

y_i = Observed Values

$\hat{y}_i$ = Predicted values

Cosine Similarity – To verify the recommendations for user profile

❖ Cosine Similarity is a measurement that quantifies the similarity between two or more vectors.

$$\text{cosine similarity} = S_C(A,B) := \cos(\theta) = \frac{\mathbf{A}\cdot\mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}},$$

Where,

A – Weight vector of recent highest rated movie by a user

B – Weight vector of the movie recommended to a user

Loss Plots

- The MF model without bias has given a validation loss value 0.827 after training 100 epochs.
- MF model without bias has not given satisfactory results with validation data, which shows a diverging trend for generalization.
- The MF model with bias has given a validation loss of 0.756 after training 100 epochs.
- This model with bias has shown a good performance in training data and good generalization to validation data.

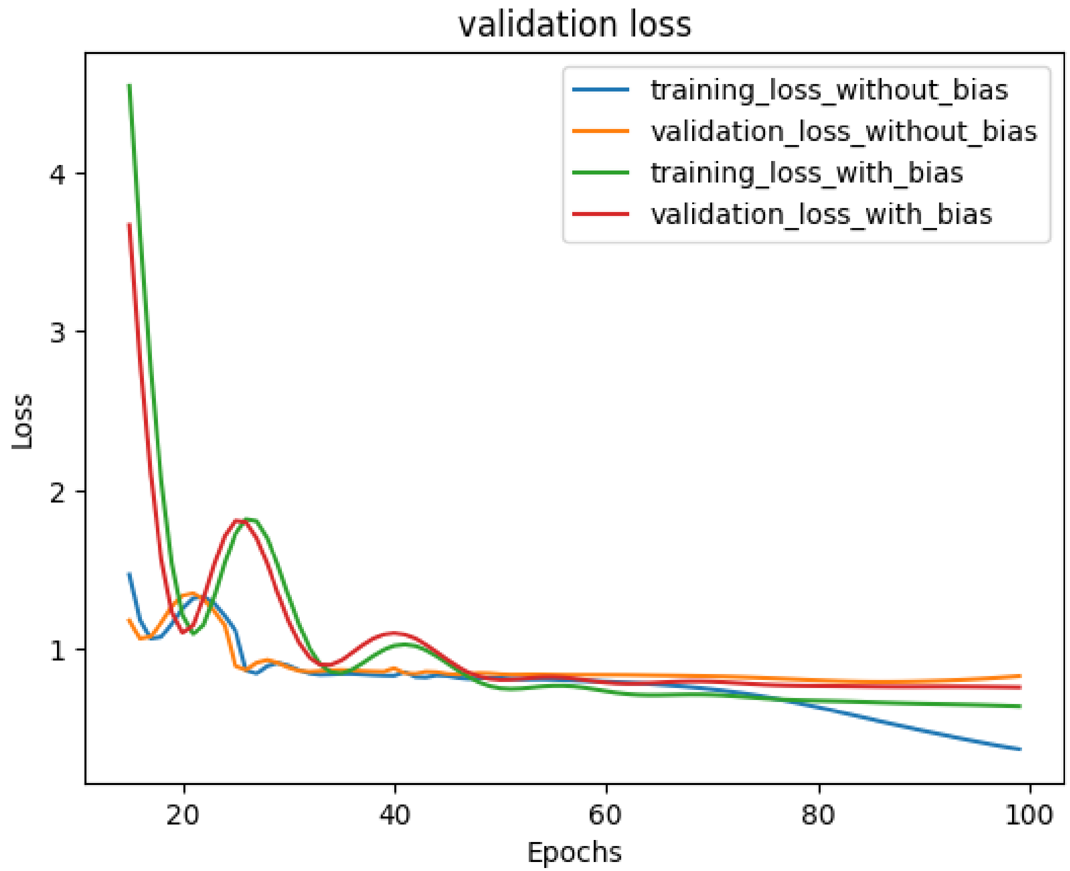


Figure-7: Training and Validation Losses vs Epochs

MF model Results

- ❖ We have tested our model with both 15% and 20% of ML-1m dataset as test data.
- ❖ The movie name suggestions given by our model for a particular user are verified using Cosine Similarity, by comparing the predictions with most recent highest rated movie of that user.
- ❖ The cosine similarity score indicates that the recommendations are relatable to the user.
- ❖ We have observed that there was no significant change in the MSE value, after addition of new ratings to the test data.

Using 15% of ML-1m dataset for testing

	userId	movieId	rating	timestamp	title
499942	270	374	5	972669952	Mr. Smith Goes to Washington (1939)
209370	270	634	5	972669914	Monty Python and the Holy Grail (1974)
80454	270	578	5	972669864	Mister Roberts (1955)
735445	270	427	5	972669786	Midnight Run (1988)
859527	270	151	5	972669786	Midnight Cowboy (1969)
...	...	...	...	...	...
925695	270	2498	2	968624228	Braddock: Missing in Action III (1988)
775416	270	1497	2	968624162	Bonfire of the Vanities (1990)
853193	270	3031	2	968623891	Beyond the Poseidon Adventure (1979)
540273	270	1452	2	968623809	Beneath the Planet of the Apes (1970)
659116	270	259	2	968623519	Avengers, The (1998)

379 rows × 5 columns

```
Most recent high rated movie by the user : Mr. Smith Goes to Washington (1939) 5
```

Fig-8: User watch history

```
Godfather, The (1972) --> 0.84962887
Raiders of the Lost Ark (1981) --> 0.85307413
Casablanca (1942) --> 0.92780215
Rear Window (1954) --> 0.917664
Citizen Kane (1941) --> 0.86343676
Sanjuro (1962) --> 0.8830031
Wrong Trousers, The (1993) --> 0.9251889
One Flew Over the Cuckoo's Nest (1975) --> 0.9056849
Star Wars: Episode IV - A New Hope (1977) --> 0.8322398
To Kill a Mockingbird (1962) --> 0.9567722
```

Fig-9: Recommendations given by our MF model

```
[50] print(mean_squared_error(model,test))

0.7568132776446861
```

Fig-10: Mean Squared Error value

MF model Results

Using 20% of ML-1m dataset for testing

	userId	movieId	rating	timestamp	title
499942	270	374	5	972669952	Mr. Smith Goes to Washington (1939)
209370	270	634	5	972669914	Monty Python and the Holy Grail (1974)
80454	270	578	5	972669864	Mister Roberts (1955)
735445	270	427	5	972669786	Midnight Run (1988)
859527	270	151	5	972669786	Midnight Cowboy (1969)
...	...	...	...	...	...
925695	270	2498	2	968624228	Braddock: Missing in Action III (1988)
775416	270	1497	2	968624162	Bonfire of the Vanities (1990)
853193	270	3031	2	968623891	Beyond the Poseidon Adventure (1979)
540273	270	1452	2	968623809	Beneath the Planet of the Apes (1970)
659116	270	259	2	968623519	Avengers, The (1998)

379 rows × 5 columns

```
Most recent high rated movie by the user : Mr. Smith Goes to Washington (1939) 5
```

Fig-11: User watch history

```
Schindler's List (1993) --> 0.9122081
Godfather, The (1972) --> 0.9106181
Usual Suspects, The (1995) --> 0.90106136
Shawshank Redemption, The (1994) --> 0.8918897
Sanjuro (1962) --> 0.9224855
Star Wars: Episode IV - A New Hope (1977) --> 0.88264054
Raiders of the Lost Ark (1981) --> 0.8885121
To Kill a Mockingbird (1962) --> 0.9591541
Rear Window (1954) --> 0.9610451
Paths of Glory (1957) --> 0.9378479
```

Fig-12: Recommendations given by our MF model

```
[51] print(mean_squared_error(model,test))

     0.7604294228337461
```

Fig-13: Mean Squared Error value

Content Based Filtering

- Content-Based Filtering is a Machine learning technique that uses similarities in features to make decisions.
- This technique is often used in recommender systems algorithms which are designed to advertise or recommend things to users based on knowledge accumulated about the user.
- BERT4Rec model is employed to achieve content-based filtering.

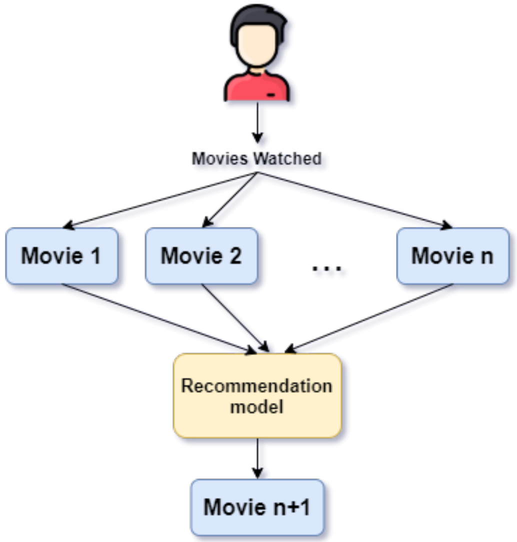


Figure-14: Content based model

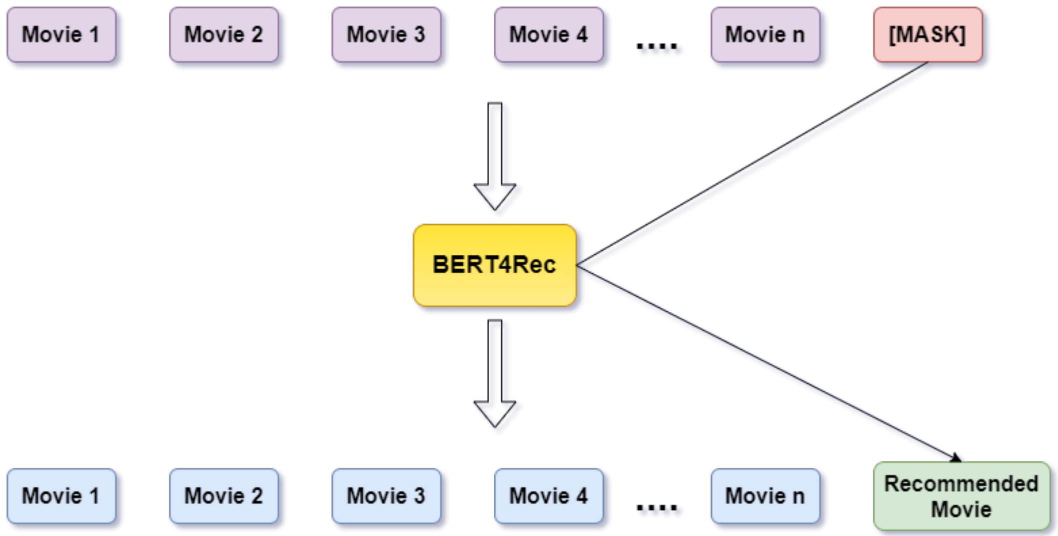


Figure-15: BERT4Rec abstract model

BERT Overview

❖ Bidirectional Encoder Representations from Transformers (BERT) is designed to pre-train deep bidirectional representations from unlabeled text by jointly conditioning on both left and right context.

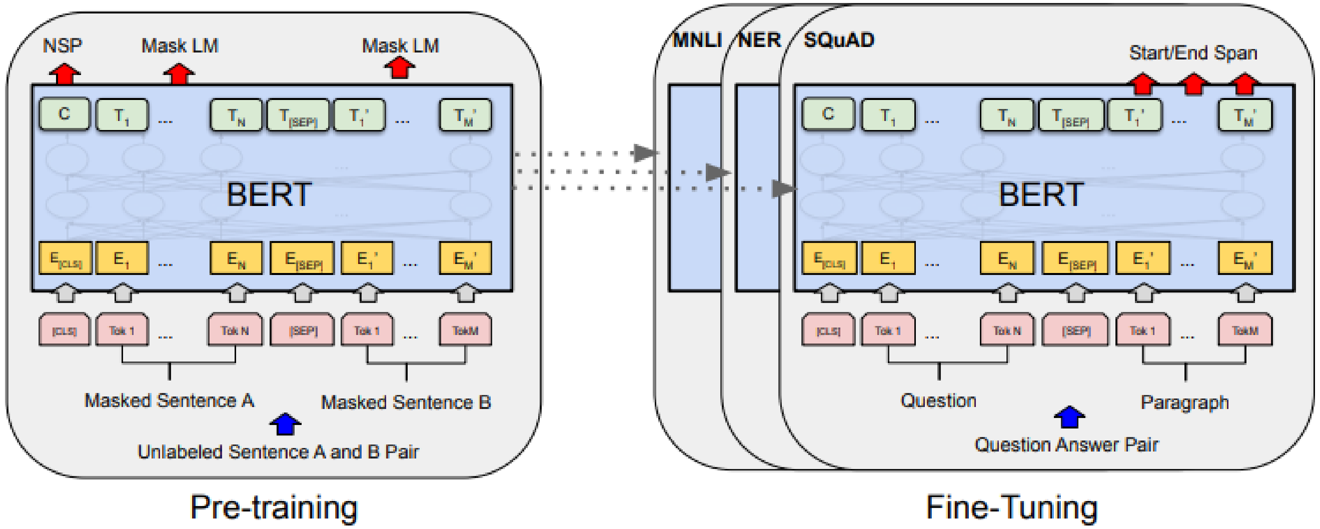


Figure-16: BERT Architecture

❖ BERT trained using two tasks :

- Mask Language Model (MLM) : predict the masked words.
- Next Sentence Prediction (NSP) : predict *IsNext or NotNext*

BERT vs BERT4Rec

BERT	BERT4Rec
• Commonly used in NLP tasks	• Used in Recommendation tasks
• Pretrained model that can be used for many NLP use cases.	• Model specific for a set of products in a platform
• Multitask learning (MLM and NSP)	• One task learning (MLM)
• Token represents word or sub-word.	• Token represents a product

BERT4Rec

- ❖ BERT4Rec uses bidirectional self-attention to simulate user's behavior sequences.
- ❖ We can easily capture item-item interactions across the entire user behavior sequence using self-attention mechanism.
- ❖ Transformer layer (Trm) is not aware of the order of the input sequence, without any recurrence or convolution module.
- ❖ In order to make use of the sequential information of the input, Positional Embeddings are added to input embeddings.

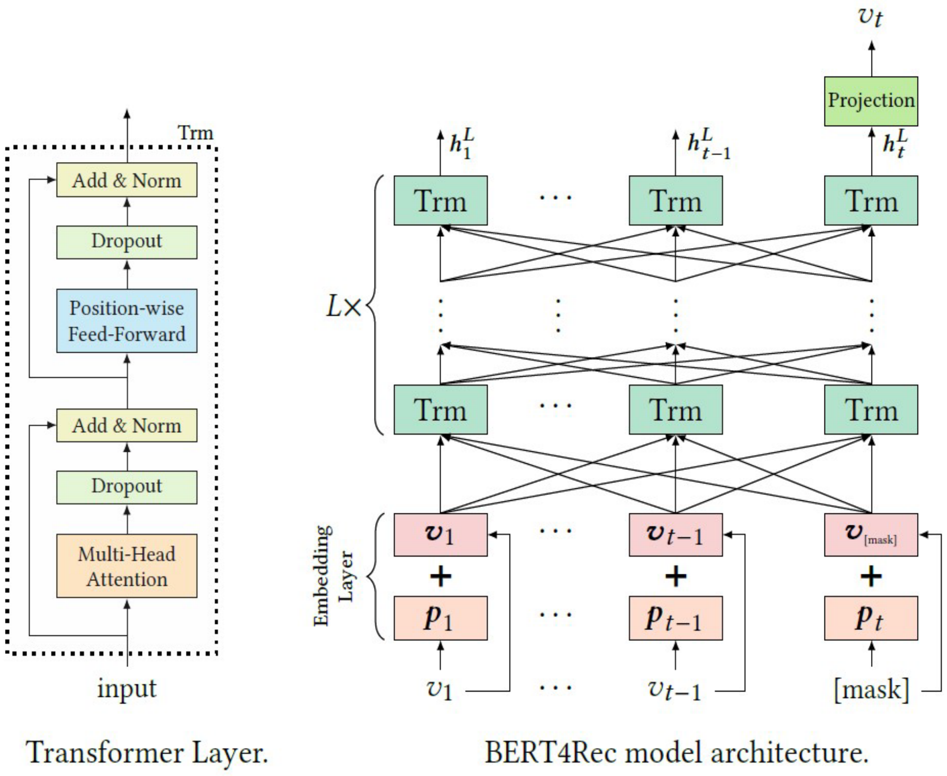


Figure-17: BERT4Rec model architecture

BERT4Rec Architecture

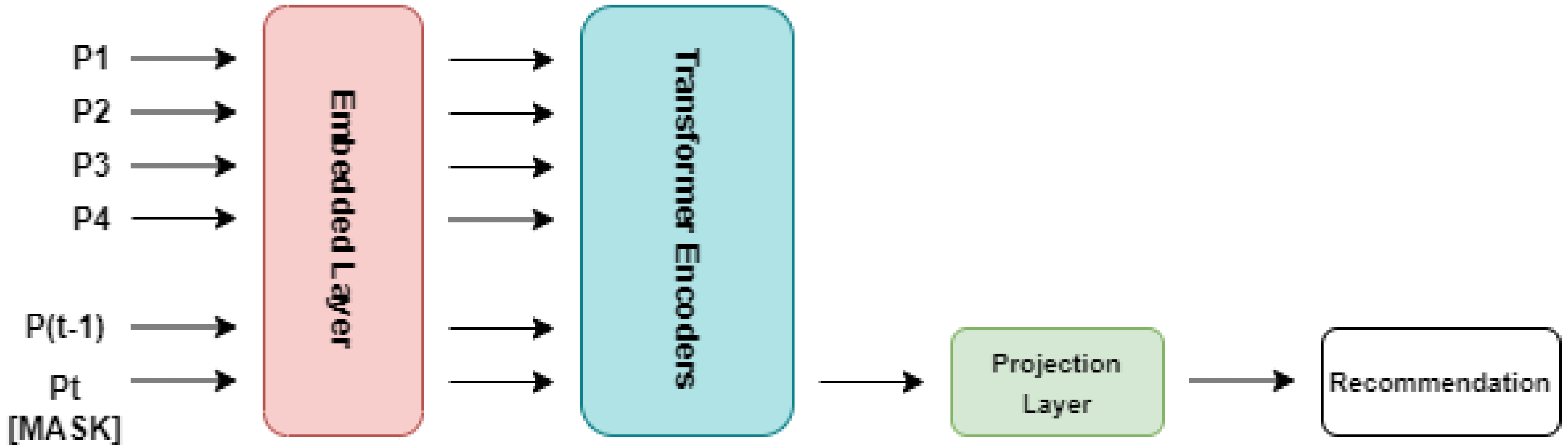


Figure-18: BERT4Rec block architecture

- Result of the embedding layer is the embedding vectors of the related products injected with learnable sinusoid positional embeddings.
- The encoder layer in the BERT4Rec model stacks multiple transformers layers to gather information from different position in the sequence.
- The projection layer is the final part of the model and it helps decide which product will be recommended.

Model Learning

Training

❖ For each training step, we randomly mask a proportion of all items in the input sequence, and then predict the original ids of the masked items based solely on its bidirectional context.

❖ Eventually, we define the loss for each masked input as the negative log-likelihood of the masked targets.

$$\mathcal{L} = \frac{1}{|\mathcal{S}_u^m|} \sum_{v_m \in \mathcal{S}_u^m} -\log P\left(v_m = v_m^* \mid \mathcal{S}_u'\right)$$

Testing

❖ Append the mask token to the end of user's behavior sequence, and then predict the next item based on the final hidden representation of this token.

Loss Plots

Fig-19: Epochs vs X-axis values

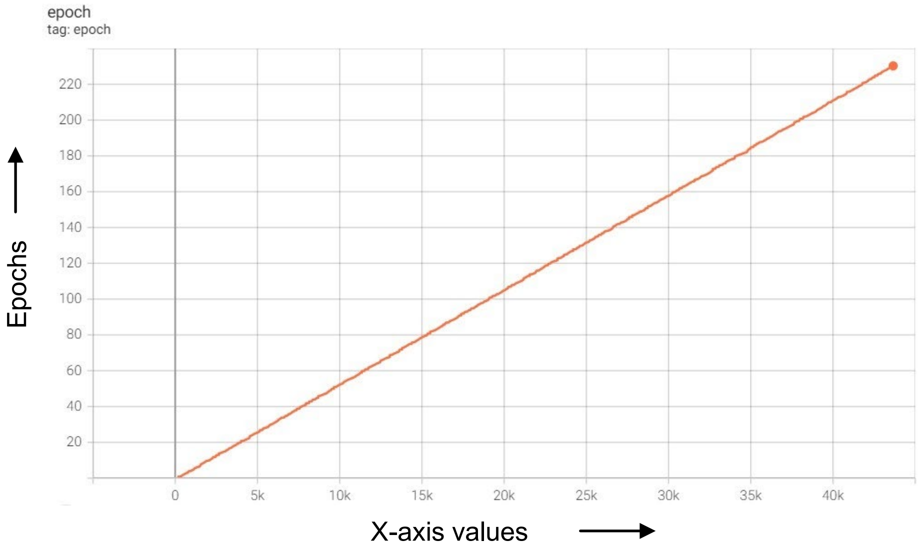
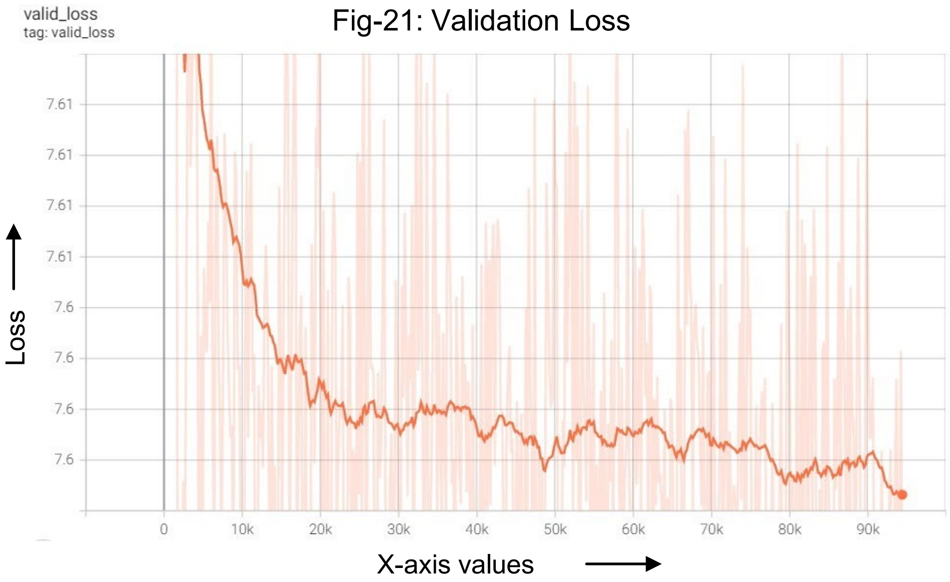


Fig-20: Training Loss

Fig-21: Validation Loss



Evaluation Metric

- Precision: It is the ratio of recommended items that are relevant to the total number of recommended items.
- Recall: It is the ratio of recommended items that are relevant to the total number of relevant items.

Mean Average Precision

- Average Precision is a measure of the number of relevant products recommended in the first “K” results.

$$AP = \sum_{1}^{x} (precision\ at\ i) \times (change\ in\ recall\ at\ i)$$

- AP is calculated using precision and recall at every position in the ranked sequence of recommendations.
- MAP is the mean of average precisions for all users.

$$MAP = \frac{\sum_{u=1}^{N} AP_u}{M}$$

Hyper-parameters

- We have used various values for total number of epochs, learning rate, and checked for better MAP value, based on trail and error method.
- We have varied learning rate and observed that LR = $1e^{-3}$ is giving good performance for less number of epochs.

Learning Rate / Epochs	1e-3	1e-4
500	0.277	0.247
1000	0.256	0.265

- Fixing the LR value at $1e^{-3}$, we have varied the number of epochs to check MAP score.

Epochs / Learning Rate	250	500	750	1000
1e-3	0.223	0.277	0.264	0.256

BERT4Rec model Results

Fig-22: BERT4Rec Input

```
['Mummy, The (1959)',
 "Mummy's Curse, The (1944)",
 'Message in a Bottle (1999)',
 'McCabe & Mrs. Miller (1971)',
 'Mary Poppins (1964)',
 'Man with the Golden Gun, The (1974)',
 'Man Who Would Be King, The (1975)',
 'Maltese Falcon, The (1941)',
 'Mad Max (1979)',
 'Lolita (1997)',
 'Lethal Weapon (1987)',
 'League of Their Own, A (1992)',
 'Goldfinger (1964)',
 'Ghostbusters (1984)',
 'Getaway, The (1994)',
 'Gaslight (1944)',
 'From Dusk Till Dawn (1996)',
 'Friday the 13th Part VII: The New Blood (1988)',
 'Friday the 13th Part V: A New Beginning (1985)',
 'French Connection, The (1971)',
 'Frankenstein (1931)',
 'Foxfire (1996)',
 'Foreign Correspondent (1940)',
 'For a Few Dollars More (1965)',
 'Force 10 from Navarone (1978)',
 'Fight Club (1999)',
 'Fighting Seabees, The (1944)',
 'Falcon and the Snowman, The (1984)',
 'Face/Off (1997)',
 'Exorcist II: The Heretic (1977)']
```

BERT4Rec

Fig-23: BERT4Rec Output

```
["One Flew Over the Cuckoo's Nest (1975)",
 'Aliens (1986)',
 'Alien (1979)',
 'Terminator, The (1984)',
 'Gladiator (2000)',
 'Stand by Me (1986)',
 'Wizard of Oz, The (1939)',
 'Shakespeare in Love (1998)',
 'Star Wars: Episode I - The Phantom Menace (1999)',
 'Total Recall (1990)',
 'Toy Story 2 (1999)',
 'Fugitive, The (1993)',
 'Men in Black (1997)',
 'Titanic (1997)',
 'Airplane! (1980)',
 'Babe (1995)',
 'Usual Suspects, The (1995)',
 'Galaxy Quest (1999)',
 'Sixth Sense, The (1999)',
 'Breakfast Club, The (1985)',
 'Blade Runner (1982)',
 'Election (1999)',
 'Groundhog Day (1993)',
 '2001: A Space Odyssey (1968)',
 'Casablanca (1942)',
 "Bug's Life, A (1998)",
 'Pulp Fiction (1994)',
 'Godfather: Part II, The (1974)',
 'Abyss, The (1989)',
 'Forrest Gump (1994)']
```

Fig-24: MAP Score for BERT4Rec

```
print("Mean Average Precision of BERT4Rec model = ", sum1/count1)
```

```
Mean Average Precision of BERT4Rec model =  0.2778572935105253
```

Hybrid model

❖ Pipe-lining both the collaborative filtering and content based filtering models in a cascade fashion to design a novel hybrid recommendation system.

Cascade System

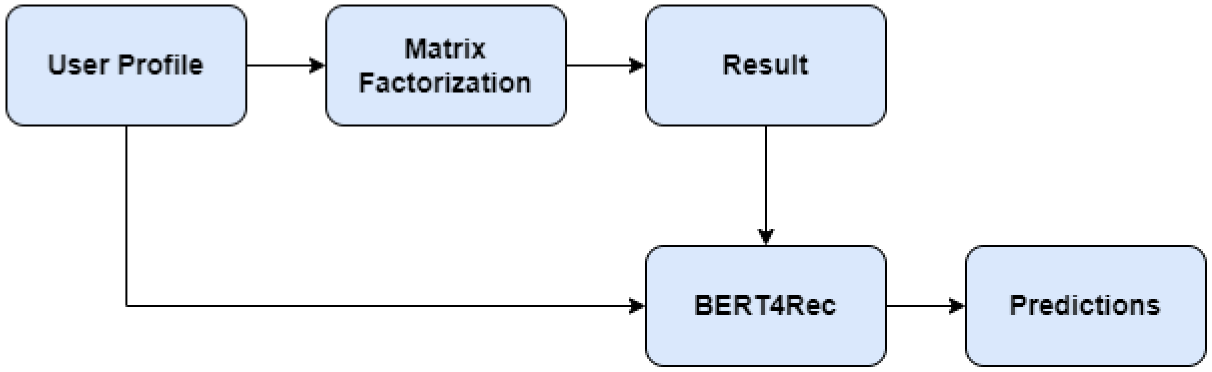


Fig-25: Cascade model

❖ The recommendations of one technique are refined by another recommendation technique.

❖ The outputs from MF model based on ratings are cascaded to the BERT4Rec model along with the user profile, which in turn produces recommendations specific to that user.

Hybrid model Results

- We have extended the same idea of "Mean Average Precision" evaluation implemented for BERT4Rec model, to the hybrid model and observed a significant increase in the MAP score.
- We have observed approximately 18% increase in the MAP score which indicates that the hybrid recommender system is enhancing the predictions for a user.

Fig-26: Input is test dataset (Example user ID =270)

	userId	movieId	rating	timestamp	title
573101	270	1639	5	972669546	Man Who Would Be King, The (1975)
296897	270	1276	5	972669521	Maltese Falcon, The (1941)
215948	270	627	5	972669474	Mad Max (1979)
507133	270	384	5	972669159	Lethal Weapon (1987)
221497	270	955	5	968626656	Goldfinger (1964)
...	...	...	...	...	...
987784	270	3238	3	968622946	Adventures of Elmo in Grouchland, The (1999)
819337	270	1615	3	968622902	Aces: Iron Eagle III (1992)
975735	270	2057	3	968622839	52 Pick-Up (1986)
967469	270	2748	2	968625819	Exorcist II: The Heretic (1977)
948243	270	404	2	968625819	Exorcist III, The (1990)

90 rows × 5 columns

Fig-27: Hybrid model output

```
['Braveheart (1995)',
 'Being John Malkovich (1999)',
 'Men in Black (1997)',
 'Fargo (1996)',
 'L.A. Confidential (1997)',
 'Princess Bride, The (1987)',
 'Shakespeare in Love (1998)',
 'Star Wars: Episode I - The Phantom Menace (1999)',
 'E.T. the Extra-Terrestrial (1982)',
 'Groundhog Day (1993)',
 'Alien (1979)',
 'Pulp Fiction (1994)',
 'Forrest Gump (1994)',
 'Terminator, The (1984)',
 'Fugitive, The (1993)',
 'Gladiator (2000)',
 'Ghostbusters (1984)',
 'Godfather: Part II, The (1974)',
 'Blade Runner (1982)',
 "Bug's Life, A (1998)",
 'Usual Suspects, The (1995)',
 'Abyss, The (1989)',
 'Aliens (1986)',
 'Galaxy Quest (1999)',
 'Total Recall (1990)',
 '2001: A Space Odyssey (1968)',
 'Babe (1995)',
 'Stand by Me (1986)',
 'GoodFellas (1990)',
 'Jaws (1975)']
```

Fig-28: MAP Score for Hybrid model

```
print("Mean Average Precision for Hybrid model = ",sum1/count1)

Mean Average Precision for Hybrid model =  0.32803170385133684
```

References

1) Y. Koren, R. Bell, and C. Volinsky, “Matrix factorization techniques for recommender systems,” Computer, vol. 42, no. 8, pp. 30–37, 2009

2) Fei Sun, Jun Liu, Jian Wu, Changhua Pei, Xiao Lin, Wenwu Ou, and Peng Jiang. 2019. BERT4Rec: Sequential Recommendation with Bidirectional Encoder Representations from Transformer. In The 28th ACM International Conference on Information and Knowledge Management (CIKM ’19), November 3–7, 2019, Beijing, China. ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3357384.3357895

Thank you!