Recap

11 January 2025

Hybrid Recommender system → Personalized
↳ Diverse

Phase I → Content Based ✓

Phase II → Collaborative filtering ✓

Phase III → Hybrid Recommender system

Phase I
Phase II
→ (+)

Weighted approach

w_1 → Content Based

w_2 → Collaborative

$Y = w_1 \times CB + w_2 \times CF$

Song name, Artist name, K

Content Based

Song → Attributes

Music Info → track id, Song name, artist name, attributes

Vectors → 50K → Vectorize

Similarity scores (A, B) input

Similarity scores

input song vector

Transformed matrix

Cosine similarities

Song name + Artist name → Row extract

index → transformed matrix → input vector

Collaborative filtering → User History

track ids → 30K

track id | User id | Playcount

Rows → track ids → 60 GB

Columns → User ids

Dask → Chunking → sparse matrix → integer

Categorical → track ids array store → 50K

Song name + artist name

50k → 30k → User history

Songs dataset filtered → Row → track id

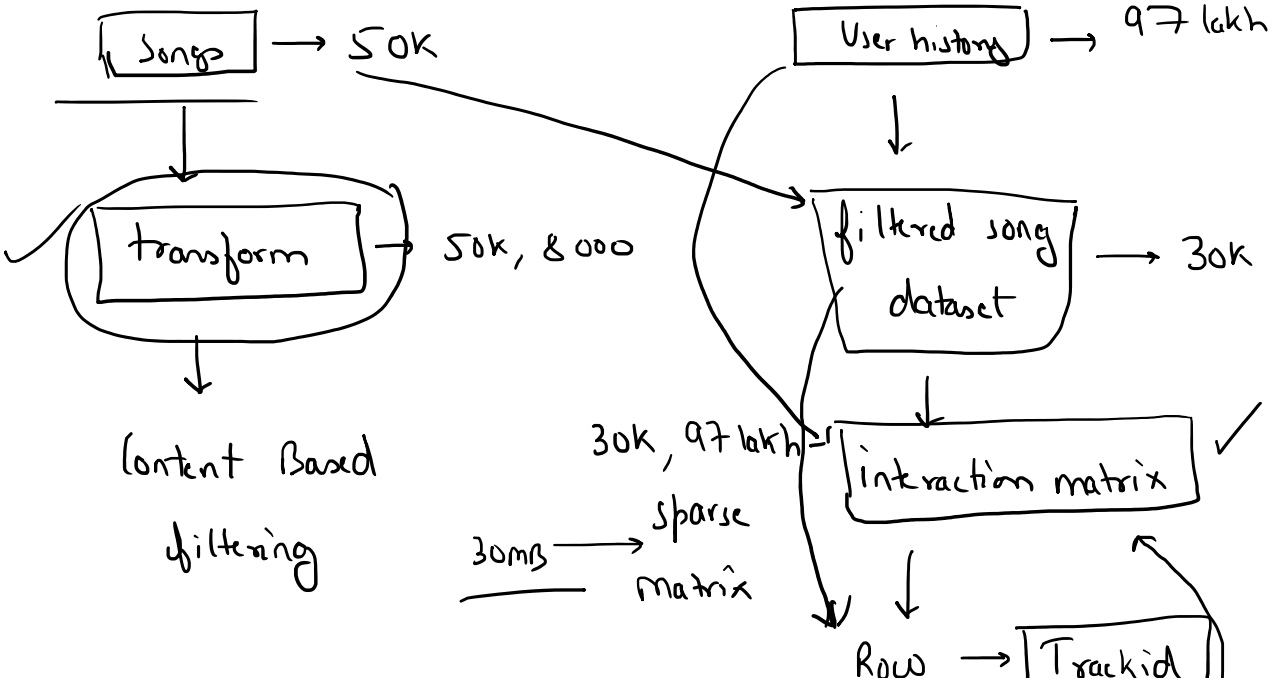
track ids array → ind → interaction matrix

input vector + interaction matrix

similarities → index wise sort

track ids → filtered songs

Rows → Artist, song name, preview URL



Row → Music_info.csv → 50K

User listening history.csv → 97 lakh

Music Info → Transformed matrix → Content Based.

User history → Filtered → 30K

interaction matrix → 30K, 97 lakh → 30MB

track ids array → numpy array

Hybrid Recommender system

Content Based + Collaborative

Similarity scores

W_1 (CB) + W_2 (CF) → Similarity scores

Song name
+
Artist name
≥ number of recommendations = 10

CF → 30K

CB → filtered songs → 30K

ind → Transformed matrix → 30K, 8000

CF → Filtered song → track id

track ids → ind

interaction matrix → Similarity scores

$W_1 = 0.3$ $W_2 = 0.7$

$w_1 = 0.3$

$\gamma = 0.3 \times$ Sim scores CB $+ 0.7 \times$ Sim scores of CF

2 Problems.

Song A Song B Song C

Problem 1 → Sim score CB Sim score CF

Matched	Song B	Song A
	Song A	Song B
	Song C	Song C

track ids → categorize. → lexical order

Song A , Song B , C ; D - - - -

A _ _ _ _ _ _

B _ _ _ _ _ _

C _ _ _ _ _ _

D _ _ _ _ _ _

filtered data. → sort → track ids.

indexes

transform.

CB CF

Song A
Song B

+

Song A
Song B

Problem 2

0.2 x CB → 0.9 x 0.2 = 0.18 ✓

0.8 x CF → sparse → 0.2 x 0.8 = 0.16

Normalize

$$\text{Normalization} = \frac{\text{Sim score} - \min}{\max - \min}$$

$Y = W_1 \times \text{Normalized CB} + W_2 \times \text{normalized CF}$

$Y = W_1 \times \text{normalized CB} + W_2 \times \text{norm CF}$

→ sort → top recommendation.

✓

show top k recommendations