Recap

12 January 2025 21:11

Hybrid Recommender

- ✓ Phase I → Content Based
- ✓ Phase II → Collaborative
- ✓ Phase III → Hybrid system

(+) → $W_1 \times$ Content Based + $W_2 \times$ Collaborative

Improve User engagement → Time spent ↑ → Personalized → Content Based

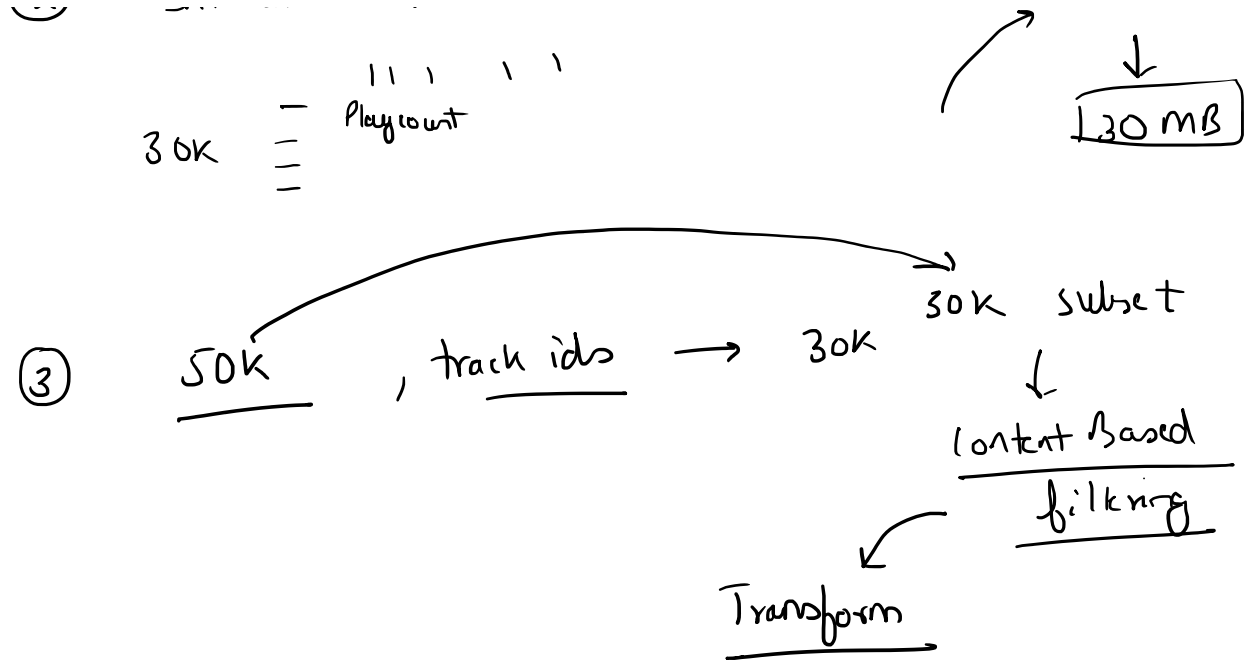
User retention → Subscribers → Renewals ↑

Diverse → Collaborative

Hybrid Recommender system → Improvements → Deployment

① Vectorize / Transform → sparse matrix → Content Based

② Interation matrix → 60GB → sparse matrix


30K
Playcount
130 MB
3
50K
, track ids
30K
30K subset
Content Based filtering
Transform

Improvements in the system

12 January 2025 21:11

1) Dynamic → Weight values.

similar ← W_1 , W_2

User → more dynamic → W_2 → Collaborative filtering

more personalized → W_1 ↑ → content based

2) 30K W_1 × Content Based + W_2 Collaborative → Combined similarity scores

50K → 20K → User history of
- geographical
- time range
- niche

Song name / Artist name → 30K → Hybrid recommender

50K → Only content based

Error

✓ Cold start → New song → Users

✓ Cold start → New song → Users

↓

Content based ←

3) Streamlit → Python script

↓

loading → load memory

↓

Faster

Metrics to evaluate

12 January 2025 21:11

→ labelled data.

accuracy (y-true , y-pred)

Precision and Recall

Precision @ k Recall @ k → number of recommendation

(k = 10

3 [User A]

[Item 1 Item 2 Item 3 Item 4 Item 5]

[Item 2 Item 4 Item 5] → Prediction.

[Item 1 Item 2 Item 4] Precision@3 = 2/3

Precision → Accuracy of Recommendations.

→ 3 Recommendation → Relevant quality

Precision @ 3 = No. of relevant Recom / No. of top k Recomm

Recall → 10 Relevant Recommendation

Recommended

$\frac{TP}{TP+FP}$ $\frac{TP}{TP+TN}$

$$\text{Recall@k} = \frac{\text{No. of Recommendation in top k}}{\text{No. of all relevant Recommendations}} \qquad \text{Recall@10} = \frac{7}{10}$$

Model Based Recommender Systems

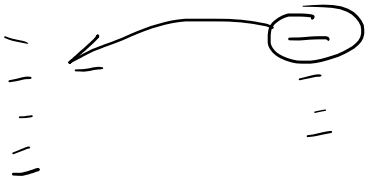
12 January 2025 21:12

Collaborative filtering

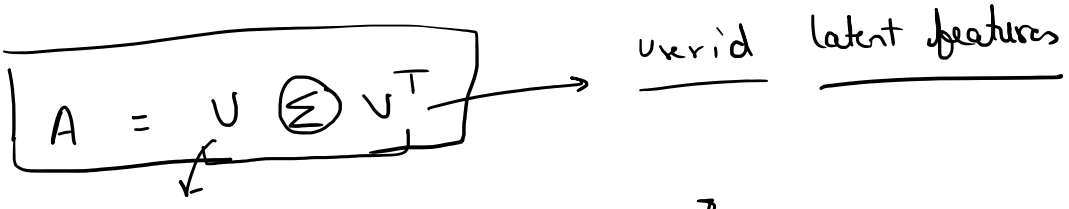
SVD factorization

User history

track id	Userid	Playcount



latent features → hidden features about our tracks



relevance

sparse	→ Playcount → Predict

User A Play → Sort

↓

top k recommendations.

Regression → MAE, MSE, RMSE

↓

Improve our model → Good Recommendations

Boosting → XgBoost, LightGBM

Collaborative filtering

Surprise, Recsys