Chrome File Edit View History Bookmarks Profiles Tab Window Help 10:00

Community Dashboard - Applie × | Airmeet: Scenario based In ×

airmeet.com/event/52601870-e8ca-11ec-9f61-dd73ad3844ed

Update

Scenario based Interview Questions on RecSys-part 2 | View Schedule

Moderation People Messages Alerts Settings

LIVE 00:19 Scenario based Interview Questions on RecSys-part 2 37 people in session REC

Pause Session End Session

You
Founder, Applied Roots · an hour ago

Hi Folks, we will start the session at 10AM, sharp. Thank you for joining early.

Please wait for a few seconds for everyone to join
Connecting...

2 / 2

Rec-Sys-2

Recap:

1. problem-def ; scale ; key aspects

2. A/B test ; metrics

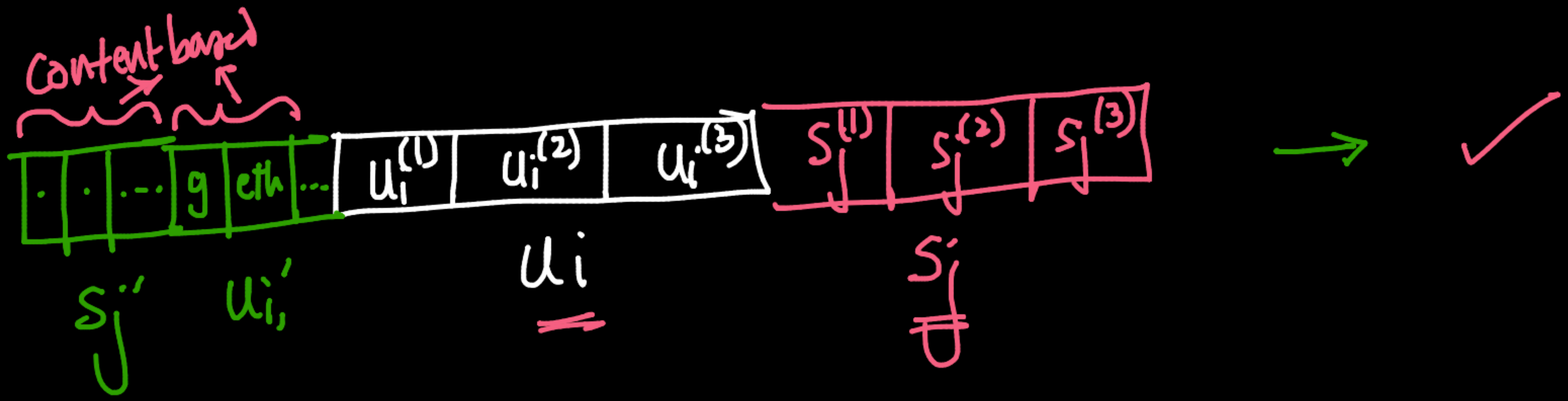
3. data : $\langle t_k, u_i, s_j \rangle \rightarrow$ L/Pl/S.... (actions)

4. MF $\rightarrow$ time
   MF $\rightarrow$ many actions $\rightarrow$ multiple MF

✓ $L_{n\times m} = A^{(1)}_{n\times d} \; B^{(1)}_{d\times m} \; : \; \underline{\underline{u_i^{(1)}}} \quad \underline{\underline{s_j^{(1)}}}$

$P_{n\times m} = A^{(2)}_{n\times d} \; B^{(2)}_{d\times m} \rightarrow u_i^{(2)} \quad s_j^{(2)}$

$S_{n\times m} = \ldots \rightarrow u_i^{(3)}, \; s_j^{(3)}$



[previous - session]

accounting for multiple actions

Chrome File Edit View History Bookmarks Profiles Tab Window Help 10:09

Topics:

→ RecSys → time (recency)

→ scale of MF (training)

→ cold-start (optional)

→ productionize (retrain ; serving ....)

→ Interpretable

6 / 6

1 Recsys → recency (Youtube, Spotify ....)

↳ recency + multiple actions

$$\begin{cases} R \rightarrow u_i^R, s_j^R \\ L \rightarrow u_i^L, s_j^L \\ \vdots \end{cases}$$

Chrome File Edit View History Bookmarks Profiles Tab Window Help

$$\min_{\substack{A^R, B^R \\ A^L, B^L}} \underline{\lambda_1} \underbrace{\| R - A^R B^R \|_F^2} + \underline{\lambda_2} \underbrace{\| L - A^L B^L \|_F^2}$$

Works if its only an offline system

$$R_{ij} = \text{func}(\text{actual rating}; \text{recency})$$

↳ ∝ actual rating

→ ∝ recency

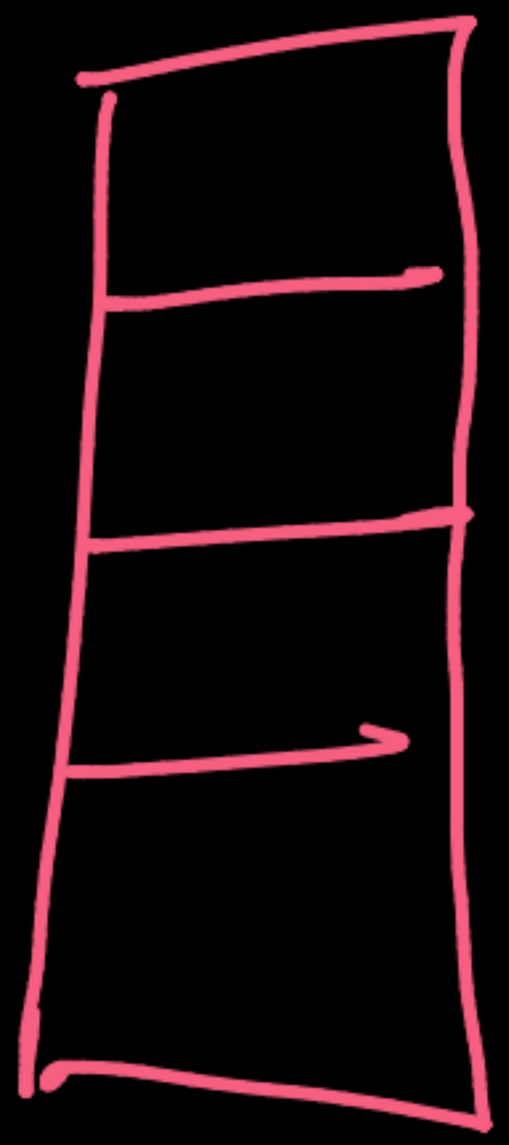
8 / 10

Problem:

1 hr

6 videos

DL



redo the whole MF in real-time

SOTA : DL

Neighborhood-based → embedding or represent... for a song

Summary:

Spotify:

$\boxed{\text{MF}}$: Co-factorization → ARRehman; Melodies; 90's songs . . .

→ recency → song/video/tweet/post/Image ($\overbrace{\qquad}^{1B}$)

↓

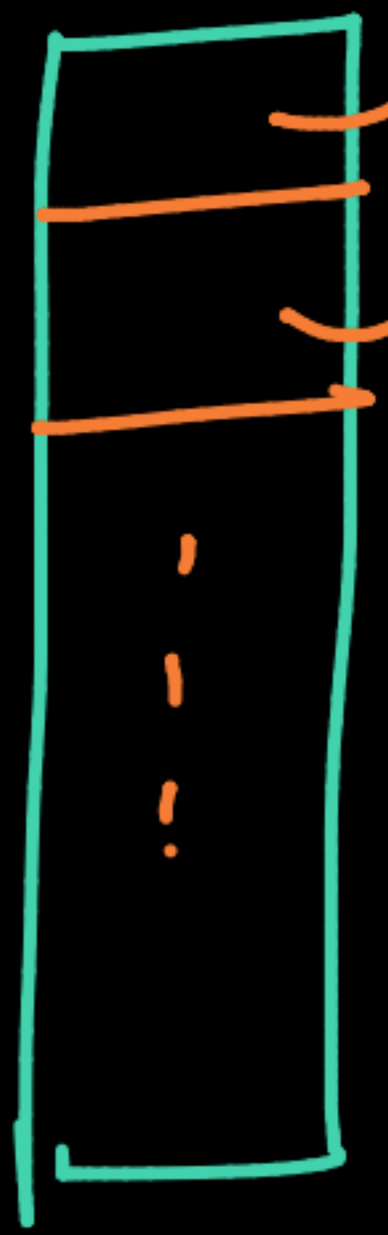
DL-embedding

Approximate-NN

~10's of MS

SCANN, LSH; FAISS;

→ Carnatic



carnatic → recency } ✓

ARRehman → MF } ✓

Q

General ML

GBDT → feature ~~vectors~~ engineering (How?)

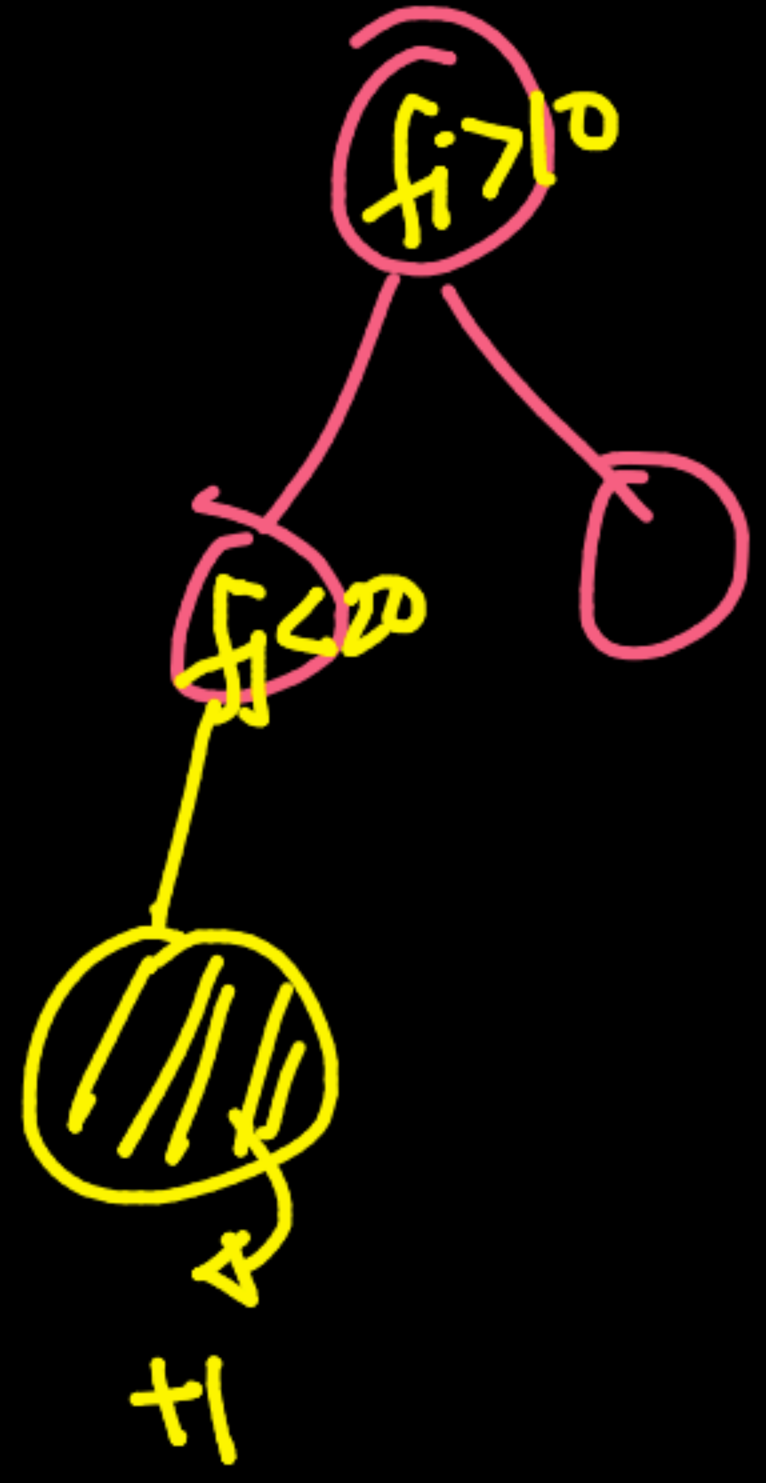
↓

logistic regression ✓

(Interactions ; scalability)

RF/GBDT/

DT → interaction feature



→ fi>10 and fj<20

Scale of training : MF

$$R_{n\times m} = \underset{n\times d}{A^R} \cdot \underset{d\times m}{B^R}$$

→ $\underline{100M}$ of users ; 100M of items

Songs

→ R

→ L

→ P

⋮

→ Sparsity - extreme

$$\min_{A^R, B^R} \| R - A^R \cdot B^R \|_F^2$$

$\rightarrow R^{[G_1]} \quad R^{[G_2]} \quad - - - -$

$L^{G_1} \quad L^{G_2} \quad - - - -$

$\}$ $\rightarrow$ cross-learning is being lost

$\rightarrow$ cluster users & items

↳ using metadata $\rightarrow$ $\boxed{\text{Indians}}$

e.g.: $<0.1\%$

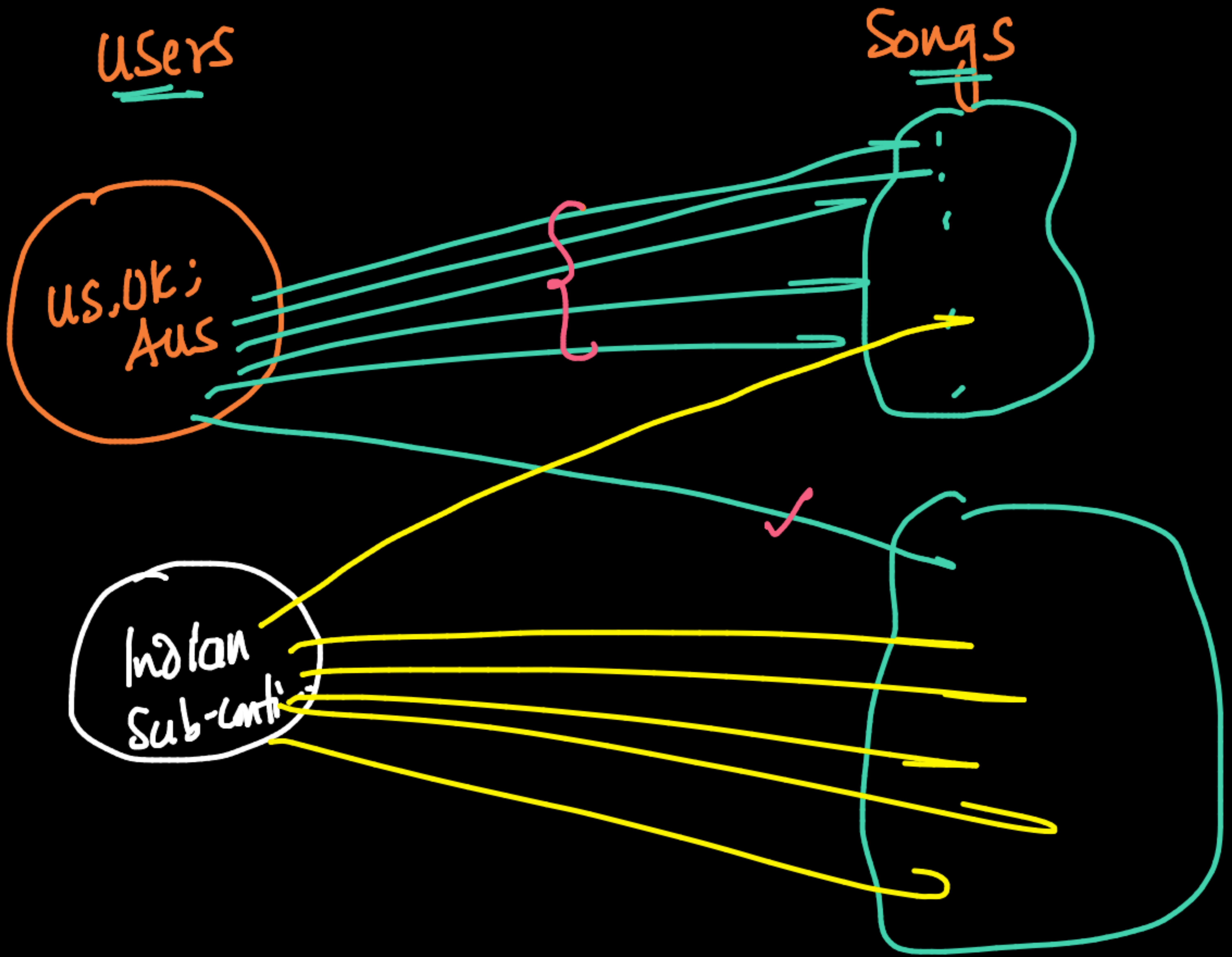
South Indians

⋮

$\boxed{\text{Careful}}$

Can: $\boxed{\text{US}}$; UK; Aus

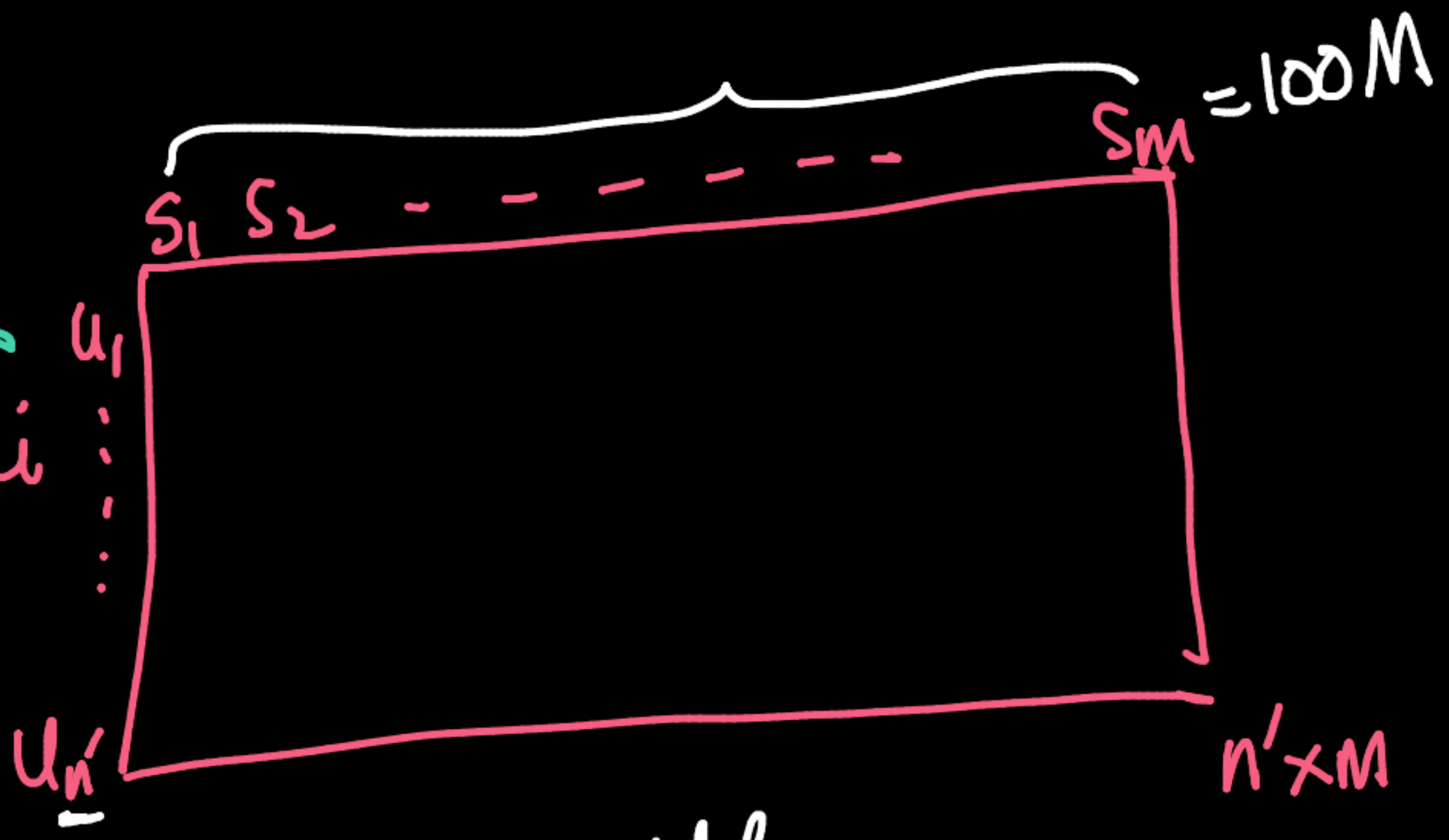
Bi-partite
Graph partitioning



Clustering based on PINCODE / City → (granularity)

problem:

Mumbai



n' is reasonable

$$R_{n' \times m} = A_{n' \times d} \quad B_{d \times m}$$

$n' = 1MM$

$m = 100MM$

Chrome File Edit View History Bookmarks Profiles Tab Window Help

1st

Cluster
→ age
→ geo
→ Hindi/Punjabi
Urdu
(Linguistic)
→ Spanish
→ Oriya/Bengali

make sure that we are not loosing much on cross learning

↓

bi-partite graph partioning.

18 / 22

$\boxed{2}$ alt. ideas for scale & sparsity

MF

$R_{100M \times 100M}$

Data:

$S_1 \; S_2 \; S_3 \ldots$

DL

d-dim embedding

$S_1 \quad S_2$

recency + NN

Chrome File Edit View History Bookmarks Profiles Tab Window Help 11:09

n : 100MM

m : 100 MM

Sparse $R_{n \times m}$

2

sampling:

top 100 ratings per user (X)

randomly sample 1000 ratings per user

alt :- 1% of ratings

recency (1-3 yrs)

non-trending but played

also consider #listens/replays..

20 / 22

Chrome File Edit View History Bookmarks Profiles Tab Window Help 11:10

smaller d ✓

3

$u_i \in \mathbb{R}^d$

$s_j \in \mathbb{R}^d$

$$\min_{u_i, s_j} \sum_{R_{ij} \neq NaN} \left(r_{ij} - u_i^T s_j \right)^2$$

✓ Sparse

$100 \times 10^6 \times 100 \times 10^6 = 10^{16}$ entries

21 / 22

Final System → MF + recency-NN + Trending

MF ↓ historical patterns (enough data)

Q

Interpret your MF models

d-dim

u_i's & s_j's

k-means + manual

elbow

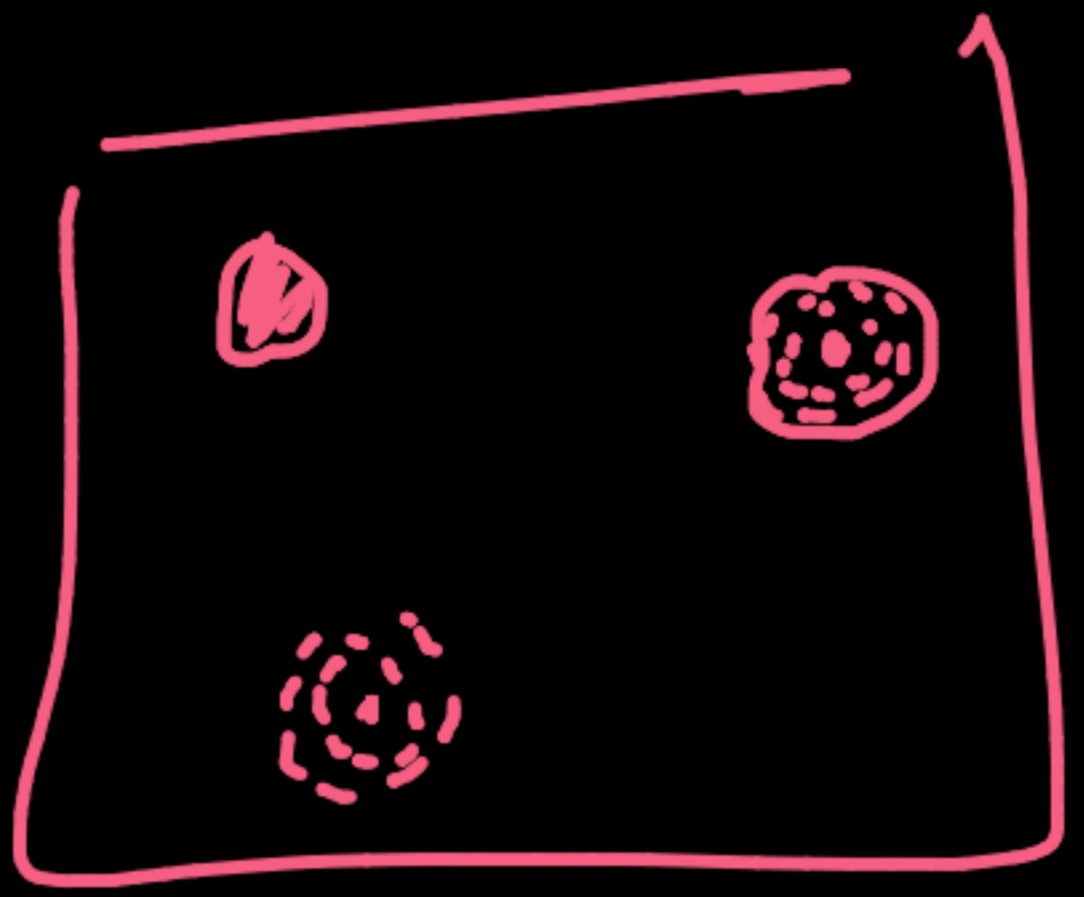
Cluster (k-medoids) ... Very expensive

DB-SCAN / Agglomerative Clustering

1

t-SNE/UMAP Visualization

2



3

d-dim

Playlists:
(user-created)

strong-signal

no additional man-power

Productionize:

MF

- Spark: daily → nightly
- retrain: incr. SGD
- Parallelize + weighted average by splitting data
→ ui → top 100/1000 recommen by Genre
(dict)

Kubernetes
Docker ✓

Kubeflow
SageMaker
Azure ML

NN-recency

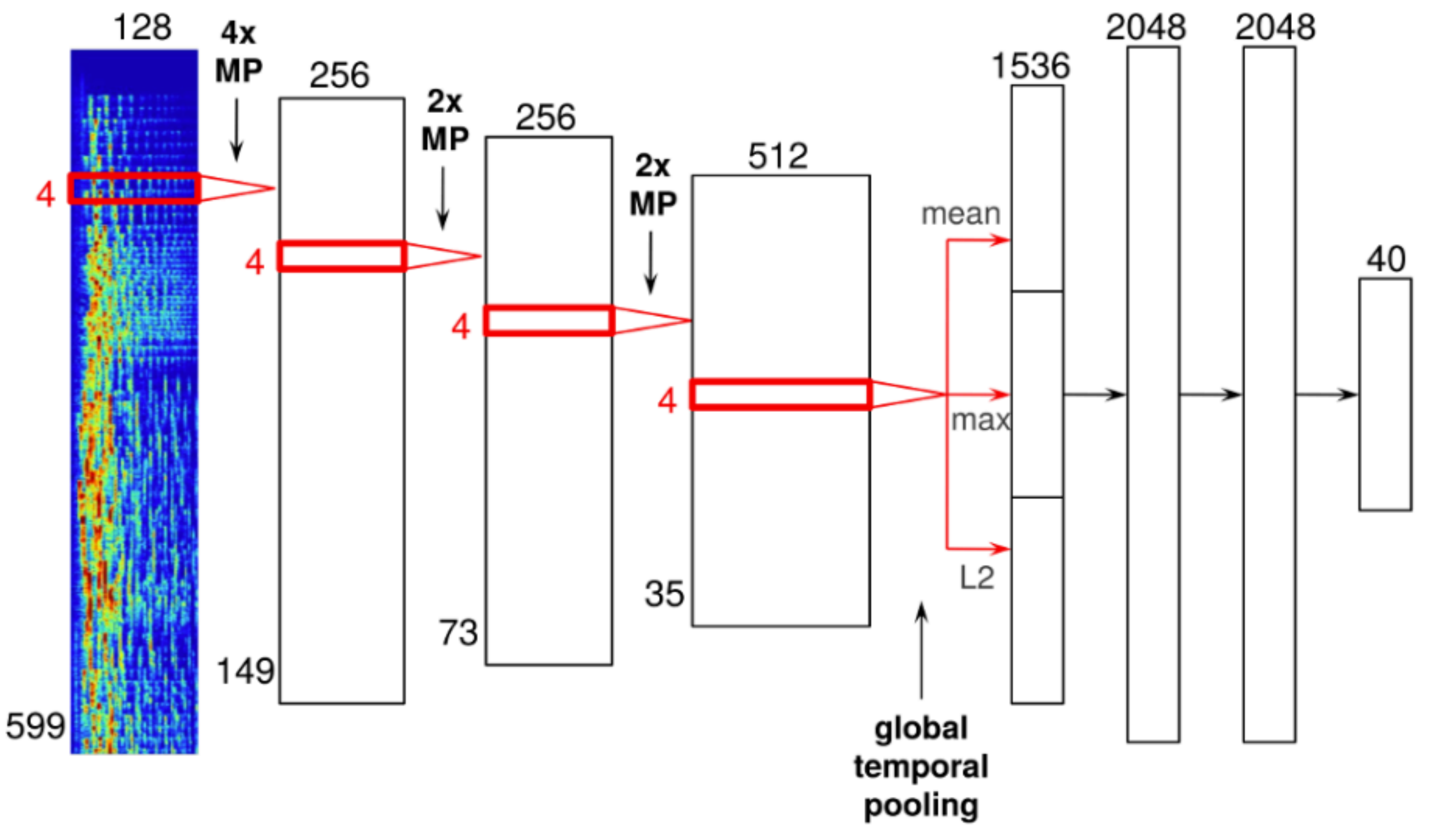
- ScaNN, FAISS
- DL-embeddings
-
-

Trending

- by loc: dict/DB
- SOTA: streaming-count algo..
-
-

25 / 25

Warning: gory details ahead! Feel free to skip ahead to 'Analysis' if you don't care about things like ReLUs, max-pooling and minibatch gradient descent.



One of the convolutional neural network architectures I've tried out for latent factor prediction. The time axis (which is convolved over) is vertical.

The input to the network consists of **mel-spectrograms**, with 599 frames and 128 frequency bins. A mel-spectrograms is a kind of **time-frequency representation**. It is obtained from an audi… ws. Each of thes… are then