Music Recommendation System

By Carols

Dataset

- Full history for 1 million of users (110 songs/user)

- Half of History 110000 users

- History Information:

  - User_id

  - Song_id

  - play_times

- Each song has the corresponding name, album, genre (386213 songs)

b80344d063b5ccb3212f76538f3d9e43d87dca9e   SOKSIKA12A6D4F5DC7   1
b80344d063b5ccb3212f76538f3d9e43d87dca9e   SOKUZGZ12A8C136F88   1
b80344d063b5ccb3212f76538f3d9e43d87dca9e   SOLFJVY12A6D4F9DCA   1
b80344d063b5ccb3212f76538f3d9e43d87dca9e   SOLGNOE12A8C139CA9   1
b80344d063b5ccb3212f76538f3d9e43d87dca9e   SOLLGNU12AF72A4D4F   1
b80344d063b5ccb3212f76538f3d9e43d87dca9e   SOLUHPJ12A8C13AB0F   1
b80344d063b5ccb3212f76538f3d9e43d87dca9e   SOLXHAI12A6D4FD6F3   1
b80344d063b5ccb3212f76538f3d9e43d87dca9e   SOMGIYR12AB0187973   6
b80344d063b5ccb3212f76538f3d9e43d87dca9e   SOMLMKI12A81C204BC   1
b80344d063b5ccb3212f76538f3d9e43d87dca9e   SOMSQJY12A8C138539   1
b80344d063b5ccb3212f76538f3d9e43d87dca9e   SOMZWUW12A8C1400BC   6
b80344d063b5ccb3212f76538f3d9e43d87dca9e   SONKSNV12A58A7F654   2
b80344d063b5ccb3212f76538f3d9e43d87dca9e   SONRXOY12AB0181E84   1
b80344d063b5ccb3212f76538f3d9e43d87dca9e   SONSAEZ12A8C138D7A   1
b80344d063b5ccb3212f76538f3d9e43d87dca9e   SONVJAW12A8C139EFB   1
b80344d063b5ccb3212f76538f3d9e43d87dca9e   SONYTAN12A8C13BF88   1
b80344d063b5ccb3212f76538f3d9e43d87dca9e   SOOCURD12A81C21C23   5
b80344d063b5ccb3212f76538f3d9e43d87dca9e   SOOGGEX12A58A7DACF   1
b80344d063b5ccb3212f76538f3d9e43d87dca9e   SOOKGRB12A8C13CD66   1
b80344d063b5ccb3212f76538f3d9e43d87dca9e   SOOSIVQ12A6D4F8AE0   1
b80344d063b5ccb3212f76538f3d9e43d87dca9e   SOPBCSY12A6D4F5DC4   1
b80344d063b5ccb3212f76538f3d9e43d87dca9e   SOPCVQE12AC468AF36   1
b80344d063b5ccb3212f76538f3d9e43d87dca9e   SOPZAUC12A58A7DB24   1
b80344d063b5ccb3212f76538f3d9e43d87dca9e   SOQDBOA12A8C13E5E6   1
b80344d063b5ccb3212f76538f3d9e43d87dca9e   SOQEMEN12A8C13BF8B   1
b80344d063b5ccb3212f76538f3d9e43d87dca9e   SOQIVUD12AB01821D2   1
b80344d063b5ccb3212f76538f3d9e43d87dca9e   SOQJLDY12AAF3B456D   1
b80344d063b5ccb3212f76538f3d9e43d87dca9e   SOQLCKR12A81C22440   1
b80344d063b5ccb3212f76538f3d9e43d87dca9e   SOQMUDW12A58A7AF03   1

All the data are downloaded from
the Echo Nest

How the System works?

❖ General Idea is collaborative filtering approach and user-based similarity

$$h_{ui}^{U} = \sum_{v \in \mathcal{U}} w_{uv} r_{vi}$$

How to calculate:
- w: user similarity?
- r: rating for each song of each user?

User Similarity

Analogies to textual information retrieval

term

occurrences of this term in the document

document (the listening history of user u)

b80344d063b5ccb3212f76538f3d9e43d87dca9e	S0KSIKA12A6D4F5DC7	1
b80344d063b5ccb3212f76538f3d9e43d87dca9e	S0KUZGZ12A8C136F88	1
b80344d063b5ccb3212f76538f3d9e43d87dca9e	S0LFJVY12A6D4F9DCA	1
b80344d063b5ccb3212f76538f3d9e43d87dca9e	S0LGN0E12A8C139CA9	1
b80344d063b5ccb3212f76538f3d9e43d87dca9e	S0LLGNU12AF72A4D4F	1
b80344d063b5ccb3212f76538f3d9e43d87dca9e	S0LUHPJ12A8C13AB0F	1
b80344d063b5ccb3212f76538f3d9e43d87dca9e	S0LXHAI12A6D4FD6F3	1
b80344d063b5ccb3212f76538f3d9e43d87dca9e	S0MGIYR12AB0187973	6
b80344d063b5ccb3212f76538f3d9e43d87dca9e	S0MLMKI12A81C204BC	1
b80344d063b5ccb3212f76538f3d9e43d87dca9e	S0MSQJY12A8C138539	1
b80344d063b5ccb3212f76538f3d9e43d87dca9e	S0MZWUW12A8C1400BC	6
b80344d063b5ccb3212f76538f3d9e43d87dca9e	S0NKSNV12A58A7F654	2
b80344d063b5ccb3212f76538f3d9e43d87dca9e	S0NRX0Y12AB0181E84	1
b80344d063b5ccb3212f76538f3d9e43d87dca9e	S0NSAEZ12A8C138D7A	1
b80344d063b5ccb3212f76538f3d9e43d87dca9e	S0NVJAW12A8C139EFB	1
b80344d063b5ccb3212f76538f3d9e43d87dca9e	S0NYTAN12A8C13BF88	1
b80344d063b5ccb3212f76538f3d9e43d87dca9e	S0OCURD12A81C21C23	5
b80344d063b5ccb3212f76538f3d9e43d87dca9e	S0OGGEX12A58A7DACF	1
b80344d063b5ccb3212f76538f3d9e43d87dca9e	S0OKGRB12A8C13CD66	1
b80344d063b5ccb3212f76538f3d9e43d87dca9e	S0OSIVQ12A6D4F8AE0	1
b80344d063b5ccb3212f76538f3d9e43d87dca9e	S0PBCSY12A6D4F5DC4	1
b80344d063b5ccb3212f76538f3d9e43d87dca9e	S0PCVQE12AC468AF36	1
b80344d063b5ccb3212f76538f3d9e43d87dca9e	S0PZAUC12A58A7DB24	1
b80344d063b5ccb3212f76538f3d9e43d87dca9e	S0QDB0A12A8C13E5E6	1
b80344d063b5ccb3212f76538f3d9e43d87dca9e	S0QEMEN12A8C13BF8B	1
b80344d063b5ccb3212f76538f3d9e43d87dca9e	S0QIVUD12AB01821D2	1
b80344d063b5ccb3212f76538f3d9e43d87dca9e	S0QJLDY12AAF3B456D	1
b80344d063b5ccb3212f76538f3d9e43d87dca9e	S0QLCKR12A81C22440	1
b80344d063b5ccb3212f76538f3d9e43d87dca9e	S0QMUDW12A58A7AF03	1

User Similarity

$$idf_t = log\frac{n}{df_t}$$

total number of users

number of users who listened to song t

- the value of idf if **high** for **rare songs**, which were listened by only a few users
- the **users** are probably **similar**, if two users **listened to the same track** and nobody else did it

How to calculate user similarity?

❖ **The similarity Wuv** between users u and v :

the sum of idf of the tracks that they have listened both

$$\sum_{\substack{i:M(u,i)>0 \\ M(v,i)>0}} idf_i$$

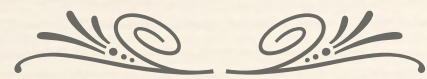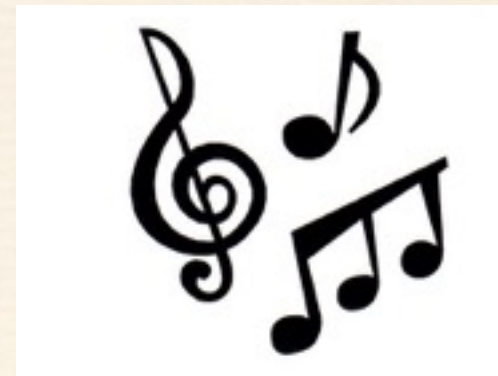# Rating for each song?

- Rvi of how much a user v likes a track i is chosen as 1 if user v listened to a track i and 0 otherwise.

- **Rvi = 1/Cv** (**Cv**: the sum of all play counts from user v's history)

- So all the tracks are ranked equally regardless of their play counts.

Recommendation!

$$h_{ui}^{U} = \sum_{v \in \mathcal{U}_s} \left(\sum_{\substack{i:M(u,i)>0 \\ M(v,i)>0}} idf_i \right) \frac{1}{c_v}$$

The Resulting score function

recommend songs
with top 20 scores!

Evaluation

$$AP(u, y) = \frac{1}{n_u} \sum_{k=1}^{\tau} P_k(u, y) \cdot M_{u, y(k)}$$

Baseline : 0.15037
Our Method : 0.21514

Let the joy of Carols Begin🎵