Music Recommendation System

TEAM – JARVIS

Aakash Bhatia – 201711031

Krutika Mahavadia – 201711068

Akash Thakrar – 201711038

Chintan Sangani – 201711055

1. Model:

- New User:

  Recommendation to new user will be done by two techniques given below, few top songs from each technique will be forming a list of 10 songs.

  i. Popularity

     Here most popular few songs of whole system are recommended to new user.

  ii. Content Based Filtering – CB

     Favorite or preferred Artist and Album will be asked (Optional) to new user and CB will be based on Artist and Album provided by new user (if any).

- Existing User:

  Recommendation to existing user will be done by five techniques given below, few top songs from each technique will be forming a list of 10 songs.

i. Collaborative Filtering – CF

Here recommendation given will be such songs that are heard by similar type of persons in system and top few songs will be considered.

ii. Content Based Filtering – CBF

Here recommendation given will be such songs that match to the favorite artist and album of user in system and top few songs will be considered.

iii. Cosine Similarity

Here recommendation given will be a song that matches to currently(ongoing) listening song.

iv. Personalized Filtering – PF

Here recommendation given will be such songs that are already heard by current user based on their listen_count*score and top few songs will be considered.

v. Popularity

Here most popular few songs of whole system are recommended to new user.

2. Decision Parameters:

- Artist
- Album
- Release Year
- Listen Count

3. Rewarding and penalizing policy:

- Initial Score – 1.0
- Reward – 5%

- Penalty – 2%

When any song is recommended to user first time then it's initial score will be 1.0. When user selects an item from a bunch of 10 recommendation, that item gets reward of 5% where as other items get 2% penalty on their score.

4. Model Description:

- POPULARITY:

  i. Create song wise groups,
  ii. Aggregate Listen count per group,
  iii. Derive proportion of aggregation with respect to total aggregation,
  iv. $\text{Aggregation}_{\text{Per\_group}}$: $\sum_{1 \text{ to } n}$(Listen Count),
  v. Total Aggregation : $\sum_{1 \text{ to } n}(\text{Aggregation}_{\text{Per\_group}})$,
  vi. $\text{Proportion}_{\text{Per\_group}}$: ($\text{Aggregation}_{\text{Per\_group}}$ / Total Aggregation)*100.

- COLLABORATIVE FILTERING:

  i. User based collaborative filtering,
  ii. Co-occurrence Matrix is used,
  iii. Jaccard similarity is used to find similar users.

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}.$$

(If A and B are both empty, we define $J(A,B) = 1$.)

$$0 \leq J(A,B) \leq 1.$$

- CONTENT BASED FILTERING:

    i. Create $\text{matrix}_{\text{User x Album}}$,
    ii. Create $\text{matrix}_{\text{User x Artist}}$,
    iii. Get User row from matrix while login,
    iv. Fetch (highest listened artist & album),
    v. Create $\text{list}_{\text{Artist}}$ -> songs ∈ Artist,
    vi. Create $\text{list}_{\text{Album}}$ -> songs ∈ Album,
    vii. Extract Highest listened two songs.

- COSINE SIMILARITY:

    i. A -> Attributes of current(ongoing) listening songs. (Like Artist, Album)
    ii. B -> Attributes ∈ (All songs – A)

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

- PERSONALIZED FILTERING:

    i. Personalized_Score -> α* β,
    ii. α = Listen count,
    iii. β = Score,
    iv. Sort user list according to Personalized_Score in Descending order,
    v. Fetch (Top few songs).

5. Dataset:

- Million song datasets from MusicBrainz
- Two files: triplet file and metadata file.
- 10,00,000 entries
- Split ratio – 70:30