Summer Research 2019

Analysis of Recommender System

Professor Fahmida Hamid

Personalized Hybrid Artist Recommender System

Linh Tang

Department of Computer Science

Grinnell College

Introduction

Recommender System (RS) utilizes various sources of data to infer user interests and recommends possible list of items so that users conveniently find what they are looking for.

How does a recommender system know you well enough so it quenches your thirst by recommending a list of items of your choice? Researchers are investigating these basic questions by looking into this potential research area from different domains to offer a reliable model for achieving user satisfaction. Every system is significantly different than others; the features and aspects of consideration in various domains are also quite dissimilar.

In this work, we introduce **a personalized hybrid artist recommender system (PHARS)** that uses user's listening history and their bias towards gender and types of musicians to recommend artists. Our model offers a novel approach to update personalized music recommender system as it extracts implicit information from the users listening pattern.

Datasets

We mainly used two large datasets: the LastFM360K and the MusicBrainz. The **LastFM360K** dataset has listening history of **358855** unique users about **147669** unique artists with **17,559,530** listening history. **MusicBrainz** dataset is a **mega-database** that provides information of songs and artists' features such as artist type, artist area, and artist gender.

Table 1 and table 2 present the results of preliminary analysis of the LastFM360K dataset.

Gender	Male	Female	Other
Percentage	23.63	67.25	10.12

Table 1: Gender Demographics: LastFM360K Dataset

Age-range	Percentage
[0 ~ 12]	0.37
[13 ~ 17]	5.61
[18 ~ 24]	40.91
[25 ~ 34]	23.82
[35 ~ 44]	5.78
[45 ~ 64]	1.57
[65 ~ 100]	0.29

Table 2: Age Demographics: LastFM360K Dataset

Motivation & Challenges

Music is very powerful as it can improve one's mood and metal state quick and easily. Music comes with lots of variations and genres. Music is different than movies, books, or other daily necessities. A music recommender system should be more sensitive and sophisticated than other system, such as book or movie recommender system., as it has to overcome various major challenges:

Information Age

- The amount of digital data stored appears to be growing approximately exponentially (Moore's law).
- The explosion of published information would be moot if the information could not be found, annotated, and analyzed

Modeling Human Subjects

Humans psychological make-ups depend on an array of emotional and motivational parameters, e.g spatial (area living in, area born and brought up, friends and family) and temporal dimensions (age, current location, weather, etc.)

Availability & Scalability of Personal Information

- We do not always have access to all the private information that someone is comfortable sharing with.
- We need to find the dominant features to represent a human in a scalable manner.

Framework

Two most common paradigms to approach recommender system are Collaborative Filtering (CF) and Content-based (CB). CF system simply means that people collaborate to help each other perform filtering by recording their reactions to items that they use (Goldberg et al., 1992). In content-based (CB) recommender systems, the description of the attributes of items are used to make recommendations. As these traditional systems only consider similar users or similar items, users have fewer chances to be exposed to new items of different types that potentially may become their next favorite. Hybrid model have been designed to overcome these limitations and explore more possibilities. In our design, we use one of the most popular hybrid recommender system approaches: weighted feature combination (Aggarwal, 2016), to cascade the output of CF to CB, and then finally use personal to produce recommendations list.

Collaborative Filtering

- Build a user-artist implicit ratings matrix based on users listening history
- Implement baseline model using Alternating Least Square (ALS) matrix factorization paradigm.

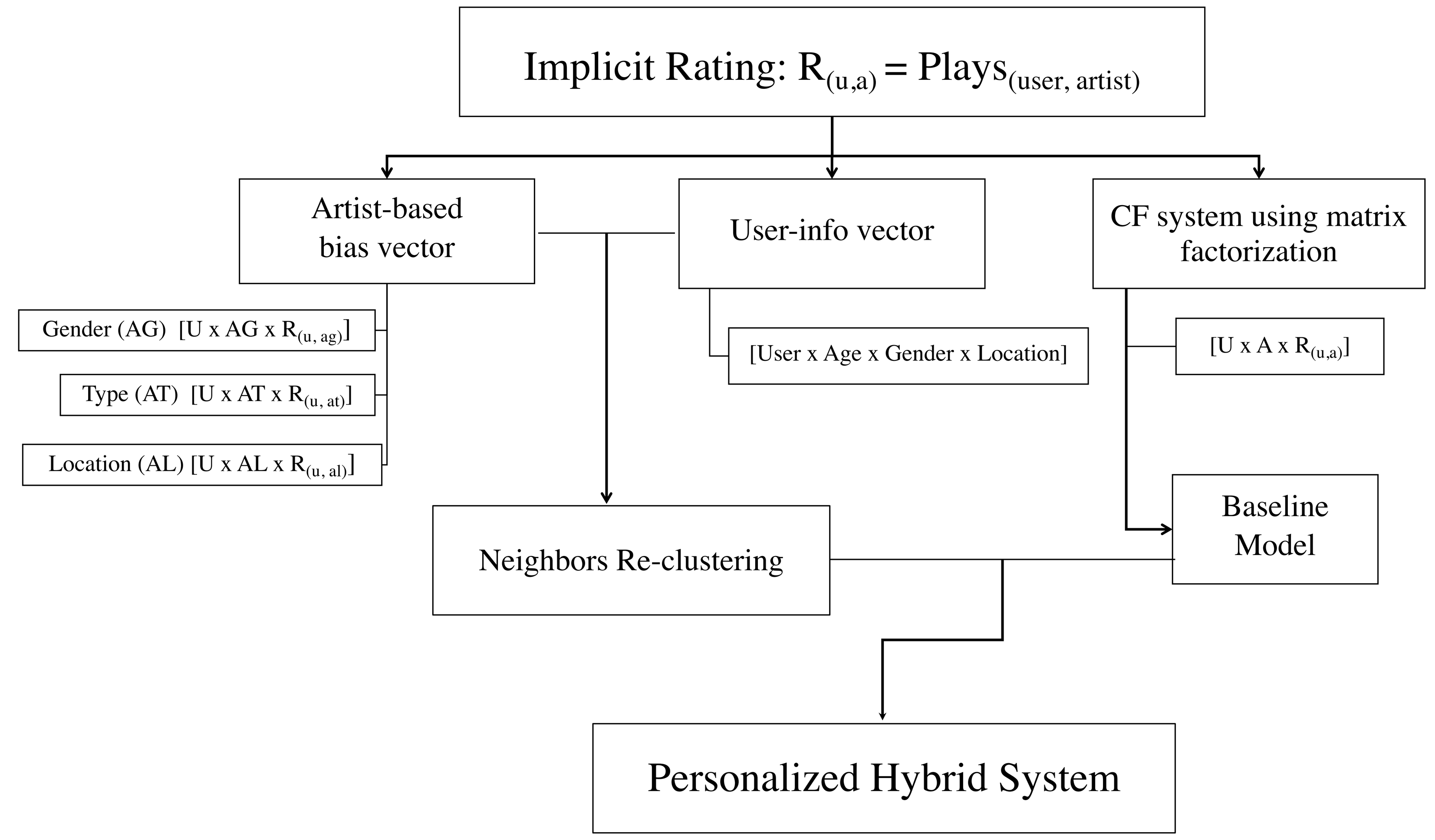
Personalization

- User-based: Build user-info vector from user's age, gender, location
- Item-based: Build artist-bias vector from user's preference towards different artist's types and genders.

Hybrid Model

- Build user-user similarity matrix from user-based and item-based vectors
- Combine the collaborative filtering model with the personalized neighborhood matrix

- ➢ **Implicit Feedback:** We consider users' artist preferences based on how many times they listen to some music (Yifan et al., 2008)
- ➢ **High-dimensional Space:** We consider users age-group, gender-group, user's gender-bias, and artists type-bias in a high-dimensional space to model users. We use KD-tree for this purpose.
- ➢ **Infusing two systems:** We incorporate the outcome of a basic collaborative filtering approach with K-nearest neighbors of a user.



Demonstration

Given an user-id, our model will suggest a personalized list of artists for the user with a default of 10 artists if not specified. An example of our output is shown below:

```
10 artists recommended for user c1eb6566cc1691664e2810f8f224e9122281a572

Collaborative Filtering System                  Personalized System

1 Blue Scholars , with a score of 1.137         1 Daniel Lanois , with a score of 5.549
2 D12 , with a score of 1.137                   2 Savath & Savalas , with a score of 5.225
3 Non Phixion , with a score of 1.134           3 Project Pat , with a score of 5.211
4 OuterSpace , with a score of 1.126            4 Necro , with a score of 5.174
5 Lil' Wyte , with a score of 1.122             5 Natalie Merchant , with a score of 5.12
6 N.W.A , with a score of 1.114                 6 Rivers Cuomo , with a score of 5.085
7 P.O.D. , with a score of 1.098                7 Adam Sandler , with a score of 5.084
8 Fort Minor , with a score of 1.095            8 Lisa Loeb , with a score of 5.082
9 Celph Titled , with a score of 1.092          9 Seu Jorge , with a score of 5.068
10 Alien Ant Farm , with a score of 1.082       10 Lil' Wyte , with a score of 5.055
```

Performance Evaluation

To evaluate the performance of our system, we use precision, recall and F-measure (F_1 score). In particular, we compare the recommendations list for an user and the actual lists that user listened to, and count the number of matches between these two lists and use the common formula to calculate F_1 score (shown below).

We split the dataset into 80-20 train-test ratio and calculate ***Precision, Recall, & F-measure*** on the randomly-chosen 1000 users in the test set and used the basic CF system as the baseline model. The result is presented in table below, PHARS shows significant improvement over the basic collaborative filtering model.

	CF System	Personalized
Precision	0.357143	0.476190
Recall	0.075552	0.140103
F-measure	0.124696	0.215922

$$\text{Precision} = \frac{\text{Correct Recs}}{\text{Total Recs}}$$

$$\text{Recall} = \frac{\text{Correct Recs}}{\text{Real Listenings}}$$

$$F_1 = \frac{2* \text{Precision}* \text{Recall}}{\text{Precision}+\text{Recall}}$$

Conclusion & Future Work

The goal of this research is to bring diversity in the recommendation list. We aim to keep the list interesting and relevant while making it more variable. We've worked on several directions and tried many different approaches. Our final personalized model find similar users by considering the their own group and gender group and then relies on the hybrid model to recommend new artists. The model showed promising results as it greatly improved the performance of the baseline model.

We are exploring different datasets for identifying dominating user features to offer a better music recommender model. Our next step is to use NLP techniques to relate song-lyrics to common human emotions and use it to offer a mood-based personalized recommender system.

Acknowledgements

I would like to thank Prof. Hamid for her guidance and mentorship; my friends and family for their financial and mental support; Grinnell College for the research grant, and my research group for their support throughout the project.

References

Charu C. Aggarwal. 2016. Recommender Systems: The Textbook, 1st edition. Springer Publishing Company, Incorporated.

David Goldberg, David Nichols, Brian M. Oki, and Douglas Terry. 1992. Using collaborative filtering to weave an information tapestry. *Commun. ACM*, 35(12):61–70.

Metabrainz, F. 2015. Musicbrainz database.

Yifan Hu, Yehuda Koren, and Chris Volinsky. 2008. Collaborative filtering for implicit feedback datasets. *2008 Eighth IEEE International Conference on Data Mining*.