What Makes Songs Popular?

Cheng-Chun Lee, I Made Sanadhi Sutandi and Skander Hajri

School of Computer and Communication Sciences, EPFL



 Utilize million songs dataset to discover interesting facts prior to songs' popularity.

Introduction

Song is one of the greatest creations of human kind in the course of history. Music has transformed itself into an area of industry and constantly changes over the time. Through this project, it is exciting for us to elaborate what factors influence songs' popularity at most.

Million Songs Dataset

We mainly use **Million Song Dataset** that has collection of audio features and metadata of popular and unpopular songs. In addition we also make use two additional datasets:

- The musiXmatch Dataset: Containing lyrics.
- The Echo Nest Taste Profile Subset:
 Containing users' profiles with their play count.
 Important fields of Million Song Dataset:
- track_id The primary key identifier.
- song_hotttnesss the popularity of a song measured between 0 1.

Data Preprocessing

- **Preprocessing** the big raw datasets using cluster operations (Spark, Hadoop).
- Removing unused attributes based on analysis of features correlation; 52 → 28 features.
- Filtering songs with missing songs hotness entries; 1 million → 581,965 datapoints.
- Grouping songs to popular songs (popularity
 ≥ 0.8) and unpopular songs (popularity ≤
 0.40)

Important Features of Popular Songs

We use random forest classifier to predict whether a song is popular/unpopular. We get a high accuracy of 97.52% and thus we observe the attribute feature_importances to see which features matter the most.

We figure out that two most important features are:

- artist_hotttnesss The popularity of an artist (last for shorter period)
- artist_familiarity The indication of how well-known an artist is (remain for longer-term)

Popularity Matters For Song Hotness

Are songs from popular artists usually popular? We collect same number of popular and unpopular artists in 2010, and analyze the song hotness of their songs. Surprisingly, it differs a lot!

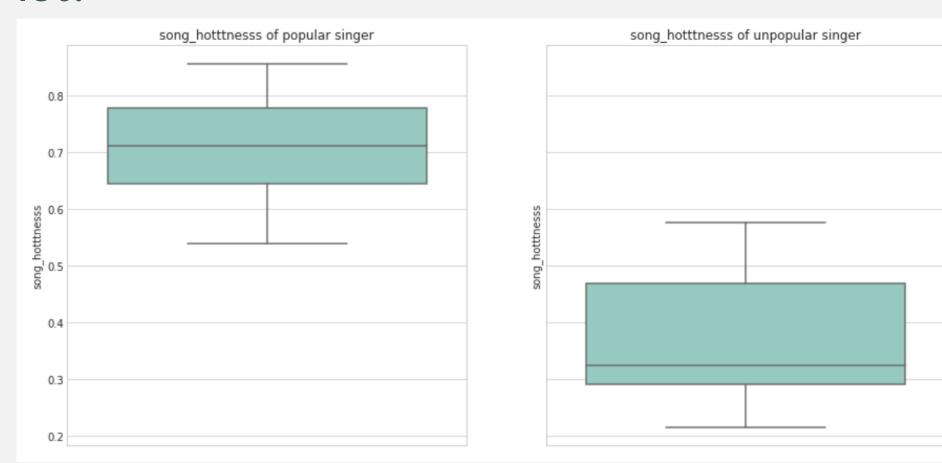


Figure: Song hotness from popular & unpopular artist

The Existence of Herding Bias

Do you notice that several songs from your playlist are actually from certain artists? We define this phenomenon as **herding bias** and we measure users' level of herding bias with the following formula:

LevelOfHerdingBias =
$$\frac{\sum_{m} p_{m} I_{m}}{\sum_{m} p_{m}}$$
 (1)

 p_m : total play count of songs from singer m I_m : indicator, = 1 if more than 2 songs from artist m and 0 otherwise

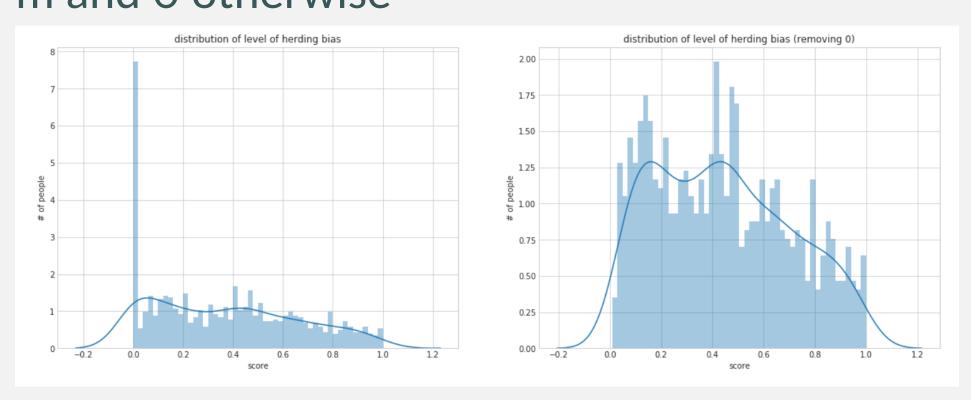


Figure: Distribution of herding bias
160 people out of 1022 people (only 16%) do
not have herding bias, and the median of
herding bias is 0.38.

First Performance Is Important

Do the songs in the first year matter a lot for artist? We choose several popular and unpopular artists during 1995-2000, 2000-2005 and 2005-2010, and observe the song hotness of their songs in their first year.

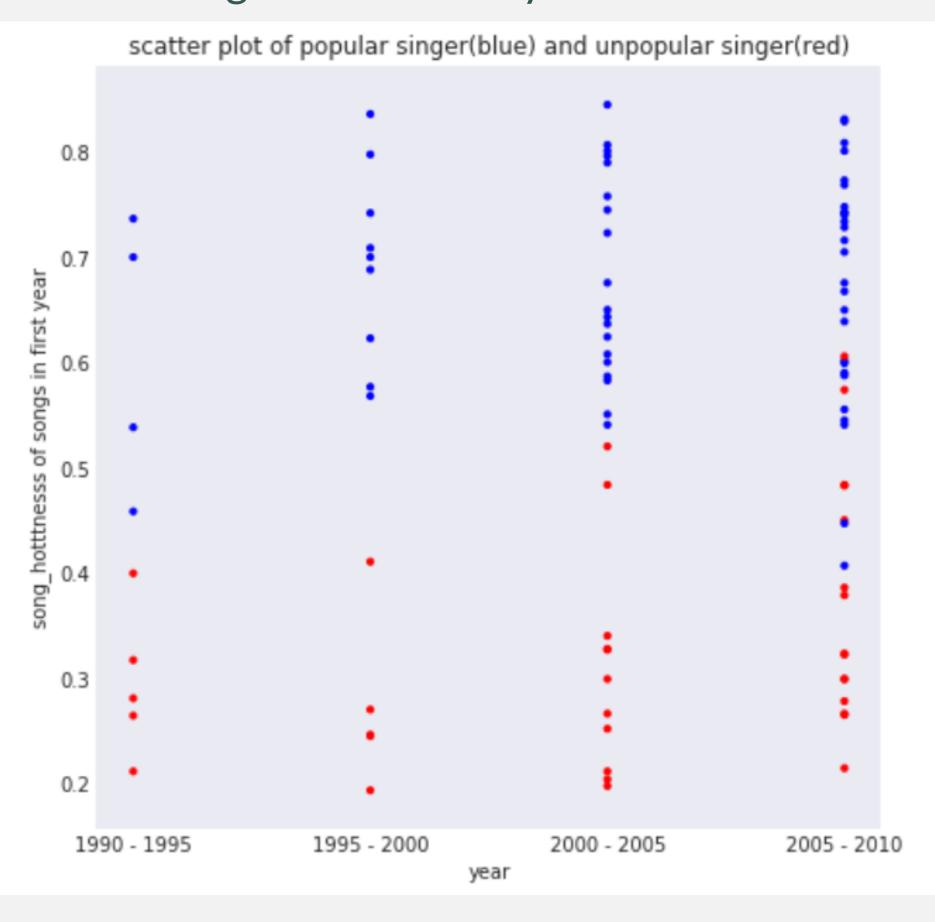


Figure: Song hotness of artist in their 1st year

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Lyrics Distribution

By exploiting MusiXmatch dataset, we can see many high-ranked words that recall **feelings** and **emotions**.



Figure: Composition of lyrics in both group

Presence of Slang Words

"Slang words", such as insults, are gathered in frequencies within popular/unpopular songs to give an estimation of the lyrics quality. Are borderline songs more interesting?

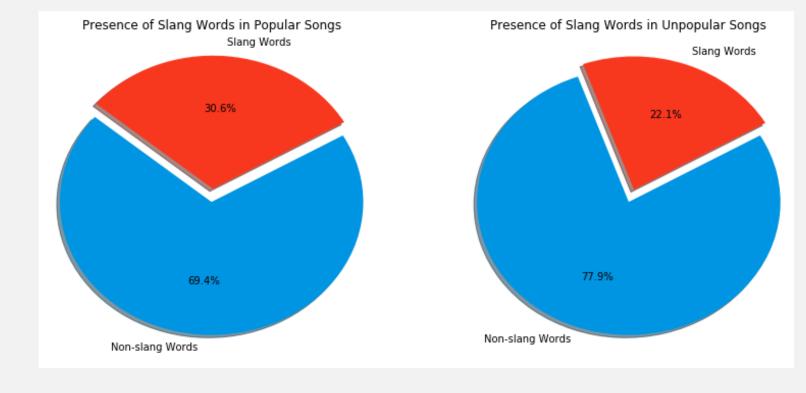


Figure: Ratio of slang words in both group

Conclusion

- People listen to his/her favorite artists. This is proven by important features of random forest classifier and analysis of herding bias phenomenon. "Rich become Richer"
- People care less about slang words in lyrics.

The project

• Github:

https://github.com/sanadhis/ITT-ADA-2017

1/