BanG Dream! Characters and Songs Analysis

0616023夏軒安

0616042呂明哲

0616210張皓鈞

Outline

- Data
  - Character data
  - Song data
- Model
  - Decision Tree
  - Random Forest
  - KNN
  - SVM-Classifier
- Result
- Analysis

BanG Dream

- What is BanG Dream!?
  - A music game
  - With anime characters and reality band formed by voice actresses
  - Most of the customers are Otaku

Source:https://bang-dream.com/about

Data-Character

- 25 main characters
- Features
  - Band
  - Band Type
    - Band: Exist in reality
    - Vocal: Virtual
  - Part: Vol. Gt. ….
  - AvgDiffExpert
  - OriginalInGame
  - SongInGame
- Popularity
- PTT vote
- CharAge
- CharHeight
- 4Stars(~12/8)
- Attribute(Cool, Powerful, Pure, Happy)
- Hair :RGB, HSL
- Eyes : RGB, HSL

Data-Song

- 51 original songs of 5 main bands
- Features
  - SingleNo
  - SingleTopRanking(oricon)
  - WeeksOnRankList(oricon)
  - EX/SPDifficulty
  - EX/SPNotes
  - SongLengthInGame
  - NotePerSecond

Source:

https://ja.wikipedia.org/wiki/BanG_Dream!%E3%81%AE%E3%83%87%E3%82%A3%E3%82%B9%E3%82%B3%E3%82%B0%E3%83%A9%E3%83%95%E3%82%A3

https://appmedia.jp/bang_dream/696454

Target

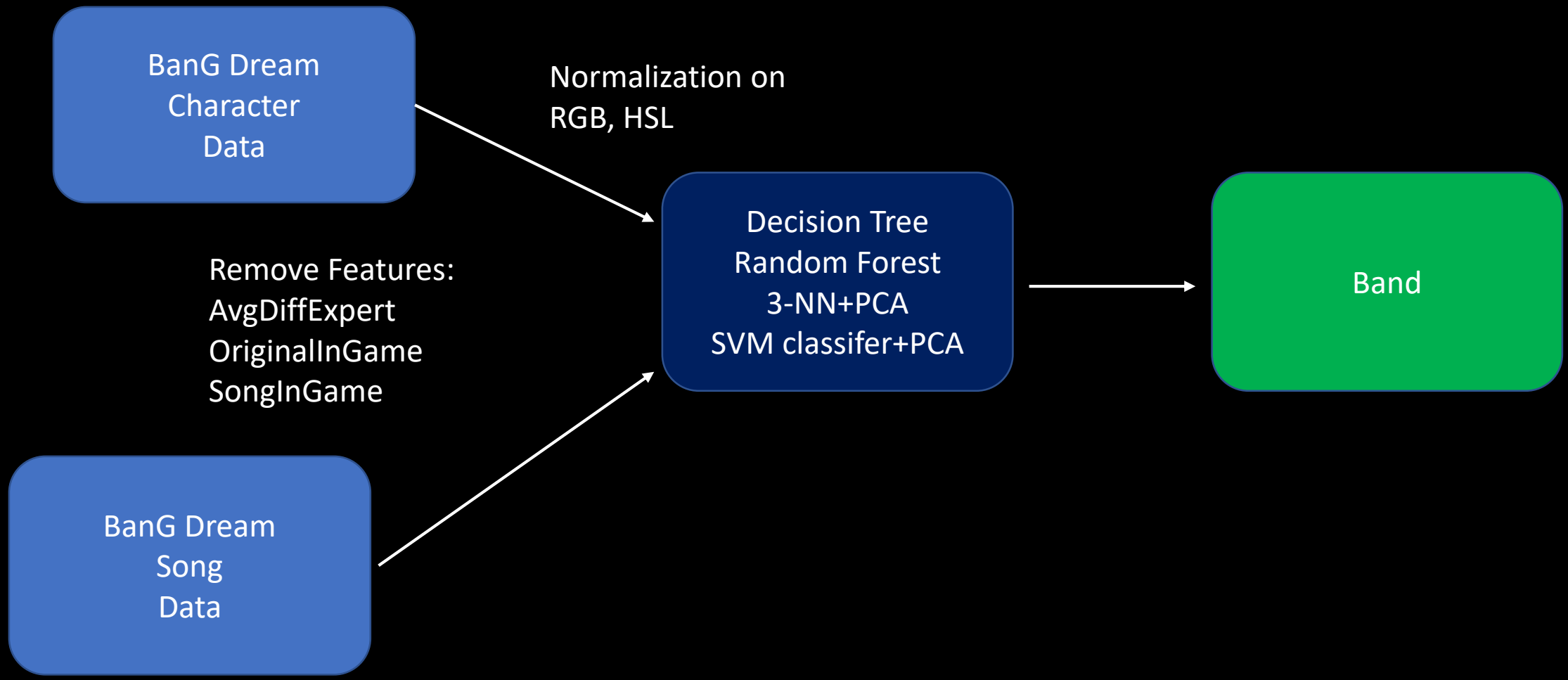
- We would like to predict the **band** that the song or the character belongs to

Why?

- Each band has strong characteristic settings
  - Easier for human to differ each band
  - Collect lovers of different music type
  - Easier for otaku to remember his “waifu”

- However, can machine do the same thing by using objective data?

Experiment


BanG Dream
Character
Data
Normalization on
RGB, HSL
Remove Features:
AvgDiffExpert
OriginalInGame
SongInGame
Decision Tree
Random Forest
3-NN+PCA
SVM classifer+PCA
Band
BanG Dream
Song
Data

Model-setting

We use Sklearn to built all of our models

- Decision Tree
  - Criterion = gini impurity
- Random Forest
  - Criterion = gini impurity
  - Number of trees = 10
- 3-NN
- SVM classifier
  - Kernel = rbf, poly(deg = 3), linear

Training and Testing Dataset

- We use ratio 0.2 to divide training and testing dataset
- Character dataset
  - 20 instances in training dataset
  - 5 testing dataset

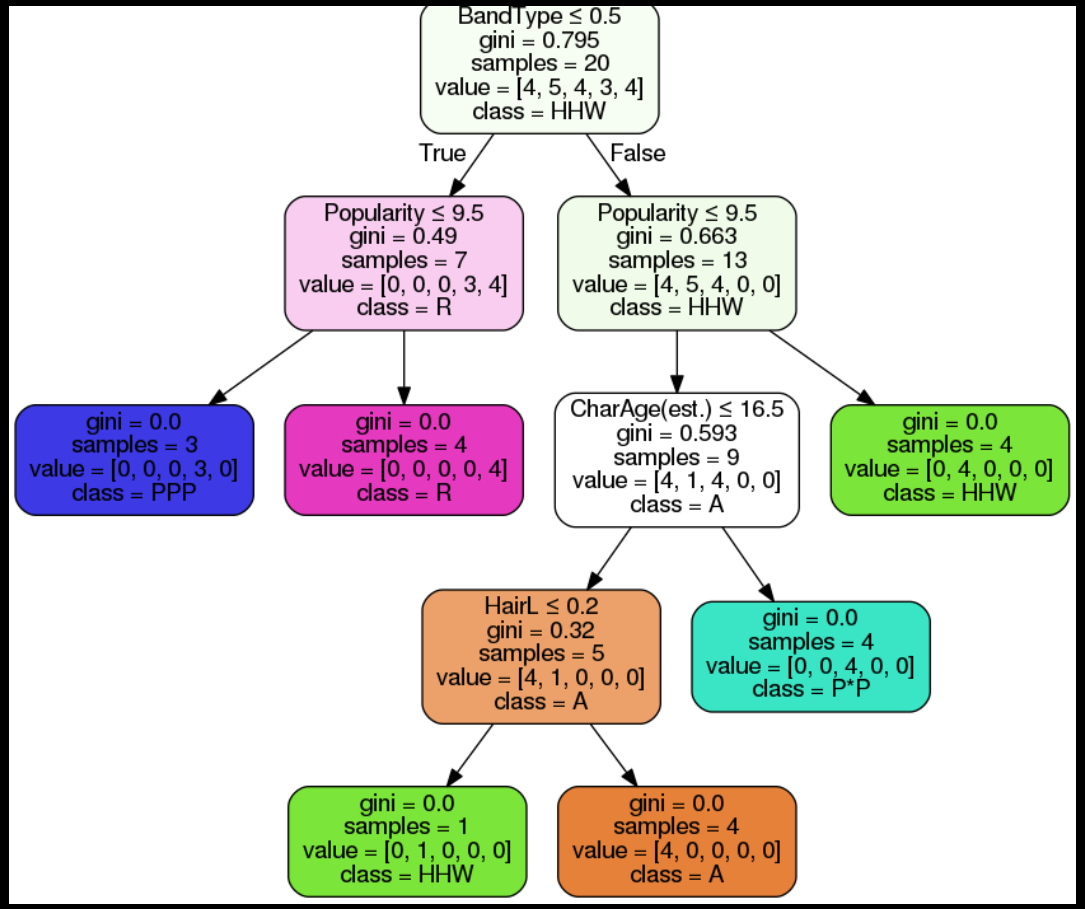
- Song dataset
  - 40 instances in training dataset
  - 11 instances in testing dataset

Result

Model	Decision Tree	Random Forest	3-NN	SVM classifier
Character	**1.0**	0.8	0.2	0.4
Songs	0.18	**0.55**	0.18	0.27

Analysis-Information Based Model

- Decision Tree in character data
  - 100% Accuracy
  - Just Luck
- Band Type and Popularity are important features
- Random Forest may be interfered by unimportant features
- We can differ Afterglow and Pastel*Palettes by age



Analysis-Information Based Model

- Song data
  - Decision Tree
    - Bad performance
    - The result is easily interfered by instances of training dataset
  - Random Forest
    - Best performance
    - The model is good at generalizing

Analysis -3NN

- Character Dataset
  - Each band has 5 members, and there are 5 bands, so the probability of we guess the correct band that the character belongs to is 0.2
  - The result of this model is equal to random
  - Dataset is too small
  - Although we do dimension reduction (PCA) the performance doesn’t improve
- Song Dataset
  - Although we do dimension reduction (PCA) the performance doesn’t improve
  - I think the dataset is also to small for the this model to work

Analysis-SVM Classifier

- We tried different type of kernels
  - RBF has the worst performance
  - Linear and polynomial kernel almost has the same performance
- Applying PCA does slightly improve the performance

Conclusion

- Information based model is strong on small dataset
- kNN model is weak on this task
- Machine has difficulty on classifying which band the song or the character belongs to, although information based models give acceptable performance

Further works

- Normalization on more features
- Reduce features by PCA
- Wait for more data
- Try DNN

#Applying Fourier transform on each song and do the same task (Spectrum)

- RNN or LSTM model adding dense layer

Any Question?

Thanks for listening