

BanG Dream! Characters and Songs Analysis

一、BanG Dream!

BanG Dream! 是由日本武士道公司所開發的二點五次元女子樂團企劃，因為其角色、聲優豐富的個性以及樂團有各自強烈的音樂風格，在動漫圈掀起了一股風潮。而本篇想藉由其同名音樂手機遊戲內的資料，進行分析和預測。

二、Data

在專題中我們收集了 Character 和 Songs 兩種資料集。

1. Characters

在這個資料集我們收集了目前在遊戲內登場的 25 個主要角色以及 27 種不同的特徵：

- Band 角色所屬樂團
- Band Type 所屬樂團是否有在現實進行樂團活動
- Part 角色在樂團中所擔任的角色
- AvgDiffExpert 角色參與歌曲的平均難度
- OriginalInGame 角色參與的原創歌曲數量
- SongInGame 角色參與的歌曲數量
- Popularity 日本和台灣人氣投票結果(加權比例 15.6 : 1)
- PTT vote PTT 上的投票結果
- CharAge 角色年紀
- CharHeight 角色身高
- 4Stars(~12/8) 截止至 2018/12/8 為止遊戲內的四星卡片數量
- Attribute 角色屬性(遊戲內的四種屬性)
- Hair 角色頭髮的 RGB 和 HSL
- Eyes 角色眼睛的 RGB 和 HSL

2. Songs

在這個資料集我們收集了來自 5 個樂團的 51 首原創歌曲和 7 種特徵：

- SingleNo 所屬單曲為該樂團第幾張單曲
- SingleTopRanking 該曲在 Oricon 上的最高排名
- WeeksOnRankList 該曲在 Oricon 上的週榜最高排名
- EX/SPDifficulty 該曲在遊戲內的最高難度
- EX/SPNotes 該曲在最高難度下的音符數量
- SongLengthInGame 該曲在遊戲內的時間長度
- NotePerSecond 該曲平均每秒通過的音符數

三、Target and Model

1. Target

我們希望我們的模型可以預測出該角色或樂曲的所屬樂團。

因為對於人類而言，每個樂團的角色個性鮮明而且樂曲風格強烈，辨別並不是一件太過困難的事情。因此我們好奇是否機器可以用較為客觀的數據進行區分？

2. Preprocessing and Model Settings

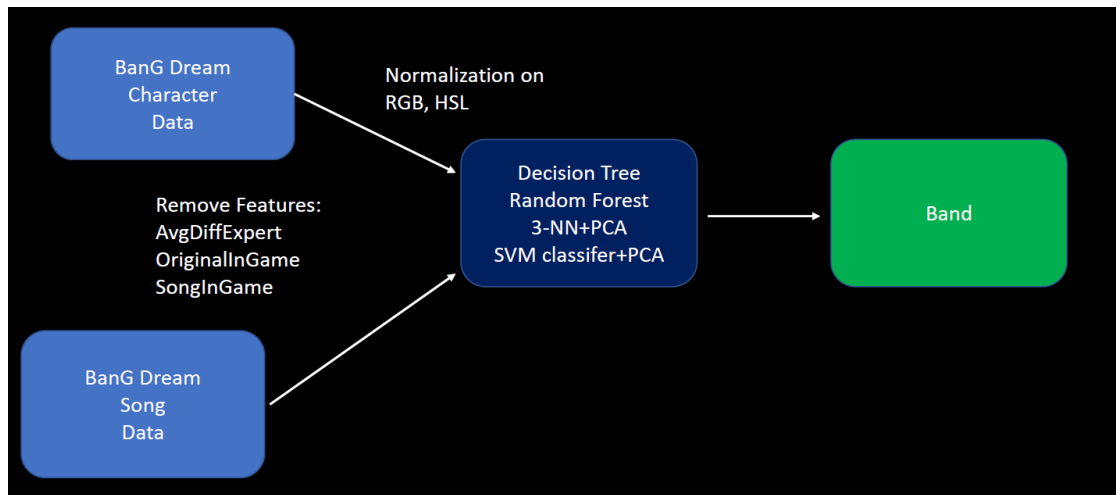


Fig.1 實驗流程

首先，我們必須對資料做一些前處理。我們移除了在 Characters 資料中的 AvgDiffExpert、OriginalInGame、SongInGame 這三種特徵，因為此三種特徵和所屬樂團是完全正相關，而且三種特徵的數值乃因樂團決定，若參考此三種特徵，則屬於作弊。

除此之外，我們也對 RGB 和 HSL 進行了 Range Normalization 使數值落於零到一的區間。

前處理完後，就可以開始設定模型，我們使用 Sklearn 構築所有所需模型，在本篇我們採用了下面四種不同的模型：

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

四、Training and Testing

我們使用 0.2 這個比例來分割訓練資料集和測試資料集，對於 characters 這個資料集來說，則有 20 筆訓練資料 5 筆測試資料；而 songs 這個資料集則

有 40 筆訓練資料和 11 筆測試資料。

五、Results and Analysis

Table.1 Testing 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

1. Analysis-Information Based Model

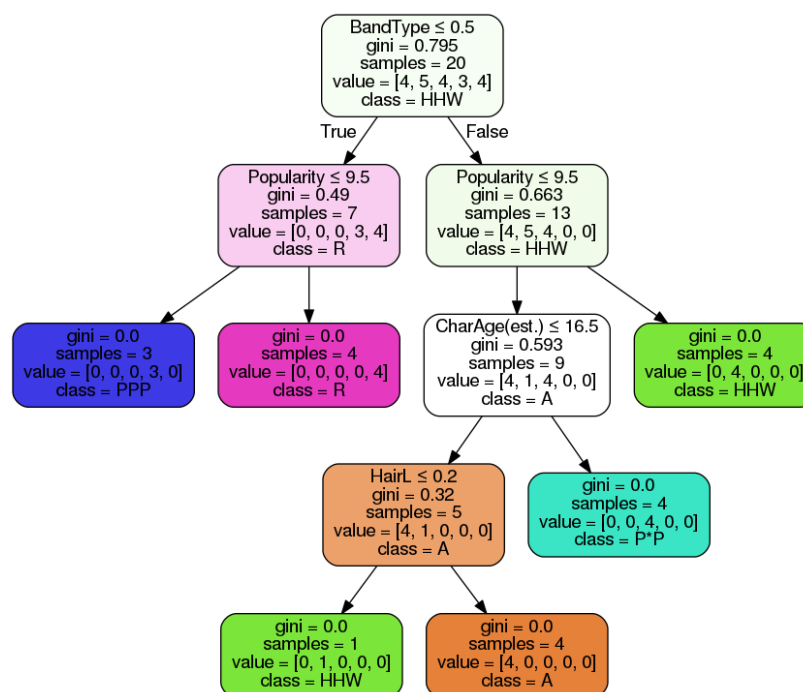


Fig.2 Decision Tree Model

Decision Tree 在 characters 這個資料集中擁有 100%的準確率，其結構如 Fig.2 所示，但這僅僅是巧合罷了，畢竟資料數量過少，我們不能就此定論 Decision Tree 在此情況下是強而有力的模型，而且此模型遇到 songs 此資料集時，表現就不盡人意，只有僅僅 0.18 的準確率。

不過我們也可以從 Fig.2 中得知，Band Type 和 Popularity 是重要的特徵，而且令人驚訝的是，我們可以用角色年齡區分出 Pastel*Palettes 此團體，這是平常較未注意到的特徵。

然而，由數個 Decision Tree 所組成的 Random Forest 模型在 characters 此資料集中略遜 Decision Tree，但在 songs 中卻擁有最高的準確率，證明了此模型有很強的泛化能力。

2. Analysis-3NN

由於每個樂團擁有相同數量的成員，而總共有五個樂團，所以以機率來說隨機猜樂團的正確率應該是 0.2，而此模型在 **characters** 資料集中擁有和隨機一樣的準確率，在 **songs** 也只有可悲的 0.18 準確率。雖然我們有使用 PCA 進行降維，但結果依然沒有太大的改變。

個人認為，此結果可以歸咎於過小的資料集，導致模型無法訓練出好的結果。

3. Analysis-SVM Classifier

我們總共是了三種不同的 kernels：rbf, polynomial 和 linear。而我們發現 rbf 有最差的表現，而 linear 和 polynomial kernel 有相近的表現，同時，我們發現 PCA 可以稍微提升成效，或許和 PCA 是線性降維有些許關聯。

六、Conclusions

我們可以發現 Information based model 在規模小的資料集下依然有很出色的表現，相反地，kNN 則擁有很慘不忍睹的表現，但綜合所有的模型而言，其實機器從上述特徵去辨認角色或歌曲所屬的樂團依然有其難處，即使 information based model 給出了可接受的結果。

七、Reference

<https://bang-dream.com/about>

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

八、Contribution Table

成員	夏軒安	呂明哲	張皓鈞
貢獻度	70%	25%	5%