

知能情報実験 III（データマイニング班）  
Spotify API を用いた楽曲判別

215730G 嶋田 圭吾  
215741B 許田 雅怜

2023 年 6 月 17 日

## 目次

|     |                                       |   |
|-----|---------------------------------------|---|
| 1   | はじめに                                  | 0 |
| 1.1 | 実験の目的と達成目標 . . . . .                  | 0 |
| 1.2 | テーマ「Spotify API を用いた楽曲判別」とは . . . . . | 0 |
| 2   | 実験方法                                  | 0 |
| 2.1 | 実験目的 . . . . .                        | 0 |
| 2.2 | データセット構築 . . . . .                    | 1 |
| 2.3 | モデル選定 . . . . .                       | 3 |
| 2.4 | パラメータ調整 . . . . .                     | 3 |
| 3   | 実験結果                                  | 3 |
| 4   | 考察                                    | 4 |
| 5   | 意図していた実験計画との違い                        | 4 |
| 6   | まとめ                                   | 4 |

## 概要

概要本文書は知能情報実験 III（データマイニング班）のグループ 3 の開発テーマである「Spotify API を用いた楽曲判別」についての中間報告レポートである。

## 1 はじめに

### 1.1 実験の目的と達成目標

知能情報実験 III は、情報工学分野のより専門的な知識を理解・習得することを目的として、半年間でシステムの開発やデータ解析等に取り組む実施される。その中の一つデータマイニング班においては機械学習外観ならびにその応用を通し、対象問題への理解、特徴量抽出等の前処理、バージョン管理やデバッグ・テスト等を含む仕様が定まっていない状況下における開発方法、コード解説や実験再現のためのドキュメント作成等の習得を目指す。

### 1.2 テーマ「Spotify API を用いた楽曲判別」とは

TikTok ではテンポが良い曲に合わせて踊ったり、耳に残ったり雰囲気のある曲を使って TikTok ユーザーの耳に残るような選曲がされている。本グループではそんな Tiktok で任意に選んだ楽曲が TikTok の投稿者に使われるかそうでないか分類することを対象問題として設定した。クラス分類とは [2] によると、様々な対象をある決まったカテゴリーまたはクラスのいずれかにデータを割り当てることである。任意に選んだ曲を TikTok で使われる曲かそうでないかを判別することで、TikTok 投稿者の再生回数に寄与することができると考える。また [3] によると教師あり学習の一種とも述べられており、カテゴリ（離散値）を予測することが目的である。

## 2 実験方法

### 2.1 実験目的

1. 与えた楽曲が将来的に Tiktok で使われる曲かどうかを判別をする
2. Tiktok で実際に使われている曲を与えた際に、Tiktok で使われる曲と判別できるのかどうかの検証する

上記の二つを通して、TikTok に使われている曲、選ばれている曲にはこういった共通点があるのかを明らかにすることが目的である。

## 2.2 データセット構築

Spotify API を用いてデータセットを構築した。具体的には、Tiktok でよく使われる曲が含まれた複数のプレイリストから曲の情報を取得し、csv ファイルに書き出すプログラムを Python で作成した。このプログラムを実行して Tiktok でよく使われる曲のデータセット「tiktok\_data.csv」を得た。

また、Tiktok で使われていない楽曲の情報は、「jpop」で検索して出てきた曲の中から「tiktok\_data.csv」に含まれる曲を除外して取得することで、Tiktok で使われない曲のデータセット「not\_tiktok\_data.csv」を得た。

取得した情報の詳細は以下の表 1 の通りである。

表 1 Audio Features から使用するデータ一覧

| データ名             | 詳細  |
|------------------|---|
| acousticness     | 曲の音響的な特徴を表す指標。0.0（音響的でない）から 1.0（音響的）の範囲で表されます。                |
| danceability     | 曲がダンスに適しているかどうかを表す指標。0.0（ダンスに適さない）から 1.0（ダンスに適している）の範囲で表されます  |
| energy           | 曲のエネルギーの強さを表す指標。0.0（エネルギーが低い）から 1.0（エネルギーが高い）の範囲で表されます。       |
| instrumentalness | 曲がインストゥルメンタルかどうかを表す指標。0.0（ボーカル曲）から 1.0（インストゥルメンタル曲）の範囲で表されます。 |
| key              | 曲のキーを表す数値。0 から 11 までの整数で表されます。                                |
| liveness         | 曲の演奏がライブ録音かどうかを表す指標。0.0（スタジオ録音）から 1.0（ライブ録音）の範囲で表されます。        |
| loudness         | 曲の音量を表す指標。-60 から 0 の範囲で表されます。                                 |
| mode             | 曲のモード（メジャーかマイナーか）を表す数値。0（マイナー）または 1（メジャー）で表されます。              |
| speechiness      | 曲の音声の存在を表す指標。0.0（音声が少ない）から 1.0（音声が多い）の範囲で表されます。               |
| tempo            | 曲のテンポ（BPM）を表す数値。  |
| time_signature   | 曲の拍子記号を表す整数。  |
| duration_ms      | 曲の再生時間をミリ秒単位で表す整数値。   |
| valence          | 曲のポジティブさまたはネガティブさを表す指標。0.0（ネガティブ）から 1.0（ポジティブ）の範囲で表されます。      |

表 1 の中から二つの特徴量を使用して分類を行う。集めたデータを探索し、特徴量の分布やクラス間の差異を確認し図を作成して、特徴量の特性を理解する。データの 70~80% をトレーニングセットに、残りの 20~30% をテストセットに割り当て、トレーニングセットを使用してモデルを学習していく。

## 2.3 モデル選定

今回はラベル付きデータのクラス分類を行うので、scikit-learn algorithm cheat-sheet で調べた結果、Linear SVC の中の KNeighbors Classifier(K 近傍法) を用いることにした。

## 2.4 パラメータ調整

データセットのサイズや特性、クラスのバランスなどに応じて最適な K の値は異なる場合がある。そのため異なる K の値 (1~10 を想定) を試し、最良の結果を示す K の値を選択する。

# 3 実験結果

事実として得られた結果を示そう。なお、以下の点に留意すること。

- 「思う」「思われる」のような主観ではなく、客観的事実を述べること。
- 図表には適切なキャプションを付けること。
- 挿入した図表について、本文中でその読み方を述べること。その際には label, ref により相互参照すること。
- レポートにおけるグラフの作成においては、以下の点に注意する。
  - 軸目盛および軸ラベルに関する注意事項
    - \* 必ず軸ラベルを表示する
    - \* 軸に単位がある場合には、ラベルに単位を付記する
    - \* 軸目盛は適切な感覚で表示する
    - \* 軸目盛は述べていたい内容に応じて線形スケールと log スケールを使い分ける
    - \* 印刷時に明瞭に読むことができるサイズで表示する
  - 線・点・ポイントおよび凡例に関する注意事項
    - \* 線・点・ポイントは、印刷時に明瞭に識別できる太さやサイズで表示する
    - \* 1つのグラフに複数にデータを表示する際には、データごとに異なる線種、線の太さ、ポイント形状などを使用する
    - \* モノクロ印刷でも識別できるように線・点・ポイントを使用することが望ましい
    - \* 凡例は線・点・ポイントに重ならないように注意する

## 4 考察

実験課題への取り組みを通し、実験の意義、実験からわかったこと、今後の展望などを述べる。失敗やつまづきがあれば、それらについての失敗分析を含めると良い。

## 5 意図していた実験計画との違い

グループワークとして2ヶ月程度の時間が用意されていた。ガントチャート??等、何かしら工夫して全体の計画を述べよう。これらの期間をどのように使おうとし、実際どうだったのかについて自己評価（振り返り）してみよう。大きなズレがある場合それは何故起きたのか、どうやればそのギャップを縮められそうか検討してみよう。

## 6 まとめ

データマイニング班の達成目標を振り返り、選んだテーマに対する機械学習の適用を通して得られた知見や学んだことをまとめよう。また今後やるべきことや後進に伝えたいこと等あれば自由に述べよう。

## 参考文献

- [1] レポート作成の手引き レポートの基本的形式に関するガイド, <https://www.kanazawa-u.ac.jp/wp-content/uploads/2015/01/tebiki2.pdf>, 2020/07/02.
- [2] クラス分類 AI 研, <https://ai-kenkyujo.com/term/classification/>.
- [3] キカガク, [https://free.kikagaku.ai/tutorial/basic\\_of\\_machine\\_learning/learn/machine\\_learning\\_classification/](https://free.kikagaku.ai/tutorial/basic_of_machine_learning/learn/machine_learning_classification/).
- [4] テーマ出典 2, 出典は半角, . で書こう.
- [5] ガントチャート, <https://ja.wikipedia.org/wiki/ガントチャート>, 2020/07/02.