

知能情報実験 III（データマイニング班）
K 近傍法を用いた楽曲判別

215730G 嶋田 圭吾
215741B 許田 雅怜

2023 年 7 月 30 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「K 近傍法を用いた楽曲判別」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 | 2 |
| 2.4 | パラメータ調整 | 2 |
| 2.5 | 説明変数の選定 | 2 |
| 3 | 実験結果 | 3 |
| 3.1 | 説明変数の組み合わせとパラメータの調整 | 3 |
| 3.2 | モデルの生成 | 5 |
| 3.3 | 曲の判別 | 5 |
| 4 | 考察 | 7 |
| 5 | 意図していた実験計画との違い | 7 |
| 6 | まとめ | 8 |

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「K 近傍法を用いた楽曲判別」についての最終報告レポートである。本実験は、Spotify API から作成したデータセットを K 近傍法を用いて学習させ、得たモデルをもとにクラス分類（楽曲判別）を行うというものである。

1 はじめに

1.1 実験の目的と達成目標

本実験は、機械学習のより専門的かつ実践的な知識を理解・習得することを目的として、半年間でデータセットの構築や K 近傍法による機械学習、モデルを用いたクラス分類等に取り組むために実施される。

本グループにおいては本実験を通し、機械学習（K 近傍法）への理解、特徴量抽出等の前処理、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「K 近傍法を用いた楽曲判別」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 [1] 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 で流行するであろう楽曲が判り、次のトレンドを知ることができる。

2 実験方法

2.1 実験目的

1. データセットから Tiktok で使われる曲とそうでない曲の特徴を捉える。
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
3. 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

上記の三つを通して、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 の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、そこから「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。

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

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、[2] で調べた結果、Linear SVC 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

ここで言うパラメータとは K 近傍法における k のことである。K 近傍法は k の値によって精度が変わるため、最適な k の値を設定する必要がある。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 (1~30 を想定) を試し、最良の結果を示す k の値を選択する。

2.5 説明変数の選定

Audio Features から取得した説明変数 (特徴量) の中から、使用する変数を 2 つに絞るために様々な組み合わせで学習させ、一番良い精度を出した組み合わせを調べていく。一番精度の良い説明変数の組み合わせで実際の予測を行なっていく。2 つに絞る理由は、Tiktok に使われる曲とそうではない曲とを区別するための特徴をはっきりとさせるためである。

データセット作成の際、Audio Features から 13 種類の説明変数を取得したが、これら変数を全て用いると学習させる量がとても多いため、区別するための説明変数として必要ない理由が明確である変数をあらかじめ除外した。

除外した説明変数と実験で用いない理由は下記の通りである。

表 2 実験で用いない説明変数とその用いない理由

| 実験から外す説明変数 | 説明変数の説明・外す理由 |
|------------------------|---|
| ライブ録音か (liveness) | Tiktok で使われる曲はおおむねスタジオ録音であるから。 |
| 音量 (loudness) | 聞く人や投稿者の編集によって音量は調整できるから。 |
| 音声の存在 (speechiness) | 投稿者によってサビを使うのか、イントロを使うのか決められるから。 |
| 拍子記号 (time_ signature) | ほとんどの曲が 4 拍子であるから。 |
| 再生時間 (duration_ ms) | Tiktok の動画時間は投稿者が決めれるので、曲の再生時間に依存しないから。 |

3 実験結果

3.1 説明変数の組み合わせとパラメータの調整

データの 80% をトレーニングセットに、残りの 20% 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。選定された 8 つの説明変数から 2 つを選ぶ組み合わせ全てについて、K 近傍法を用いてそれぞれの組み合わせにおける最適な k の値と評価値を算出した。その結果を「scores.csv」ファイルに保存した。

内容は以下の通りである。

表 3 scores.csv

| Feature1 | Feature2 | Best_k | Best_Score |
|-------------|-------------|--------|--------------------|
| 音響的な特徴 | ダンスに適しているか | 12 | 0.6489795918367347 |
| 音響的な特徴 | エネルギーの強さ | 1 | 0.5877551020408164 |
| 音響的な特徴 | インストゥルメンタルか | 23 | 0.5469387755102041 |
| 音響的な特徴 | キー | 1 | 0.563265306122449 |
| 音響的な特徴 | モード | 22 | 0.5387755102040817 |
| 音響的な特徴 | テンポ (BPM) | 4 | 0.5510204081632653 |
| 音響的な特徴 | ポジティブさ | 5 | 0.563265306122449 |
| ダンスに適しているか | エネルギーの強さ | 10 | 0.7020408163265306 |
| ダンスに適しているか | インストゥルメンタルか | 30 | 0.6612244897959184 |
| ダンスに適しているか | キー | 30 | 0.6816326530612244 |
| ダンスに適しているか | モード | 18 | 0.636734693877551 |
| ダンスに適しているか | テンポ (BPM) | 27 | 0.6530612244897959 |
| ダンスに適しているか | ポジティブさ | 22 | 0.6571428571428571 |
| エネルギーの強さ | インストゥルメンタルか | 26 | 0.563265306122449 |
| エネルギーの強さ | キー | 16 | 0.5346938775510204 |
| エネルギーの強さ | モード | 30 | 0.5755102040816327 |
| エネルギーの強さ | テンポ (BPM) | 23 | 0.6 |
| エネルギーの強さ | ポジティブさ | 21 | 0.5714285714285714 |
| インストゥルメンタルか | キー | 17 | 0.5551020408163265 |
| インストゥルメンタルか | モード | 12 | 0.5469387755102041 |
| インストゥルメンタルか | テンポ (BPM) | 4 | 0.5673469387755102 |
| インストゥルメンタルか | ポジティブさ | 11 | 0.563265306122449 |
| キー | モード | 19 | 0.5387755102040817 |
| キー | テンポ (BPM) | 23 | 0.5428571428571428 |
| キー | ポジティブさ | 23 | 0.5714285714285714 |
| モード | テンポ (BPM) | 3 | 0.5183673469387755 |
| モード | ポジティブさ | 5 | 0.5877551020408164 |
| テンポ (BPM) | ポジティブさ | 30 | 0.5673469387755102 |

この結果から、「ダンスに適しているか」と「エネルギーの強さ」の組み合わせで $k = 10$ の時、評価値が約 0.7020 となり、この値が全ての組み合わせの中の最大値であった。

「ダンスに適しているか」と「エネルギーの強さ」の組み合わせの時の $k = 1 \sim 30$ の値は、以下のグラフの通りである。

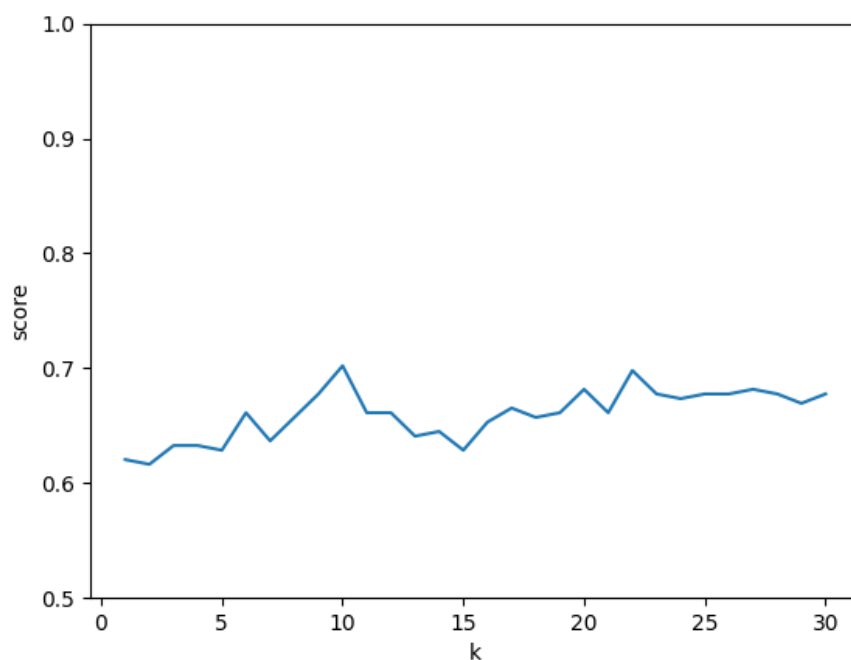


図 1 $k = 1 \sim 30$ の値

3.2 モデルの生成

説明変数の組み合わせとパラメータの調整により求められた、最適な組み合わせと k の値で再度 K 近傍法による機械学習を行い、モデルを学習させる。学習したモデルは Python のライブラリである joblib で「model.joblib」に保存する。

3.3 曲の判別

生成したモデルを用い曲の判別を行う。説明変数の組み合わせとパラメータの調整から正答率は約 70% であると考えられる。今回は TikTok で使われている曲と使われていない曲をそれぞれ 10 曲ずつ用意し、実際にはどの程度の正答率で判別できているかを確認する。

判別結果は以下の通りである。

表 4 判別結果

| 曲名 | 曲の ID | 0 or 1 | 判別結果 |
|------------------|------------------------|--------|------|
| トウキョウ・シャンディ・ランデヴ | 2MmgGuQzr0kkiA04lMdPXW | 1 | 1 |
| 丸の内サディスティック | 1mWN1USPKg632qQXG9sfDL | 1 | 0 |
| オトナブルー | 4fMewVWtew1oB6d3E8x9pp | 1 | 1 |
| 酔いどれ知らず | 4oPCfEDVIKnPJ6UPkhJPCa | 1 | 1 |
| すーぱーぬこになれんかった | 16UIKx3sHzQDwCFd7KxkzT | 1 | 1 |
| HAPPY BIRTHDAY | 4FMz2RFrbDGzJO7K4D0vS3 | 1 | 0 |
| アイドル | 7ovUcF5uHTBRzUpB6ZOmt | 1 | 1 |
| ダンスホール | 4NaaF28BeO9WzjDrSS71Nz | 1 | 1 |
| 可愛くてごめん | 3K9RzumH1VCDgIZuwuT9bq | 1 | 0 |
| エジソン | 223e30Nk8UY0HKezFtakUS | 1 | 1 |
| ドライフラワー | 7dH0dpi751EoguDDg3xx6J | 0 | 0 |
| 夜に駆ける | 6MCjmGYlw6mQVWRFVgBRvB | 0 | 0 |
| 心予報 | 4KcTqk3YtbJNVhvbhBDVw2 | 0 | 0 |
| 不可幸力 | 18nkY3pJTub8WwEGiQAGh4 | 0 | 1 |
| 怪獣の花唄 | 4W29WmnGN1FlufURaBt8yw | 0 | 0 |
| 脳裏上のクラッカー | 3FRU8JI3EK9G7hXSP3P4R8 | 0 | 0 |
| カメレオン | 0J0P8iSE6l3pYHbUdNFYdS | 0 | 0 |
| Cinderella Boy | 4UALTBWz0OdQcjbTsjFceV | 0 | 0 |
| ただ君に晴れ | 3wJHCry960drNlAUGrJLmz | 0 | 1 |
| まちがいさがし | 4Jj31lihp5Nvdi4FRjrqc | 0 | 0 |

正答率はそれぞれ、Tiktok で使われる曲が 70%、使われない曲が 80% という結果になった。

4 考察

本実験の意義として、機械学習の前処理である特徴量抽出やデータの標準化等、パラメータの調整などを通して機械学習に対しての理解を深めることが挙げられる。

今回の実験はこれから Tiktok で使われる可能性の高い曲をモデルから判別することで、判別結果から次のトレンドをいち早く掴むことができると考え実施した。今後の展望として、音響的な特徴量だけではなく再生数・高評価数なども評価の際に加えると、より高い精度で判別ができると考えられる。

Tiktok で使われる曲かどうかを左右する説明変数について、実験前の段階で Tiktok で使われる曲の特徴として踊りやすい・テンポがいい・明るいなどが挙げられたので、「ダンスに適しているか」、「テンポ (BPM)」、「ポジティブさ」、「エネルギーの強さ」が重要な説明変数であると予想された。実際に最適な組み合わせは「ダンスに適しているか」と「エネルギーの強さ」であったので概ね予想通りであった。

5 意図していた実験計画との違い

本実験の進行は以下の通りである。

| No. | 内容 | 予定 | 期間(～週目) | | | | | | | | | |
|-----|------------|-------|---------|---|---|---|---|----|----|----|----|----|
| | | | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
| 1 | テーマ決め | 5～7 | | | | | | | | | | |
| 2 | データセット作成 | 7～9 | | | | | | | | | | |
| 3 | K近傍法実装 | 8～10 | | | | | | | | | | |
| 4 | パラメータ調整 | 9～11 | | | | | | | | | | |
| 5 | 判別プログラム実装 | 10～12 | | | | | | | | | | |
| 6 | 学習済みモデルの保存 | 11 | | | | | | | | | | |
| 7 | 動作確認 | 12 | | | | | | | | | | |
| 8 | 最終レポート作成 | 13、14 | | | | | | | | | | |
| 9 | 発表資料作成 | 13、14 | | | | | | | | | | |

図2 本実験の進行（ガントチャート）

図2 のオレンジに塗られた部分は予定を超過したことを表している。

私たちのグループは早々にテーマを「K 近傍法を用いた楽曲判別」に決めることができたが、その後の Spotify API を用いたデータセット作成や K 近傍法の実装、動作確認などに当初の想定よりも時間がかかってしまった。これは授業時間外での調べが少し足りなかったからと考えられる。自分にとって新しいことには時間的余裕をしっかりと持って挑戦するべきであると反省した。

後半の動作確認や最終レポート作成、発表資料作成はしっかりと役割分担をした上でスムーズに進めることができ良かった。

6 まとめ

本実験では Linear SVC の中の KNeighbors Classifier(K 近傍法) を用いた楽曲分類・判別を行った。本実験により機械学習に対する理解が深まるとともに、正確にラベル付けがされたデータを K 近傍法を用いて機械学習させた際の精度がかなり高いことがわかった。

参考文献

- [1] クラス分類 AI 研, <https://ai-kenkyujo.com/term/classification/>, 2023/07/27.
- [2] scikit-learn algorithm cheat-sheet, https://scikit-learn.org/stable/tutorial/machine_learning_map/index.html, 2023/07/27.