

# ソーシャルメディアの特徴量を考慮した 機械学習による音楽トレンドの検知

佐藤 豪栄<sup>†</sup> 関岡 冨俊<sup>‡</sup> 秦野 亮<sup>‡</sup> 西山 裕之<sup>‡</sup>

東京理科大学 創域理工学部 経営システム工学科<sup>†</sup>

東京理科大学大学院 創域理工学研究科 経営システム工学専攻<sup>‡</sup>

## 1 序論

近年、音楽の消費方法としてストリーミングサービスやオンラインプラットフォームを活用したものが主流になり、デジタルメディアを通じた露出、Twitter や YouTube などのプラットフォームでの話題性も成功の重要な要素となっている。また、TikTok や YouTube Shorts をはじめとしたショート動画プラットフォームがソーシャルメディア上で拡大している。実際、「TikTok」は、近年急激にユーザー数を増やしており、日本では1700万人以上、世界では約10億人以上のユーザー数を誇る。

そこで、本研究は音楽産業におけるソーシャルメディアの役割と特に TikTok などショート動画プラットフォームの影響を調査する。特に、ショート動画プラットフォームのデータを利用することで、音楽トレンドの変化の予測を機械学習を用いてより正確にできるのではないかと、いう仮説について、比較実験を実施する。

## 2 関連研究

Alberto[1] は、ソーシャルメディアと新聞データから音楽アルバムのランキングを予測し、学習器としてランダムフォレストを用いて高い予測精度を達成した。また、Araujo[2] は、ポジティブ・ツイート（アルバム・リリースの30日前）の量は、単純な線形モデルで音楽ランキングの変動の95.5%を説明できると発表し

た。さらに、ショート動画プラットフォームに関する研究として、Eliza[3] は、TikTok が国際市場での認知度、トラフィック、売上を増加させることができると明らかにした。

したがって、本研究では、ショート動画プラットフォームから得られる情報が、音楽ランキングの予測にどの程度重要な特徴量となるのか考察することを目的とする。

## 3 提案手法

本研究のデータセットは、Billboard データセット、YouTube データセット、TikTok データセットから作成される。2023年11月29日に「Billboard Japan charts hot 100」にランクインした楽曲を対象にデータを収集した。また、各データの作成方法と特徴量名は以下の通りである。

**Billboard データセット：** 2023年11月1日から2023年12月20日までの期間の Billboard Japan のランキング (only\_ranking) に関する情報をスクレイピングにより収集した。

**YouTube データセット：** YouTube API を利用して、2023年12月17日にMVの再生数、いいね数、コメント数、チャンネル登録者数など (basic\_info) を収集した。また、YouTube Charts から、2023年11月1日から2023年12月17日までのランキングデータ (youtube\_ranking, short) を収集した。

**TikTok データセット：** TikTok アプリを利用して、楽曲の投稿数、アーティストやタイ

Music trends detection using machine learning based on social media features

<sup>†‡</sup> Goei Sato, Satoshi Sekioka, Ryo Hatano, and Hiroyuki Nishiyama, Tokyo University of Science

表1 実験結果 (抜粋)

特徴量	正解率
baseline	0.715
baseline_short	0.684
baseline.tiktok	0.675
baseline.tiktok_short	0.695
basic_info	0.302
only_ranking	0.756
only_ranking_basic	0.745
only_ranking_basic.tiktok	0.755
only_ranking.tiktok	0.756
youtube_ranking	0.414

トルに関するハッシュタグの検索ヒット数、フォロワー数など (tiktok) を収集した。また、「Billboard Japan TikTok Weekly Top 20」から TikTok ランキングデータをスクレイピングにより収集した。こちらは tiktok 特徴量に含めた。

本研究では、2023年12月20日の「Billboard Japan hot 100」のランキングを目的変数とし、分類器としてランダムフォレストを利用した。評価指標には正解率を使用し、特徴量の重要度も可視化した。Billboard データセット内のランキングデータの特徴量 (only\_ranking)、YouTube データセット内のランキングデータ以外の特徴量 (basic\_info)、YouTube データセット内のランキング特徴量のうち、ショート動画ランキングを除いた特徴量 (youtube\_ranking) を特徴量としたものをベースラインとし、他の特徴量との組み合わせから、正解率を比較する。

#### 4 実験・結果・考察

実験結果を一部抜粋し、表1に示す。結果より、ベースラインを超えた組み合わせは、5つあった。only\_ranking 特徴量と tiktok 特徴量を考慮した場合の正解率が高い傾向にあった。一方で、ベースラインにその他の特徴量を加えた場合は、正解率が下がった。

これらの結果から、TikTok データは音楽ラン

キングの予測に影響を与える可能性があると言える。また、ベースラインにその他の特徴量を加えた時に正解率が上がらなかった要因として、youtube\_ranking の影響が考えられる。

さらに、重要度の可視化から、TikTok でタグ検索をした際のヒット数とタイトルのタグが影響をしていることがわかった。また、その他の重要度として、アーティストのハッシュタグの影響もあったことから、アーティストの人気のランキングに影響していることも考えられる。

#### 5 結論

本研究では、ショート動画プラットフォームから得られた特徴量をもとに、音楽ランキングを予測を機械学習を使用して行なった。結果として、TikTok データは音楽ランキングの予測に重要な特徴量となる可能性を示した。

今後の展望として、さらに多様なデータソースの統合や、長期間のデータ収集を行うことや時系列を考慮した学習器を使用することが挙げられる。

#### 参考文献

- [1] Alberto Cosimato, Roberto De Prisco, Alfonso Guarino, Delfina Malandrino, Nicola Lettieri, Giuseppe Sorrentino, and Rocco Zaccagnino. The conundrum of success in music: Playing it or talking about it? *IEEE Access*, 7:123289–123298, 2019.
- [2] Carlos V.S. Araujo, Rayol M. Neto, Fabiola G. Nakamura, and Eduardo F. Nakamura. Predicting music success based on users' comments on online social networks. In *Proceedings of the 23rd Brazillian Symposium on Multimedia and the Web, WebMedia '17*, pages 149–156, New York, NY, USA, 2017. Association for Computing Machinery.
- [3] Eliza Jennings, Fernando Pinto Santos, and Ana Paula Marques. Does your business TikTok? genuineness, proximity to customers and international expansion with short video marketing. In *International Conference on Marketing and Technologies*, pages 283–291. Springer, 2022.