

メタ情報と特徴量の関係把握のための楽曲群の可視化

渡辺 みどり[†] 黒子 なるみ^{†*} 大矢 隼士[‡] 伊藤 貴之[‡]お茶の水女子大学[†] Septeni Japan 株式会社[‡]

*現在株式会社 NTT ドコモ所属

1. はじめに

音楽を聴く手段として定額のストリーミングサービスを利用する人が近年増加している[1]. サービス上での曲の選択手段の多様化にともない, 音楽の自動分類や推薦などの研究が活性化している. 音楽を分類するための指標には, アーティスト名や年代といったメタ情報や, テンポや音量のような音響特徴量など, 多様なものが考えられる. これらの指標のうちどれを強く採用するかによって自動分類の結果は変化するが, その最適化は非常に複雑な問題である. さらに, 自動分類がどのように成功もしくは失敗しているかを観察することも課題となっている.

本研究では, 楽曲分類に効果的なメタ情報, 音響特徴量, 機械学習手法, 可視化手法を包括的に検討すること, メタ情報と音響特徴量の関係を可視化によって見出せるかを実証することを目的とし, 楽曲群を可視化する. 本報告では, 機械学習で音響特徴量を算出し, 年代やアーティストのメタ情報と合わせて可視化した.

2. 先行研究

Cheuk ら[2]は, トリプレットニューラルネットワーク(TNN)を用いた回帰ベースの音楽感情予測手法を提案した. 楽曲群の感情データセットに主成分分析, オートエンコーダー, TNN を適用した次元削減により, TNN が感情回帰タスクに効果的であることを示した. このように, 音楽分類等における適切な特徴量や次元削減手法の検討や新しいアプローチは多数報告されているが, それらを包括的に調査する研究は少ない.

Lionello ら[3]は, 34 次元の特徴量を付与した楽曲群に Parametric t-SNE(t-distributed Stochastic Neighbor Embedding)を適用し, 可視化した. 高次元データに対して Parametric t-SNE が主成分分析よりも有用であることを示したが, 具体的に用いた音響特徴量に関しては言及されていない.

Visualization of Song Collections for Understanding of the Relationship between Metadata and Features

[†] Midori Watanabe Narumi Kuroko Takayuki Itoh

[‡] Hayato Ohya

[†] Ochanomizu University

[‡] SEPTENI CO.,LTD.

3. 提案手法

図1に本手法の流れを示す.

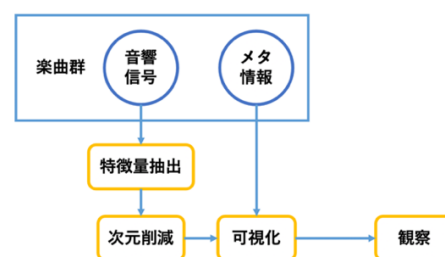


図1 提案手法の概要

3.1 メタ情報

各楽曲に対してタグとして付随させる情報をメタ情報と呼ぶ. 本報告では年代とアーティストをメタ情報として可視化した結果を紹介する.

3.2 特徴量抽出

音響信号として録音された楽曲データから, ツールや機械学習を用いて数値化された音響特徴量を算出する. 本研究では機械学習ツールである RP_extract[4]を用いて算出した各楽曲の特徴量を利用する. RP_extract は, リズムパターン(RP), リズムヒストグラム(RH), 統計スペクトル記述子(SSD)の 3 種類の音響特徴量を算出する. 本報告では, 各クリティカルバンドのビートやエネルギーの発生を統計的に測定して記述する SSD を利用した結果を紹介する.

3.3 次元削減/可視化

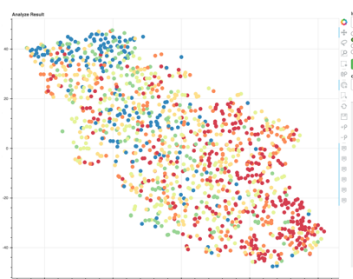
3.2節で算出した音響特徴量に次元削減を適用して2次元空間に写像し, 散布図として楽曲の分布を表示する. メタ情報を基準として散布図上の点群を色分けすることで, メタ情報と音響特徴量の関係を把握できるようにする. 本報告では, t-SNEで次元削減した結果を紹介する.

4. 実行例と考察

1986年から2018年までの日本のヒット曲1315曲からなる楽曲群に提案手法を実行した. RP_extractで算出した音響特徴量SSDをt-SNEで次元削減し, 散布図として表示した. 年代で色分けした散布図とアーティストで色分けした散布図を, それぞれ図2,3に示す. アーティストによ

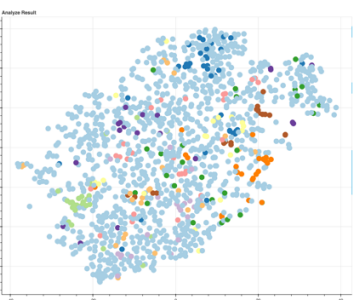
る色分けでは、楽曲群のうち、18曲以上が存在する10のアーティストについて色をつけ、残りをその他としてまとめて可視化した。

年代で色分けした可視化結果は、年代の古い楽曲（青）から新しい楽曲（赤）へのグラデーションに近い構成をしていることがわかる。アーティストで色分けした可視化結果では、Mr.Children(黄緑)に特に明確なクラスタが見られる。その他、ZARD(茶)、AKB48(青)、EXILE(オレンジ)などの一部の楽曲もクラスタを形成していることが見受けられる。これによって、1986年から2018年までの楽曲間と、アーティストごとの楽曲間では、ビートやエネルギーにおいて特徴の違いがあることがわかった。



■	1986年～1991年	■	2002年～2007年
■	1992年～1996年	■	2008年～2012年
■	1997年～2001年	■	2013年～2018年

図2 年代で色分けした結果



■	AKB48	■	工藤静香
■	Mr.Children	■	ゆず
■	安室奈美恵	■	BUMP OF CHICKEN
■	SMAP	■	ZARD
■	浜崎あゆみ	■	その他
■	EXILE	■	

図3 アーティストで色分けした結果

さらに、具体的にどのような違いがあるかを確認するため、Librosa[5]を用いて音響特徴量を算出し、個別に可視化した。その結果、年代では rms energy(音量の二乗平均平方根値)に、アーティストでは spectral flatness(周波数別の音量の均一さ)に特徴があることを発見した。各楽曲の rms

energy の平均値は、概ね年代が上がるほど大きくなった。音量の最大値は過去の楽曲も最近の楽曲もほぼ同一であるため、過去の楽曲ほど音量の変化に富み、最近の楽曲ほど音量が均一化されていることを示す。この原因として、使用される楽器や機器の変化、屋外で聴くことを意識したマスタリング工程などが考えられる。各楽曲の spectral flatness の分散値は、Mr.Children や浜崎あゆみなど一部のアーティストの値が低くなっていた。これらのアーティストの楽曲は周波数別の音量の均一さが曲を通して低かった。そのため、楽器編成や録音手法に一定の特徴があることが示唆される。

5. まとめと今後の課題

本研究では、メタ情報を付与した楽曲群について、機械学習を用いて特徴量を算出し、メタ情報と音響特徴量の分布を可視化してその結果の観察した。そして、1980年代から2010年代の各時代の楽曲と、特定のアーティストの楽曲には、ビートやエネルギーに特徴があることを発見した。さらに、これらの特徴の中でも特に、年代間では rms energy に、アーティスト間では spectral flatness に特徴的な差があることを示した。

今後の課題として、実験に扱うメタ情報、音響特徴量や可視化手法のさらなる検討があげられる。本研究に特化した機械学習モデルを構築して新たな特徴を発見することも視野に入れる。

参考文献

- [1] 一般社団法人日本レコード協会, 音楽メディアユーザー実態調査2019年度調査結果報告書 公表版.
<https://www.riaj.or.jp/f/pdf/report/mediauser/softuser2019.pdf>, 2020.
- [2] Kin Wai Cheuk, Yin-Jyun Luo, Balamurali B, T, Gemma Roig and Dorien Herremans, "Regression-based music emotion prediction using triplet neural networks," IJCNN 2020, 2020.
- [3] Matteo Lionello, Luca Pietrogrande, Hendrik Purwins and Mohamed Abou-Zleikha, "Interactive Exploration of Musical Space with Parametric t-SNE," 15th Sound and Music Computing Conference, At Limassol, Cyprus, 2018.
- [4] Vienna University of Technology, Music Information Retrieval,
<http://ifs.tuwien.ac.at/mir/musicbricks/#RPextract> (2021-02-11 参照).
- [5] Librosa development team, Librosa,
<https://librosa.org/doc/latest/index.html#> (2021-02-11 参照).