E-054

プレイリストに内在する人間の潜在的認識を反映した楽曲特徴量の提案

西原 泰宇† 市川 治†
滋賀大学データサイエンス学部

1. はじめに

近年，音楽ストリーミングサービスの発展に伴い,ユーザは膨大な選択肢の中からニーズに合った楽曲を見つけるのが難しくなっている.一方で,「その時の気分や状況に合った曲を聴きたい」「普段はあまり聴かないジャンルだが,自分の好みに合う新しい曲に出会いたい」といったニーズが高まっている.従来の推薦手法には,ユーザの視聴履歴を基にする協調フィルタリングと,楽曲の音響的特徴に基づくコンテンツベースフィルタリングがある.しかし,前者は人気曲に偏りやすく,後者は推薦が画一的になりやすいという課題がある.これらの問題は推薦された曲がユーザの嗜好や感性に合致せず,推薦曲が捨てられてしまうことにつながる.本質的に,ユーザの嗜好や背景によって楽曲に対する認識は異なっており,その認識は必ずしも音響特徴量のような直接的なデータには反映されない.例えば,「夏」と聞いて連想する情景や感情（海,恋,ドライブなど）は,既存の特徴量には表れにくい.そこで本研究では,楽曲に対するユーザの認識を集約した「プレイリスト」に注目する.プレイリストは,ユーザが「嗜好」や「状況」「感情」に応じて選んだ楽曲の集合であり,特定の文脈で意味的に関連する曲が含まれていると考えられる.本研究の目的は,こうした集合知を活用し,楽曲の“潜在的な認識”を表す新たな特徴量を抽出・提案することである.これにより,より感情に寄り添った音楽推薦や自動作曲の実現を目指す.

2. 関連研究

楽曲推薦では,ユーザの多様なニーズや文脈に応じた高度な手法が提案されてきた.近年は,ユーザの属性や状況（時間帯・活動・感情）を考慮した推薦が注目されており,Neural Collaborative Filtering (NCF) [1]は深層学習によりユーザとアイテムの非線形関係を学習可能とした.Sakurai ら[2]は知識グラフと音響特徴を統合した強化学習により Cold Start 問題に対応し,Kang ら[3]は MERT 音響埋め込みと音楽理論的特徴を用いた感情認識モデルを提案している.

音響特徴に基づく手法としては,Lee ら[4]のエンドツーエンド埋め込み学習や,Magron と Févotte[5]による音響特徴とユーザ埋め込みの同時学習がある.また,Naseri ら[6]は歌詞と音響のムード認知への影響を分析し,歌詞由来の感情的特徴の有効性を示した.Li らの MAP-MUSIC2VEC[7]は Transformer ベースの軽量モデルであり,音声からタグ・感情・ジャンルなどを効率的に抽出可能である.このように音響的特徴に基づく意味的表現の抽出は進展しているが,本研究はプレイリストというユーザ行動に内在する意味構造を中間表現として活用し,感性に基づく特徴量への変換を試みる点で補完的である.

プレイリストを利用した研究としては,Paprej ら[8]が曲を単語,プレイリストを文章とみなし,Seq2Seq モデルでジャンル予測などを行っている.本研究もプレイリストを分析対象とする点は類似するが,分類ラベルとして活用し,感性を反映した特徴量を獲得するという点で新規性がある.また,Chen ら[9]は嗜好の多様性と人気偏向を考慮した Top-N 推薦を提案しているが,本研究ではプレイリストを「意味的分類ラベル」として用いる点で異なる.

以上より,本研究は音響特徴や明示的文脈ではなく,プレイリストに内在する集合知的な意味・感性構造に着目し,それを中間表現として抽出・再構成することで,より人間の感性的認知に近い楽曲特徴の獲得を目指す.

3. 提案手法

本研究では,集合知に基づく新たな特徴量を得るために,図 1 の深層学習モデルを設計した.このモデルは,楽曲の音響特徴などを入力として受け取り,その曲が含まれるプレイリストのクラスを予測するマルチクラス分類タスクを行う.モデルには,活性化関数に GERU,損失関数にカテゴリカルクロスエントロピー,出力層に softmax 関数,オプティマイザに Adam を使用している.

Proposal of Music Features Reflecting Human Affect Inherent in Playlists.

Teu NISHIHARA†, Osamu ICHIKAWA†

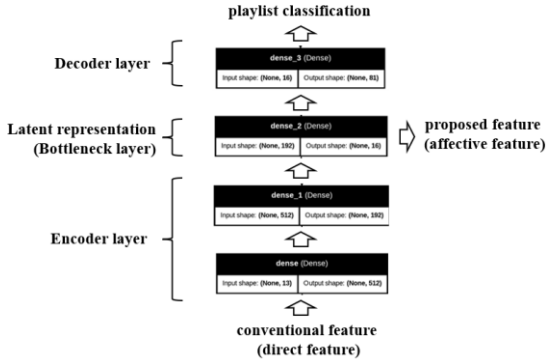

図 1：提案モデルのアーキテクチャ

本研究では，最終出力の一つ前の隠れ層（13 次元）から得られるベクトルを，新たな楽曲特徴量として利用する．この中間層は，プレイリスト分類において重要な情報を圧縮・抽出しており，次の仮定が成り立つと考える：

・モデルが十分に学習されていれば，この層は人間の概念的特徴に近い潜在表現を獲得している．

・プレイリストを通じて共有される「集合知的」な意味構造が，この層に反映されている．

つまり，この特徴量は単なる音響の類似ではなく，ユーザの認識や文化的背景を反映した意味的な表現である．

4. 評価実験 1

本実験では，提案手法によって得られた新たな特徴量（提案特徴量）が，楽曲の特徴空間をより重複なくコンパクトに表現できているかを検証することを目的とする．評価には，spotify_data[10]（n = 1,159,764）を用いた．本データセットは，Spotify に収録されている膨大な楽曲を収集したものであり，13 種類の特徴量と楽曲のジャンルが含まれている．

本研究では，これら 13 種類の数値特徴量を音響的特徴量，ジャンル情報をプレイリスト情報とみなして実験に用いた．提案手法の有効性を理論的に評価するため，クラスタリング手法およびシルエットスコアを用いた比較を行った．まず，従来の特徴量を用いた評価（A）として，13 次元の元データに標準化処理を施した後，x-means クラスタリングを適用し，各クラスタにおけるユークリッド距離に基づいてシルエットスコアを算出し，その平均値を求めた．

次に，提案手法による評価（B）では，まず提案特徴量を取得するため，13 次元の入力と 82 ジャンルの出力をもつマルチクラス分類モデルを spotify_data 上で 30 エポック学習し（最終的な accuracy = 0.22），その最終層の一つ手前に位置する隠れ層（13 次元）の出力を提案特徴量として抽出した．得られた提案特徴量に対して標準化処理を施し，StepA と同数のクラスタ数を用いてクラスタリングを実施した後，各クラスタにおけるシルエットスコアを同様に算出し，平均値を求めた．シルエットスコアは，クラスタリングの質を測る指標であり，以下の S(i)の平均値として定義される．

$$S(i) = \big(b(i) - a(i)\big)/max\big(a(i), b(i)\big)$$

ここで，a(i)は同一クラスタ内における平均距離（凝集度），b(i)は最も近い他クラスタとの平均距離（乖離度）を表し，スコアは −1 から +1 の範囲をとる．値が大きいほど，クラスタリングの理論的な適切性が高いことを示す．

図 2 と図 3 は，それぞれの特徴量におけるクラスタリングの結果を TSNE で 2 次元に圧縮して可視化したものである．

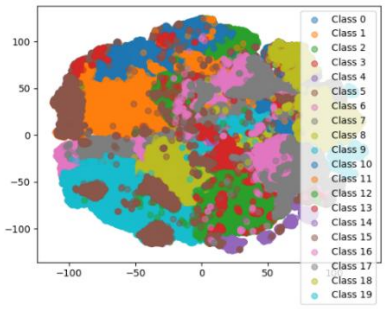

図 2：直接特徴量のクラスタリング結果可視化

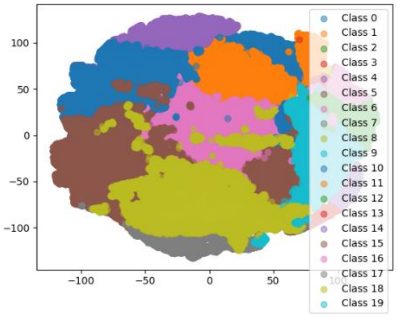

図 3：提案手法のクラスタリング結果可視化

図 2,3 から，提案特徴量を用いたクラスタリングの方が直接的な特徴量を用いたクラスタリングよりもまとまりが良いことがわかる．表 1 は，それぞれの特徴量における各クラスタのシルエットスコアの平均値である．

表 1：各特徴量におけるシルエットスコア

	シルエットスコア
提案手法	0.257
直接特徴量	0.121

この結果から，提案手法で得られた提案特徴量を用いたクラスタリングの方が，理論的により良いクラスタ構造を形成していることがわかる．これは，提案特徴量が音響的な情報に加えて，ユーザの集合知やジャンル認識といった意味的・文脈的な情報を潜在的に含んでいることを示唆しており，提案手法が理論的に有効であると判断できる．また，プレイリスト分類モデルの分類精度（accuracy）は 0.22 と，81 クラス分類において一見して高精度とは言い難いが，本研究では，高精度の分類モデルの構築を目的としてはいない．むしろ，中間層の潜在表現を得ることが目的である．

更に，この実験を実際のユーザが作成したプレイリストにも応用した．具体的には，先ほどの実験で用いたモデルを用い，ユーザが作成したプレイリストを 15 個のカテゴリに手作業で分類した「大分類プレイリスト」を用いて転移学習を行った．その際，モデルの最終層の一つ手前にある 13 次元の隠れ層から出力される情報を「提案特徴量」として抽出した．そして，プレイリスト内の各楽曲に対して，その特徴量と提案特徴量をもとにクラスタリングを実施し，クラスタリングの妥当性を評価するためにシルエットスコアの平均値を算出した．

表 2：各特徴量におけるシルエットスコア

	シルエットスコア
提案手法	0.3333
直接特徴量	0.1627

この結果から，実際にユーザが作成したプレイリストに対しても提案手法で得られた提案特徴量を用いたクラスタリングの方が，理論的により良いクラスタ構造を形成していることがわかる．

5. 評価実験 2

本実験では，提案手法により得られた特徴量が実際にユーザの感性，特にユーザの嗜好や想定している状況に沿っているかを検証するために，13 名を対象にアンケート調査を実施した．調査では，事前調査として「好みの曲」および「夏に合う曲」をそれぞれ数曲回答してもらい，その回答をもとに 2354 曲の中から以下の 2 通りの方法で各 3 曲を推薦した．

1. 直接的な特徴量（Spotify API から得られる特徴量）を用いた推薦

2. 提案手法（評価実験 1 で用いた転移学習モデルから得られた特徴量）を用いた推薦

推薦を行う際には，データベース内の各曲に対して，推薦の基準となる曲とのユークリッド距離を計算し，その距離が最も小さい（ただしタネ曲そのものは除く）3 曲を推薦対象とした．なお，評価の集計にあたっては，無効な回答を除外している．

「好みの曲」に対する推薦結果を図 4，図 5 に示す．このアンケートでは，推薦された曲が「好みに合っているかどうか」を「特に好みにあっている」，「好みに合っている」，「好みに合っていない，または嫌い」の 3 段階で，また「聞いたことがあるかどうか」を「はい」，「いいえ」の 2 段階で回答してもらった．有効回答数は 10 名であった．図 4 は，提案手法ごとに推薦された曲のスコア分布を可視化したものである．図 5 は提案手法によって得られた『好みの曲』に対するユーザごとの平均スコアを可視化したものである．また，表 3 は手法ごとにスコアの平均をとったものである．

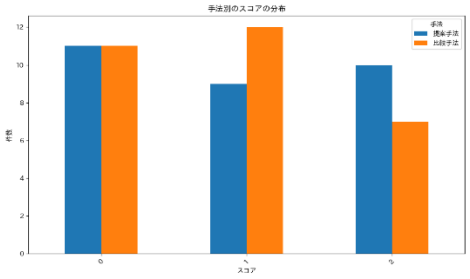

図 4：手法別のスコアの分布

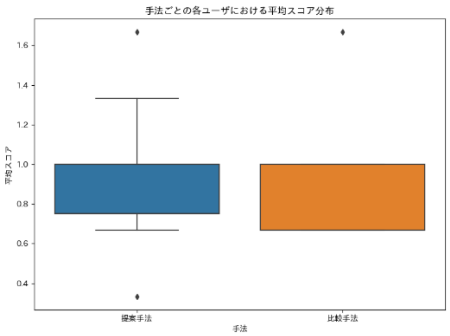

図 5：手法別のユーザごとの平均スコアの分布

表 3：手法ごとのスコアの平均

	平均スコア
提案手法	0.967
従来手法	0.867

図 4 および図 5，表 3 から，わずかながら提案手法の方が従来手法よりも好みに合った推薦の精度が高いことが判明した．また，図 4 から提案手法

は従来手法よりも特に好みにあった曲を推薦していることが読み取れる.

更に,推薦された曲がユーザにとって既知の曲であったか否かで層別に手法ごとのスコアの平均を計算した.表 4 はその結果である.なお,今回は用紙の都合で未知の曲についてのみ表記する.

表 4:未知の曲における手法ごとのスコア平均

	平均スコア
提案手法	0.533
従来手法	0.388

表 4 より,提案手法は未知の曲についてユーザの好みに合う曲を比較手法よりも高い精度で推薦していることが判明した.

「夏に合う曲」に対する推薦結果を図 6,表 5 に示す.このアンケートでは,推薦された曲が「夏に合っており,かつ聞きたいと思える曲であったか」を「はい」,「いいえ」の 2 段階で回答してもらった.有効回答数は 11 名であった.図 6 は, 提案手法による『夏に合う曲』に対するユーザごとの平均スコアの分布を可視化したものである.また,表 5 は,手法ごとのスコアの平均値を示している.

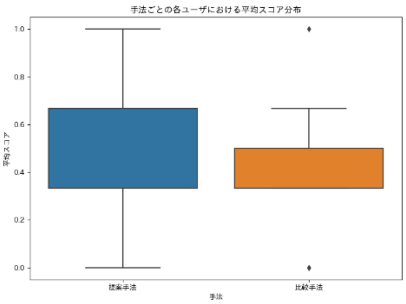

図 6:手法別のユーザごとの平均スコアの分布

表 5:スコアごとのスコアの平均

	平均スコア
提案手法	0.454
従来手法	0.286

表 5 及び図 6 により,状況に対する推薦に対しても提案手法のほうが従来手法と比較して精度が高いことが示された.

6. 終わりに

本研究では,ユーザが作成したプレイリストに内在する意味構造に着目し,感性や文脈を反映した楽曲の新たな特徴量を提案した.プレイリストを「集合知」として扱うことで,音響特徴だけでは捉えにくい文脈的・感情的側面を抽出可能とした.評価実験では,提案手法によりクラスタリング精度と推薦の適合度が向上し,特に未知の楽曲に対する推薦で有効性が確認された.今後は,Music2Vec のような手法に対して本手法を適用し,有効性のさらなる検証を行う予定である.

7. 謝辞

本研究は JSPS 科研費 25K15402 の助成を受けた.

8. 参考文献

[1]He, X., Liao, L., Zhang, H., Nie, L., Hu, X., Chua, T.-S.: Neural collaborative filtering. In: Proceedings of the 26th International Conference on World Wide Web (WWW 2017), pp. 173-182 (2017).
[2]Sakurai, K., Togo, R., Ogawa, T., Haseyama, M.: Deep reinforcement learning-based music recommendation with knowledge graph using acoustic features. ITE Transactions on Media Technology and Applications 10(1), 8-17 (2022).
[3]Kang, J., Herremans, D.: Towards unified music emotion recognition across dimensional and categorical models. In: Proceedings of the 23rd International Society for Music Information Retrieval Conference (ISMIR 2022), pp. 783-790 (2022)
[4]Lee, J., Lee, K., Park, J., Park, J., Nam, J.: Deep content-user embedding model for music recommendation. arXiv preprint arXiv:1807.06786(2018)
[5]Magron, P., Févotte, C.: Neural content-aware collaborative filtering for cold-start music recommendation. Data Mining and Knowledge Discovery 36(5), 1790-1810 (2022).
[6]Naseri, S., Reddy, S., Correia, J., Karlgren, J., Jones, R.: The contribution of lyrics and acoustics to collaborative understanding of mood. In: Proceedings of the 16th International AAAI Conference on Web and Social Media (ICWSM2022),pp.687-698(2022)
[7]Li, Y., Yuan, R., Zhang, G., Ma, Y., Lin, C., Chen, X., Ragni, A., Yin, H., Hu, Z., He, H., Benetos, E., Gyenge, N., Liu, R., Fu, J.: MAP-Music2Vec: A simple and effective baseline for self-supervised music audio representation learning. arXiv preprint arXiv:2212.02508(2022)
[8]Papreja, P., Venkateswara, H., Panchanathan, S.: Representation, exploration and recommendation of playlists. In: Cellier, P., Driessens, K. (eds) Machine Learning and Knowledge Discovery in Databases - ECML PKDD 2019,* CCIS** 1168*, pp. 543-550. Springer, Cham (2020).
[9]Chen, S.-H., Sou, S.-I., Hsieh, H.-P.: Top-N music recommendation framework for precision and novelty under diversity group size and similarity. Journal of Intelligent Information Systems 62(1), 1-26 (2024)
[10]Spotify for Developers: Get Audio Features. https://developer.spotify.com/documentation/web-api/reference/get-audio-features, last accessed 2025/04/17