

ホームビデオへの自動 BGM 付与のための 心理学に基づく音楽分類手法

小野佑大† 石先広海‡ 帆足啓一郎‡ 滝嶋康弘‡ 甲藤二郎†

早稲田大学大学院基幹理工学研究科† KDDI 研究所‡

1. はじめに

大規模な音楽データベースから映像に適切な BGM を選定する作業は多大な労力を要する。一般に BGM の選択作業は映像と音楽との雰囲気的一致に基づき行われる。そこで本稿では映像に適切な BGM を効率よく探し出すために楽曲を心理学に基づく特徴空間上に写像し音楽の明示的な雰囲気を表現するクラスタを生成する。

2. 提案手法

2.1 音楽のムードを表現する特徴空間

音楽のムードを表す特徴空間として Russell モデル[1]を用いる。Russell モデルは人間の感情を、活性・不活性(Arousal)、快・不快(Valence)を二軸とした二次元平面で定義し、楽曲をモデル上に直接写像して得られた座標値(Arousal and Valence 値、以下 AV 値)を音楽の雰囲気を表現するパラメータとして用いる。

2.2 AV 値の計算

Russell モデルにおける AV 値を計算するために、楽曲の雰囲気と関連する特徴量全 29 次元[1](フレームごとに抽出しその平均と標準偏差を使用)を抽出し平均 0、分散 1 に正規化後、主成分分析を行う。各主成分の因子負荷量を分析すると第一主成分は Arousal に負の相関、第二主成分は Valence に正の相関を示した。したがって Arousal として第一主成分得点に-1 を乗じたもの Valence として第二主成分得点を用いる。

2.3 AV 値に基づく音楽分類手法

BGM を効率よく選定するにはより明示的なクラスタの生成が重要である。そこで本稿ではクラスタの特徴を強調するために AV 空間の性質を考慮した楽曲分類を行う。一般に同一象限内の AV 値間の距離が近い楽曲はムードが類似している。しかし単純に k-means 法を適用しても象限を無視して無作為にクラスタが生成され、その特徴を弱めてしまう(図 1)。そこで各軸に対し階層的にクラスタを生成する。

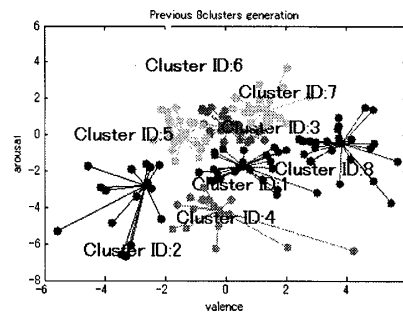


図 1 従来手法のクラスタ生成結果

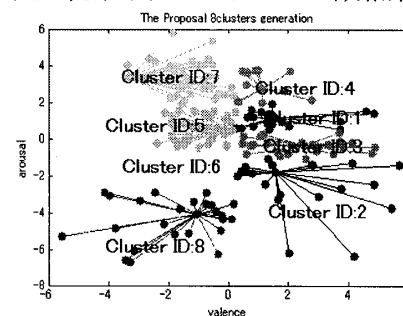


図 2 提案手法のクラスタ生成結果

例として 8 クラスタ生成する方法を示す。初めにクラスタ数 2 で Valence 値に k-means 法を適用する。次に各クラスタに対してクラスタ数 4 で Arousal 値に k-means 法を適用する(図 2)。これにより各象限を考慮した楽曲分類を行える。

3. 実験

3.1 データセット

大規模音楽データベース上には様々な雰囲気の楽曲が混在する。そこで本稿では、楽曲データとしてコロンビア大学より提供される us pop data set[2]から音楽情報サイト AMG(all music guide.com)[3]でムードラベルが付与されている楽曲 206 曲を選出し、AMG で提供される全 179 種類のラベルの内 133 種類利用する。

3.2 クラスタ評価方法

本稿では 8 クラスタ生成し、各手法のどちらで得たクラスタがより明示的な特徴を示すか評価する。クラスタの特徴はそれに所属する複数楽曲に付与されるラベル(本稿では重複ラベルと呼ぶ)で説明される。よって各クラスタの重複ラベルの平均出現頻度を求め全クラスタの平均を評価値とする(表 1 の評価値 1)。

Music mood classification for automatic background music synchronization for a home video

†Yudai Ono, Jiro Katto

Waseda University

‡Hiromi Ishizaki, Keiichiro Hoashi, Yasuhiro Takishima

KDDI R&D Laboratories Inc.

3.3 クラスタ評価

表 1 評価値の算出結果

	Previous	Proposal
評価値 1	2.33	2.35
評価値 2	5.08	5.50

表 2 各手法の Cluster ID 1 の重複ラベルの例

Previous	Proposal
Brooding	Boisterous
Campy	Energetic
Elegant	Rousing
Theatrical	Theatrical

表 1 の「評価値 1」を見ると各手法の重複ラベルの平均出現頻度はほぼ同値を示す。これは本稿のデータセットが楽曲数の割にユニークなラベルが多く、重複ラベルを適切に得られないことが原因である。ここで全楽曲におけるラベルの出現頻度(図 3)を見ると、1 ラベル当たりの付与楽曲数が少ない重複ラベルが多く存在する。つまりクラスタ内に多様な重複ラベルが混在し互いの関連性が不明確となるためラベルの多様性を考慮した評価方法を検討する必要がある。

本稿では重複ラベル間の類似性に着目し、まず各クラスタの楽曲に付与されるラベルから重複ラベルとそれに類似した雰囲気のリベルを取得する。ラベル間の雰囲気のリベルの類似性は AMG の Similar Mood(本稿では類似ムードと呼ぶ)を参照する。類似ムードは 1 ラベルに対し全 179 種類の中から約 13 種類用意される。次にクラスタ毎に各重複ラベルの類似ムードを結合したリストを作り(本稿では類似リストと呼ぶ)そのリストにおける重複ラベルの平均出現頻度を求め、全クラスタの平均を評価値とする(表 1 の評価値 2)。

表 1 の「評価値 2」に対し、手法間の差が開いた要因として、図 1 の Cluster ID 1, 3, 5 の様に特徴空間の中央付近に象限を跨いで生成されたクラスタの楽曲の雰囲気が多様で、その特徴を弱めてしまった事が示唆される。その例として重複ラベルが同数のクラスタである、図 1, 2 の Cluster ID 1 の類似リストに出現するラベルの出現率を図 4, 5 に示す。図 4 は全ラベルが万遍なく出現しピークが目立たない。つまり各ラベルが同程度出現しクラスタを特徴的に示せない。それに対し、図 5 は複数ピークが際立ちクラスタの特徴を表すラベルの存在が示唆される。表 2 に Cluster ID 1 の重複ラベルの例を示す。従来手法の重複ラベルは「憂鬱」「優雅」など類似性が低いのに対し提案手法は「騒々しい」「活気のある」など類似性の高い重複ラベルが存在している。よって提案手法で得たクラスタは直感的に分かりやすく、検索効率向上に繋がる。

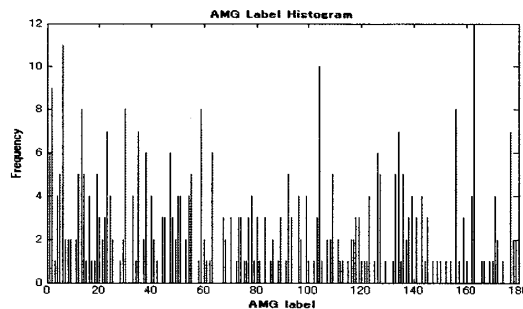


図 3 全楽曲に対するラベルの出現頻度

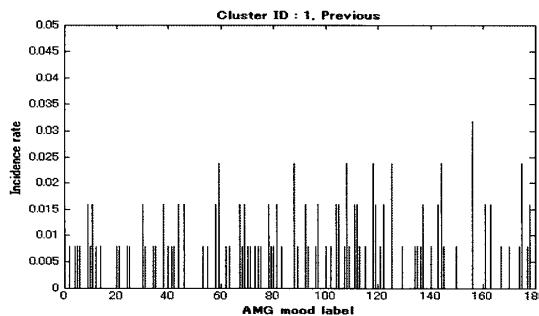


図 4 従来手法の Cluster ID 1 の類似リストに出現する重複ラベルの出現率

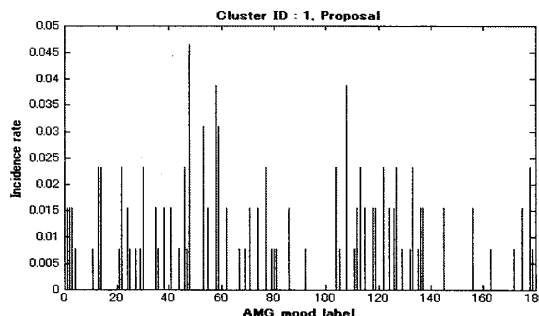


図 5 提案手法の Cluster ID 1 の類似リストに出現する重複ラベルの出現率

4.まとめ

本稿では BGM 検索の効率化のために、楽曲の雰囲気を表現する特徴空間において最適な音楽分類法を検討した。その結果、提案手法で生成されたクラスタはその特徴を明示的に示していることが分かった。今後はホームビデオと音楽とを対応付ける方法を検討し、最適な BGM を自動付与するシステムの構築を目指す。

参考文献

- [1] Tuomas Eerola, Olivier Lartillot, Petri Toiviainen, "Prediction of multidimensional emotional ratings in music from audio using multivariate regression models," *International Society for Music Information Retrieval Conference (ISMIR)*, 2009
- [2] D. Ellis, "The uspop2002 Pop Music data set," Available at <http://labrosa.ee.columbia.edu/projects/musicsim/uspop2002.html>
- [3] all music guide, Available at <http://www.allmusic.com/>