

AAC 圧縮楽曲データに対するジャンル分類手法

湯澤幸一朗

小早川倫広

東京都立産業技術高等専門学校

1 はじめに

近年,Apple の iTunes などのインターネットを介した楽曲配信サービスが普及し,ユーザは大量に楽曲を入手することが可能となった. これに加え,楽曲圧縮技術の開発により,楽曲配信サービスで扱われる楽曲データは AAC や MP3 などの楽曲データとなっている. したがって,膨大な圧縮楽曲データに対する効率的な管理や検索手法の要求が高まってきている. 楽曲を検索する手法として,テキスト検索やハミング検索などがある. テキスト検索をするうえで重要なのが,楽曲のメタデータである. しかし,メタデータ中のジャンル情報に関しては,ユーザの楽曲に対する解釈の違いなどにより,必ずしもユーザの意図するものと一致しない場合がある.

そこで,楽曲の内容に基づくジャンル分類手法が求められる. 楽曲の内容に基づくジャンル分類は国際会議のコンテストになり, [1-3] などで盛んに行われている. 本稿では,圧縮された楽曲データの内容に基づくジャンル分類手法を提案する. MPEG-4 AAC(Advanced Audio Coding) で圧縮された楽曲データのビットストリームに格納されたパラメータを用いて楽曲の内容を表す特徴量を抽出し,それを用いてジャンル分類を行う.

2 特徴量抽出

AAC 圧縮楽曲データはフレーム単位でビットストリームが形成されており,ヘッダ部分と各パラメータから構成されている. 本稿では,MPEG-4 AAC 圧縮楽曲データのビットストリームに含まれる spectral data パラメータに着目する. spectral data パラメータは,MPEG-4 AAC 圧縮過程で算出される MDCT 係数を量子化したものである. MDCT 係数は時間領域の音楽波形を周波数帯域の値に変換したものであり,どの周波数帯域がどれだけのエネルギーを持っているかを表している.

特徴量の抽出手法は以下のとおりである. 特徴量の抽出には,周波数解析と時系列解析を行うことができる離散ウェーブレット変換を用いた. まず, N フレームからなる MPEG-4 AAC 圧縮楽曲データのビットストリーム内に格納されている spectral data パラメータを得る. spectral data パラメータは可変長で,最大で 1024 次元である. パラメータの次元数を揃えるため,1024 次元に満たない場合は 1024 次元になるまで 0 を割り当てる. したがって, n フレームにおける 1024 の spectral data パラメータ $s_n = (s_{n1}, s_{n2}, \dots, s_{n1024})^T$ を得る. ここで, n フレームにおける $m(m = 1, 2, \dots, M)$ 番目のサブバンドエネルギーを

$$e_n^m \equiv \sum_{k=w(m-1)+1}^{wm} (s_{nk})^2, \quad w = \frac{1024}{M},$$

とする. すなわち, n フレームにおけるサブバンドエネルギー $e_n = (e_n^1, e_n^2, \dots, e_n^M)^T$ を得る. N フレームからなる AAC 圧縮楽曲データのサブバンドエネルギー系列は $e \equiv (e_1, e_2, \dots, e_N)$ と表す. サブバンドエネルギー系列 e の j 行目の要素 $e^j = (e_1^j, e_2^j, \dots, e_N^j)$ に対して,Haar 基底の離散ウェーブレット変換 $DWT(e^j)$ から各ウェーブレット変換の分解レベル $l(l = 1, 2, \dots, L)$ の係数列 $v^{jl} = (v_1^{jl}, v_2^{jl}, \dots, v_{\lfloor \frac{N}{2^l} \rfloor}^{jl})$ を得る. ここで,離散ウェーブレット変換 $DWT(e^j)$ から得られる係数列 v^{jl} の平均 μ^{jl} , 標準偏差 σ^{jl} を算出し,楽曲の特徴量 $f_\mu, f_\sigma, f_{\mu,\sigma}$ とする.

$$f_\mu \equiv (\mu^{11}, \mu^{21}, \dots, \mu^{M1}, \mu^{12}, \dots, \mu^{M2}, \dots, \mu^{ML}),$$

$$f_\sigma \equiv (\sigma^{11}, \sigma^{21}, \dots, \sigma^{M1}, \sigma^{12}, \dots, \sigma^{M2}, \dots, \sigma^{ML}),$$

$$f_{\mu,\sigma} \equiv (f_\mu, f_\sigma).$$

3 実験

実験に用いるデータセットとして,ボサノヴァ,シャンソン,童謡,ダンス,ヒップホップ,ジャズ,メタル,オールディーズ,レゲエ,タンゴの 10 種類のジャンルから 1498 曲を用意し, AAC 圧縮楽曲データを得る. AAC 圧縮楽曲データから,上記の方法で楽曲特徴量 $f_\mu, f_\sigma, f_{\mu,\sigma}$ を抽出する. 得られた特徴量を用いて以下の手順で実験を行った. また,特徴量抽出を行う周波数帯域の分割数 M を 16,32,64,128, ウェーブレット変換のレベル L を 9 として実験を行った.

1. 無作為に各ジャンル毎に楽曲を選択し,学習データセットを作る. 学習データセットのサイズは各ジャンル毎の楽曲総数に対して 50% とする.
2. 学習データセットに対して特徴量を用いて線形判別分析を行い,判別空間を作る.
3. 学習データセットとして使われなかった楽曲をテストデータセットとする. 各楽曲の特徴量を用いて,判別空間に写像してジャンル分類を行う.
4. 学習データセットとテストデータセットを入れ替えて 1-3 を行い,テストデータセットに対する平均正答率を算出する.

4 結果と考察

図 1 に各サブバンド分割に対して特徴量を用いた場合の最も高かった平均正答率を示す. 図 1 から,分割数が 16 以外の場合に $f_{\mu,\sigma}, f_\sigma, f_\mu$ の順で平均正答率が高くなっており,分割数を増やすことで,平均正答率が高くなることが分かる. 平均正答率が最も高かったのは,分割数が 128 で特徴量が $f_{\mu,\sigma}$ の場合で,80.31%であった. こ

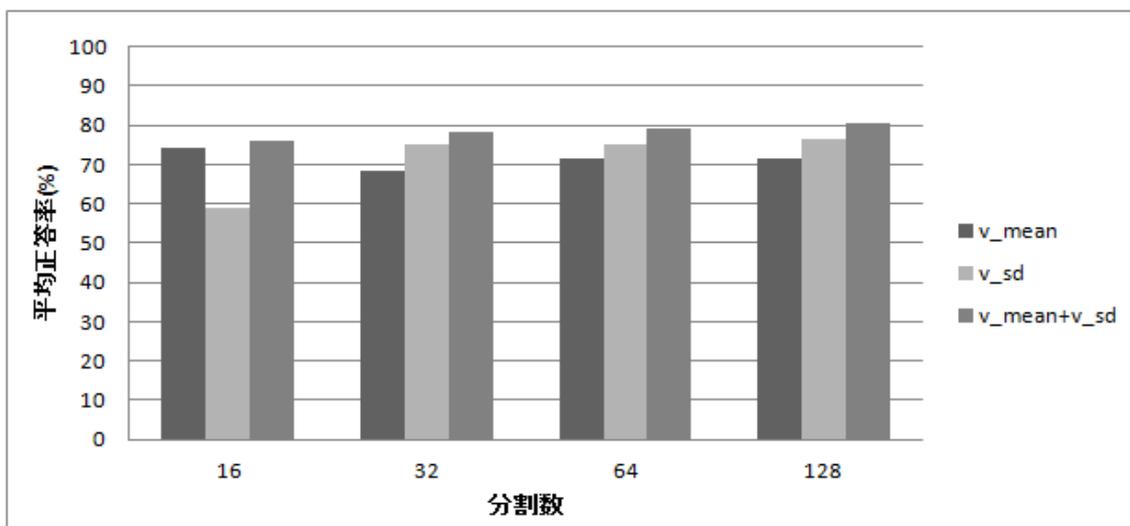


図1 平均正答率

表1 $f_{\mu,\sigma}$ を用いた場合のテストデータの分類結果

	bossa	chanson	children	dance	hiphop	jazz	metal	oldies	reggae	tango
bossa	59.3	16.9	0.0	3.4	5.1	3.4	1.7	6.8	1.7	1.7
chanson	5.2	81.8	2.6	0.0	1.3	0.0	0.0	3.9	0.0	5.2
children	0.0	6.1	90.8	0.0	0.0	1.0	0.0	0.0	0.0	2.0
dance	0.0	0.0	0.0	81.8	6.8	0.0	4.5	4.5	2.3	0.0
hiphop	6.0	1.0	0.0	1.0	85.0	0.0	1.0	0.0	5.0	1.0
jazz	6.0	0.0	2.0	0.0	2.0	86.0	0.0	0.0	0.0	4.0
metal	1.5	1.5	0.0	1.5	0.0	0.0	94.1	1.5	0.0	0.0
oldies	5.0	8.3	3.3	3.3	3.3	0.0	8.3	65.0	0.0	3.3
reggae	2.0	0.0	0.0	6.1	14.3	0.0	0.0	0.0	75.5	2.0
tango	1.0	11.0	2.0	0.0	0.0	0.0	0.0	2.0	0.0	84.0

れは、周波数帯域の情報が増え、平均正答率が向上したのではないかと考えられる。

図1の中で平均正答率が最も高かった場合のテストデータの分類結果を表1に示す。表中の各セルの値は、各ジャンル行のデータがどのジャンル列に分類されたかの割合を示している。表中の対角要素の値が各ジャンルのテストデータが正しく分類された楽曲の割合を表している。表1から、各ジャンルの正答率についてみると、最低でボサノヴァの59.4%、最高でメタルの94.1%、平均で80.3%となった。ボサノヴァについてみると、シャンソンとオールディーズに多く誤分類されている。これは、特徴量が楽曲のエネルギーの変化を表しているため、ボサノヴァの楽曲の緩やかなエネルギーの変化がシャンソンやオールディーズと似ているためだと考えられる。また、メタルについてみると、誤分類されている楽曲が少ない。これは、メタルという楽曲のエネルギーの変化が他のジャンルよりも激しいためだと考えられる。

5 まとめ

本稿では、MPEG4-AAC 圧縮楽曲データに対するジャンル分類手法を提案した。AAC 圧縮楽曲データのビットストリームに格納されている spectral data パラメータに対して離散ウェーブレット変換を行い特徴量を算出し、線形判別分析によりジャンル分類を行った。その結

果、80.31%の平均正答率を得ることができた。

今後の予定として、データセットの楽曲数を増やしたり、異なるジャンルのデータセットに対して実験を行うことが挙げられる。また、本稿で述べた特徴量よりもジャンル分類に対して有効な特徴量の抽出手法についても検討していく。

参考文献

- [1] MIREX2013 Results, http://www.music-ir.org/nema_out/mirex2013/results/act/mixed_report/
- [2] Ming-Ju Wu and Jyh-Shing Roger Jang, "MIREX 2013 Submissions for Train/Test Tasks (Draft)," Available: <http://www.music-ir.org/mirex/abstracts/2013/JJ1.pdf>.
- [3] Karam Byun, Seung Ryoel Baek, Jong-Seol Lee, SEI-JIN JANG, Moo Young Kim, "MUSIC GENRE/MOOD/COMPOSER CLASSIFICATION:MIREX 2013 SUBMISSIONS," Available: <http://www.music-ir.org/mirex/abstracts/2013/BBLJK1.pdf>.