修士論文要旨 (2021 年度)

冗長ウェーブレット変換を用いた バンド間相関とフレーム間相関に基づく楽曲特徴量の抽出

Audio Feature Extraction Based on Inter-Band Correlation and Inter-Frame Correlation by Using Undecimated Wavelet Transform

電気電子情報通信工学専攻　大石　智貴

1.　まえがき

音楽ストリーミング配信サービス，および動画配信サービスの普及により，ユーザは今までより多くの，そして世界中に存在する様々な種類の音楽に簡単にアクセスできるようになった．その一方で，膨大な量の楽曲の中から自分の興味のある音楽をいかに効率良く探し出せるかということが課題となっている [1].

ユーザが好みの楽曲を探す方法の一つに「楽曲推薦システム」がある．これは，ユーザの行動履歴から好みを分析し，それらに類似した未知の楽曲を推薦するものである．このシステムが推薦時に利用するデータとして，他のユーザの行動履歴を利用するものを「協調フィルタリング」，楽曲の音楽内容を考慮するものを「内容に基づくフィルタリング」，それら両者を考慮するものを「ハイブリッド型フィルタリング」と呼ぶ [5]．本研究では，内容に基づくフィルタリングよる楽曲推薦システムに向け，様々なジャンルの楽曲において適切な音響特徴量を抽出するための手法について提案し，それらの特徴量に基づく音楽ジャンル分類を行うことで，分類の正答率により評価を行う．また，正答率の向上及び特徴量の次元数削減を目的とする．

2.　先行研究

本論文の先行研究として，小林拓哉氏らの論文がある [3]．図 1 に先行研究における手法の流れを示す．この論文では，楽曲信号に対し冗長ウェーブレット変換 (UWT) を用いて抽出した楽曲特徴量と，MIR Toolbox[4] によって抽出した音響特徴量とを統合し，サポートベクターマシン (SVM) を用いた分類器へ入力することで音楽ジャンル分類を行っている．

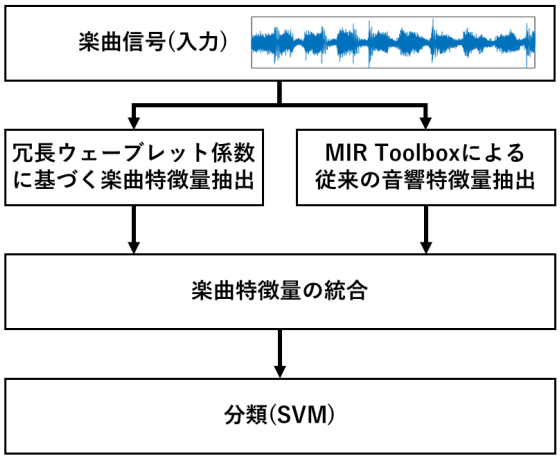

図 1: 先行研究における手法の流れ

3.　提案手法

図 2 に本論文における提案手法の全体の流れを示す．提案手法では，楽曲信号から冗長ウェーブレット変換 (UWT) を用いて楽曲特徴量を抽出する．また，同時に既存の特徴量を抽出し，UWT を用いた特徴量と統合することで，最終的な楽曲特徴量とする．この楽曲特徴量をサポートベクターマシン (SVM) を用いた分類学習器に入力することで音楽ジャンル分類を行い，評価をする．

統計計算処理においては，フレーム間相関係数の追加を行う．また，既存の特徴量として MFCC，GFCC などの追加を行う．楽曲特徴量の統合の段階では，主成分分析を用いた次元の削減を行う．

3.1　フレーム間相関係数

図 3 にフレーム間相関係数抽出の流れを示す．フレーム間相関係数では，s 番目のフレームと $s+1$ 番目のフレームとの間で相関係数を求める．ここで，UWT における分解バンド数を J，相関係数を求めるサブバンドの

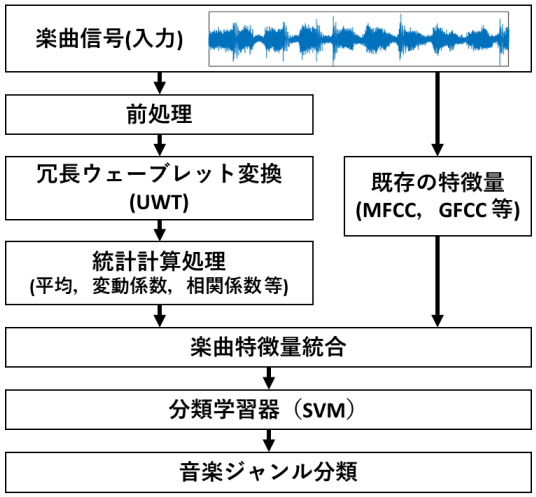

図 2: 提案手法の全体の流れ

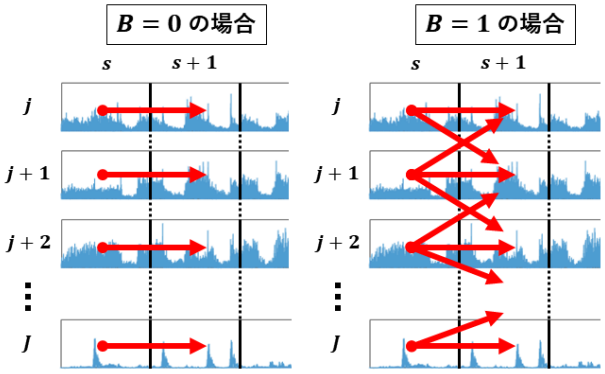

図 3: フレーム間相関係数抽出の流れ

組み合わせを決める隣接バンド幅を B とする．

$B = 0$ の場合，相関係数を求めるサブバンドは同じバンド番号同士のみとなる．$B = 1$ の場合，まず s 番目のフレームの j 番目のサブバンドと，$s + 1$ 番目のフレームの j 番目，$j + 1$ 番目のサブバンドとの相関係数を求める．次に，s 番目のフレームの $j + 1$ 番目のサブバンドと，$s + 1$ 番目のフレームの j 番目，$j + 1$ 番目，$j + 2$ 番目のサブバンドとの相関係数を求める．最終的に，1 つのフレームに対して $J + \sum_{n=1}^{B} 2(J - n)$ 次元の特徴量を抽出する．

4. 実験

特徴量の評価として，GTZAN データセットを用いた実験を行った．このデータセットには，10 種類のジャンル (Blues，Classical，Country，Disco，Hiphop，Jazz，Metal，Pop，Raggae，Rock) が各 100 トラック，計 1000 トラックのオーディオデータが含まれている．これらのトラックはすべてサンプリング周波数 22050 Hz，モノラル，16 ビット，30 秒間のオーディオファイルである．評価には K-分割交差検証法 (K-fold cross validation) を用いた．先行研究 [3] の実験結果に基づき，k=10 と設定し実験を行う．分類学習器には MATLAB のサポートベクターマシン (SVM) を用いる．

局所特徴量としては，平均 $A_k^{(j)}$，変動係数 $CV_k^{(j)}$，バンド間相関係数 $BC_k^{(j)}$，歪度 $S_k^{(j)}$，尖度 $K_k^{(j)}$，フレーム間相関係数 $FC_k^{(j)}$，MFCC $MC_k^{(j)}$，GFCC $GC_k^{(j)}$，ケプストラム係数 $CC_k^{(j)}$ がある．それぞれの局所特徴量に対して平均，変動係数，分散，歪度，尖度を算出することで楽曲特徴量とする．これらと，音量に関する特徴量 A を組み合わせ，実験を行う．

4.1 隣接バンド幅 B を変化させたときの特徴量の組み合わせによる正答率

図 4 に，フレーム間相関係数において隣接バンド幅 B を 0 から 7 まで変化させたときの正答率を示す．また，表 1 にフレーム間相関係数を用いないときと，正答率が最大となったときの特徴量の組み合わせを示す．

結果より，隣接バンド幅 B=1 または 2 のとき，正答率が 86.9%で最大となった．これは基準値に対して，B=1 のとき次元数で 80 次元，正答率で 1.3%の増加であり，B=2 のとき 128 次元，1.3%の増加である．隣接バンド幅 B=1 または 2 のときをそれぞれ次元数に基づき比較をすると，B=1 のとき 319 次元，B=2 のとき 367 次元と，同じ正答率でありながら B=1 のときの方が 48 次元少ないため，本研究においては B=1 が適していると考えられる．

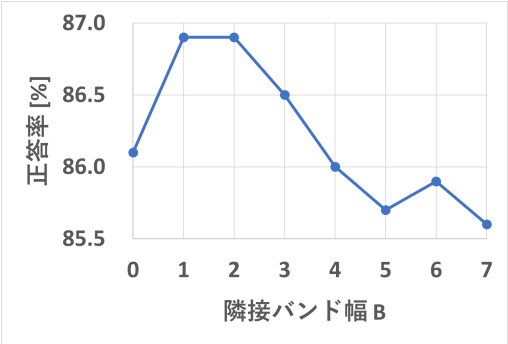

図 4: 隣接バンド幅 B を変化させたときの正答率

表 1: 隣接バンド幅 B を変化させたときの特徴量の組み合わせ

隣接バンド幅 B	特徴量の組み合わせ	次元数	正答率 [%]
なし (基準値)	$A_{ave}^{(j)}$, $A_{cov}^{(j)}$, $CV_{ave}^{(j)}$, $CV_{cov}^{(j)}$, $BC_{ave}^{(j)}$, $BC_{var}^{(j)}$, AVG	239	85.6
1	$A_{ave}^{(j)}$, $A_{cov}^{(j)}$, $CV_{ave}^{(j)}$, $CV_{cov}^{(j)}$, $BC_{ave}^{(j)}$, $BC_{var}^{(j)}$, AVG, $FC_{ave}^{(j)}$, $FC_{var}^{(j)}$	319	86.9
2	$A_{ave}^{(j)}$, $A_{cov}^{(j)}$, $CV_{ave}^{(j)}$, $CV_{cov}^{(j)}$, $BC_{ave}^{(j)}$, $BC_{var}^{(j)}$, AVG, $FC_{ave}^{(j)}$, $FC_{var}^{(j)}$	367	86.9

表 2: MFCC，GFCC，ケプストラム係数を加えた場合の正答率

特徴量の組み合わせ	次元数	正答率 [%]
$A_{ave}^{(j)}$, $A_{cov}^{(j)}$, $CV_{ave}^{(j)}$, $CV_{cov}^{(j)}$, $BC_{ave}^{(j)}$, $BC_{var}^{(j)}$, AVG	239	85.6 (基準値)
$A_{ave}^{(j)}$, $A_{cov}^{(j)}$, $CV_{ave}^{(j)}$, $CV_{cov}^{(j)}$, $BC_{ave}^{(j)}$, $BC_{var}^{(j)}$, AVG, $MC_{ave}^{(j)}$, $MC_{var}^{(j)}$, $MC_{skw}^{(j)}$, $GC_{ave}^{(j)}$, $GC_{var}^{(j)}$, $GC_{skw}^{(j)}$, $CC_{ave}^{(j)}$, $CC_{var}^{(j)}$, $CC_{skw}^{(j)}$	362	89.2

4.2 MFCC，GFCC，ケプストラム係数を加えた実験

基準値に対し，MFCC，GFCC，ケプストラム係数をそれぞれ組み合わせる実験を行う．表 2 に，正答率が最大となった時の特徴量の組み合わせとその次元数を示す．結果より，基準値における特徴量の組み合わせに，MFCC $MC_k^{(j)}$ から平均，分散，歪度，GFCC $GC_k^{(j)}$ から平均，分散，歪度を，ケプストラム係数 $CC_k^{(j)}$ から平均，分散，歪度を加えた時，次元数が 362 となり，正答率が最大で 89.2%となった．これは，基準値に対し正答率は 3.6%，次元数は 123 増加している．

4.3 すべての楽曲特徴量を組み合わせた場合の正答率

本研究におけるすべての楽曲特徴量の中から組み合わせを変えて実験を行い，正答率が最も高くなる組み合わせを調査した．表 3 に正答率が最高となったときの楽曲特徴量の組み合わせと次元数，その正答率を示す．また，表 4 にすべての楽曲特徴量を組み合わせた場合の分類結果を示す．なお，この表の縦軸は真のクラスを表し，横軸は予測されたクラスを表す．

結果より，次元数 485 のとき，正答率が最大で 91.9%となった．これは，本研究における最高正答率である．また，先行研究 [3] における最高正答率 87.1%，次元数 210 と比較すると，正答率では 4.8%，次元数では 275 次元増加した．

表 4 それぞれのジャンルにおける正答率に着目すると，Reggae や Rock においては従来手法と比べ 10%近く増加しており，またそれ以外のジャンルにおいては正答率が 90%を超えていることがわかる．

4.4 バンド間相関係数の平均 $BC_{ave}^{(j)}$ に対して主成分分析を行い他の特徴量と組み合わせる

表 5 に，主成分分析を行っていない場合の基準値と，主成分分析を行い累積寄与率が約 80%のとき，約 90%のとき，および正答率が最高値のときの，次元数と正答率を示す．

結果より，第 12 主成分まで用いたとき，合計次元数は 160，正答率は 86.6%であった．これは基準値と比較すると，次元数は約 67%削減でき，正答率は 1.0%増加した．

5. おわりに

本論文では，冗長ウェーブレット変換を用いた楽曲特徴量の抽出手法に対して，新たな特徴量抽出の手法と，正答率を維持しつつ次元削減を行う手法を提案した．

フレーム間相関係数を導入した実験では，隣接バンド幅を 0 から 7 まで変化させ特徴量を組み合わせた場合，B=1 のとき 319 次元で 86.1%の正答率が得られた．MFCC，GFCC，ケプストラム係数をそれぞれ組み合わせる実験では，次元数が 362 のとき正答率が最大で 89.2%となり，3 種類のケプストラム係数の優位性が示された．また，本論文における最も高い正答率は，すべての楽曲特徴量を組み合わせた場合の 91.9%で，次元数は 485 次元であった．

主成分分析を用いた次元削減手法では，次元数を 160 まで削減し正答率は 86.6%であった．先行研究が 210 次元で 87.1%であったのに対し正答率では及ばなかったものの，基準値に対して正答率を高めたうえで，先行研究よりも次元を削減することができた．

主成分分析を用いた次元削減手法はバンド間相関係数に有効であると考えられるため，同様に相関係数を用いているフレーム間相関係数に対しての適用を検討する必要がある．フレーム間相関係数は隣接バンド幅を大きくするほど次元数が増加するため，次元削減の効果が得られやすいと考えられる．また，すべての特徴量に対し，種類別に主成分分析を行ったうえで次元削減を行い特徴量を統合して分類を行うことで，さらなる次元削減と正答率の向上が期待できると考える．

表 3: すべての楽曲特徴量を組み合わせた場合の最高正答率

特徴量の組み合わせ	次元数	正答率 [%]
$A_{ave}^{(j)}$，$A_{cov}^{(j)}$，$A_{skw}^{(j)}$，$CV_{cov}^{(j)}$，$CV_{var}^{(j)}$，$CV_{krt}^{(j)}$，$BC_{ave}^{(j)}$，$BC_{var}^{(j)}$，AVG，$S_{ave}^{(j)}$，$K_{ave}^{(j)}$，$K_{krt}^{(j)}$，$FC_{ave}^{(j)}$，$FC_{var}^{(j)}$，$MC_{ave}^{(j)}$，$MC_{skw}^{(j)}$，$GC_{ave}^{(j)}$，$GC_{var}^{(j)}$，$GC_{krt}^{(j)}$，$CC_{ave}^{(j)}$，$CC_{skw}^{(j)}$	485	91.9

表 4: すべての楽曲特徴量を組み合わせた場合の分類結果

Genres	Bl	Cl	Co	Di	Hi	Ja	Me	Po	Re	Ro
Blues	94	0	2	0	0	3	0	0	0	1
Classical	0	96	0	0	0	4	0	0	0	0
Country	0	0	91	4	0	2	0	0	0	3
Disco	0	0	2	90	2	0	0	1	3	2
Hiphop	1	0	0	1	92	0	3	1	2	0
Jazz	0	4	0	1	0	95	0	0	0	0
Metal	0	0	0	2	0	0	94	0	0	4
Pop	0	0	0	0	0	0	0	98	2	0
Reggae	1	0	1	0	0	0	0	10	84	4
Rock	2	0	4	4	3	0	0	0	2	85

表 5: バンド間相関係数の平均 $BC_{ave}^{(j)}$ に対して主成分分析を行い他の特徴量と組み合わせたときの正答率

用いた主成分	累積寄与率 [%]	次元数	正答率 [%]
なし (基準値)	-	239	85.6
第 1〜第 12	77.55	160	86.6 (最高値)
第 1〜第 14	80.67	162	86.5
第 1〜第 25	90.52	173	85.6

参考文献

[1] J.H. Su, T.P. Hong, J.Y. Li, and J.J. Su, “Personalized Content-Based Music Retrieval by User-Filtering and Query-Refinement,”IEEE Conference on Technologies and Applications of Artificial Intelligence, Dec. 2018.

[2] 吉井和佳，後藤真孝，“音楽情報処理技術の最前線 : 7. 音楽推薦システム，” 情報処理，vol.50，no.8，pp.751-755，Aug. 2009.

[3] 小林拓哉，久保田彰，“音楽ジャンル分類における冗長ウェーブレット係数のバンド間相関に基づく楽曲特徴量抽出，” 映情学技報，vol.43，no.4，pp.43-46，Feb. 2019.

[4] O. Lartillot, “MIRtoolbox1.6.3 User’s Manual,” Apr. 2017.

[5] 吉井和佳，後藤真孝，“音楽情報処理技術の最前線 : 7. 音楽推薦システム，” 情報処理，vol.50，no.8，pp.751-755，Aug. 2009.

[6] 宮埜壽夫，大山正，心理学研究法（6），pp.35-64，誠信書房，日本，2015.

[7] G. Tzanetakis, and P. Cook, “Musical Genre Classification of Audio Signals, ”IEEE Transactions on Speech and Audio Processing, vol.10, no.5，pp.293-302，Jul. 2002.

[8] Zhouyu Fu, Guojun Lu, Kai Ming Ting, and Dengsheng Zhang, “A Survey of Audio-Based Music Classification, ”IEEE Transactions on Multimedia, vol.13, no.2, Apr. 2011.

[9] Y. V. Srinivasa Murthy and Shashidhar G. Koolagudi, “Content-Based Music Information Retrieval (CB-MIR) and Its Applications toward the Music Industry: A Review, ”ACM Computing Surveys, vol.51, no.3, Article 45, pp.1-46, May 2019.