

時系列ウェーブレット変換を用いた楽曲推薦の有効性の検証

A Study on Music Recommendation System
Using the Time-series Discrete Wavelet Transform徳山 竜也†* 吉富 康成†† 田仲 雅和‡ 田伏 正佳††
Tatuya Tokuyama Yasunari Yoshitomi Masakazu Tanaka Masayoshi Tabuse

1. まえがき

近年、携帯音楽プレーヤや Web 上でのオンライン楽曲販売サービスの普及により、音楽への関心が高まってきた。しかし、膨大な量の楽曲から好みの楽曲を探し出すのは容易でないため、ユーザにふさわしい楽曲を推薦する仕組みが求められている。楽曲推薦を実現する方策として、大別すると、協調フィルタリングを用いた手法[1]と楽曲の属性に基づく手法[2-4]が提案されている。

近年、楽曲の特徴量の抽出にウェーブレット変換を用いた研究が行われてきた[5-7]。本稿では、著者らが既報[7]で提案した、時系列ウェーブレット変換を利用した楽曲特徴量と、従来から利用されてきた MFCC(Mel-Frequency Cepstrum Coefficient)の時系列との比較を楽曲推薦に対する有効性という観点で行なった。また、ユーザによる楽曲評価から得た嗜好ベクトルと時系列ウェーブレット変換を利用した特徴量、時系列 MFCC、リズム特性を表す特徴量を利用した楽曲推薦システムを構築し、その性能を検証した。

2. 特徴量抽出

本報では楽曲の特徴量として、ウェーブレット変換を用いて得られた特徴量、MFCC の特徴量、リズム特性を表す特徴量の三種類を用いた。

2.1 ウェーブレット変換の特徴量

楽曲に離散ウェーブレット変換を施し、その多重解像度表現部のウェーブレット係数のヒストグラムをとると 0 を中心とした分布となる[8]。著者らは、楽曲及びその開始からの時間帯により、この標準偏差がかなり異なる場合があることを見出した。そこで、楽曲の音色や音量を表す特徴量として、音声信号を離散ウェーブレット変換によって多重解像度解析した各係数の標準偏差を用いた。まず、窓幅に区切った音声信号 $f(n)$ を、レベル 0 のスケーリング係数 $s_k^{(0)}$ とみなす。次に、レベル j のスケーリング係数 $s_k^{(j)}$ 、およびウェーブレット係数 $w_k^{(j)}$ を式(1)、(2)を使用して、 $s_k^{(0)}$ から、逐次、レベル J まで求める。

$$s_k^{(j)} = \sum_n p_{n-2k} s_n^{(j-1)} \quad (1)$$

$$w_k^{(j)} = \sum_n q_{n-2k} s_n^{(j-1)} \quad (2)$$

図 1 にスケーリング係数が徐々にスケーリング係数とウェーブレット係数とに分解される様子を示した。

本研究では、 p_k としてドベシーのスケーリング関数を表す数列を用いた。 q_k は式(3)から求められる。

$$q_k = (-1)^k p_{1-k} \quad (3)$$

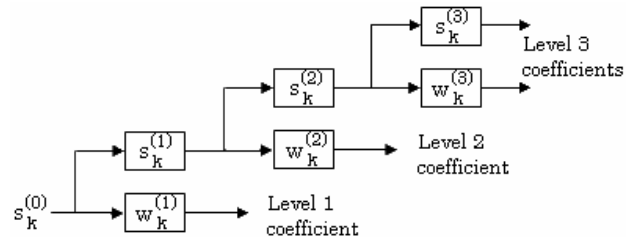


図1 多重解像度解析

ここまでの処理を窓幅ごとにスライドさせながら、楽曲に対して順次行う。ここで得られた $1 \sim J$ 次までのウェーブレット係数と J 次のスケーリング係数の各標準偏差の時系列データを楽曲の特徴ベクトルの各要素とする。

2.2 MFCC の特徴量

フレーム毎に音響信号の高速フーリエ変換(FFT)で得られるパワースペクトルに対してメルスケールのフィルタバンクを施し、周波数軸変換されたパワースペクトルに対して離散コサイン変換(DCT)を実行することにより MFCC 特徴量が抽出される。12 次元 MFCC, 1 次元対数パワー, 12 次元 Δ MFCC, 1 次元 Δ 対数パワーの各平均と各標準偏差の時系列データを楽曲の特徴ベクトルの各要素とする。

2.3 リズム特徴量

楽曲をジャンル分けする論文で提案されたリズム特性を表す特徴量[9]を採用した。そして、主のビートの強さ、リズムの規則正しさ、主のビートとサブビートの関係などを特徴ベクトルによって表す。まず、ウェーブレット変換の多重解像度解析を利用して、4 個の周波数バンドに分解する。続いて、各周波数バンドに対して、ローパスフィルタなどの前処理を施す。次に、各バンドを逆ウェーブレット変換し、自己相関を計算する。自己相関におけるピークをビートヒストグラムとして音声ファイル全体で蓄積する。ビートヒストグラムにおける次の量を、リズム特性を表す特徴量として用いる。

- A0, A1 : 1, 2 番目のピークの相対振幅
- P1, P2 : 1, 2 番目のピークの時間
- RA : 振幅の比率 A1 / A0
- SUM1, SUM2, SUM3 : BPM (Beats Per Minute) が 40~90, 90~140, 140~250 であるヒストグラムの合計

3. 楽曲推薦システム

本章では 2 章で述べた特徴量を使った楽曲の推薦方法について述べる。本報での楽曲推薦とは、対象ユーザが評価を与えていない楽曲に対して、推薦するかどうかを決定することである。いま、全楽曲のインデックス集合を $M = \{m | 1, \dots, N_M\}$ 、ユーザが主観評価 s ($1 \leq s \leq 5$) (好き: 5 ~ 嫌い: 1) を与えた楽曲のインデックス集合を $M_s = \{m_s | 1, \dots, N_{M_s}\}$ ($M_s \subseteq M$) とする。ここで N_M と N_{M_s} は全楽曲数と評価済み楽曲数を表す。

†京都府立大学, Kyoto Prefectural University
 ††京都府立大学大学院, Graduate School of Kyoto Prefectural University
 ‡株式会社 NTT データ, NTT DATA Corp.
 *株式会社松風, SHOFU INC. (現在)

3.1 1種類の特徴ベクトルを用いた方法

全評価済み楽曲と対象未評価楽曲 m' の合計 $[N_{Ms} + 1]$ 曲の特徴ベクトルを主成分分析し、累積寄与率が 80%以上になるように l 個の成分を選択する。そして、評価済み楽曲 m の特徴ベクトル $c_m = (c_{m,1}, \dots, c_{m,l})$ と m' の特徴ベクトル $c_{m'}$ との類似度(=ユークリッド距離の逆数)を計算する。

その結果、類似度が最大となる評価済み楽曲 m^* を N_{Ms} 曲の中から決定し、その評価済み楽曲のスコア s を m' のスコアとする。決定したスコアが 4 以上であるとき m' を推薦する。

3.2 複数種類の特徴ベクトルを組み合わせた方法

3.1 章と同様にして、二または三種類の特徴ベクトルそれぞれで対象未評価楽曲 m' のスコアを決定する。続いて、次の A~F の六種類の手法で推薦するかどうかを決定する。

二種類の特徴ベクトルだけを用いる、以下の A~C の三種類の手法は、「1 種類の特徴ベクトル(「主特徴ベクトル」と表記)でスコアが 4 以上となり、かつ、他方の特徴ベクトルで、平均以上の場合に「推薦」、という考え方で設定した。そして、「平均」の基準を以下のように 3 種類(スコア 3, 順位平均, 類似度平均)設けた。

- A: 各特徴ベクトルで求めたスコアが、主特徴ベクトルで 4 以上で、他方の特徴ベクトルで 3 以上であれば推薦する。
- B: 主特徴ベクトルでスコアが 4 以上であり、かつ主特徴ベクトルで類似度最大となった楽曲において、他方の特徴ベクトルにおいて全評価楽曲の中での類似度の順位が上位 $[N_{Ms}/2]$ 位以内であった場合推薦する。
- C: 主特徴ベクトルでスコアが 4 以上であり、かつ主特徴ベクトルで類似度最大となった楽曲において、他方の特徴ベクトルにおいて求めた類似度が全評価楽曲の平均より大きければ推薦する。

三種類の特徴ベクトルを用いる手法においては、上記 A,B,C における「他方」を「他の二種類」に置き換えて、手法 A,B,C に対応して各々手法 D, E, F と表記する。

4. 評価実験

4.1 条件

まず、RWC 研究用音楽データベース(音楽ジャンル)[10]の 100 曲に対して、10 名のユーザごとの 5 段階の主観評価を付与した。

ウェーブレット変換の特徴量を求める際のウェーブレットの窓幅を 524288 サンプル(約 11 秒)とし、楽曲開始から 10s ~ 65s までの信号を利用した。楽曲開始直後は、無音の時間がある場合があるので、楽曲開始 10s から、窓幅の 5 倍にあたる 65s まで 5 つの時間区間に分けサンプリングを行った。上記の窓幅および録音時間については実験的に決定した。ウェーブレット変換の関数はドベシーの $N = 4$ とし、レベル 7 で特徴量を作成した。5 つの時間区間各々について、レベル 1~7 のウェーブレット係数とレベル 7 のスケール係数の各々における標準偏差が計 8 個得られるので、40 次元の特徴ベクトルが得られる。

MFCC 特徴量を求める際は、窓幅を 30ms, シフト幅を 10ms とし、窓関数としてハミング窓を使用した。またフィルタバンク数を 24 とし、楽曲開始から 10 ~ 65 秒までの信号を利用した。特徴量として、11 秒間の信号に対して算

出された 12 次元 MFCC, 1 次元対数パワー, 12 次元 Δ MFCC, 1 次元 Δ 対数パワーの計 26 次元 MFCC の平均と標準偏差を用いた。11 秒間につき 52 個の特徴量が得られるので、55 秒の信号に対して 260 次元の特徴ベクトルが得られる。

リズム特徴量を求める際は、文献[9]と同様に信号を 22050Hz に変換し、ウェーブレットの窓幅は、65536 サンプル(約 3 秒), 跳躍サイズ(ウェーブレット変換の開始点をずらす幅)を 32768 サンプルとして、楽曲開始から 10s ~ 40s までの信号を利用した。ウェーブレット変換の関数はドベシーの $N = 4$ とした。特徴ベクトルの次数は、後述する 4.2 節に記した 3 通りの条件で各々定まる。

4.2 評価方法

3 章記載のシステムで楽曲推薦を行う。リズム特徴量については予備実験で高い推薦正答率を得ることができた a:[SUM1], b:[A0, A1, SUM1, SUM2, SUM3], c:[A0, A1, P1, P2, RA, SUM1, SUM2, SUM3] の 3 種類を用いて、計 74 通り(表 1)の推薦正答率を算出し、そのうち最良であった推薦方法の結果をシステムの推薦正答率とした(実用上は、最良であった方法だけを使うので、本報では、その結果だけを示す)。そして、以下の 3 つの条件でシステムの評価を行った。図 2, 3 に各条件での処理フローを示す。

条件 1: 評価既知曲 99 曲に対して評価既知曲 1 曲を未知曲として取り扱い、推薦するかを判定する。これをすべての組み合わせについて行う。

条件 2: 評価既知曲 98 曲に対して評価既知曲 1 曲を未知曲として取り扱い、判定する。これを 99 曲に対して行ない、そのとき選択された最良方法で残り 1 曲を推薦するかを判定する。これをすべての組み合わせについて行う。

条件 3: 条件 2 で推薦すると判定された楽曲のうち、主特徴ベクトルにおいて類似度が 2 位となった楽曲のスコアが 4 以上の場合のみ推薦する。

条件 2, 3 の場合、未知曲を推薦するか否かを決めるのに用いる方法は、既知楽曲の構成に依存する。例えば、1 曲目を未知曲として取り扱う場合、2~100 曲目の楽曲の特徴量とユーザのスコアを用い、最良方法が選ばれる。

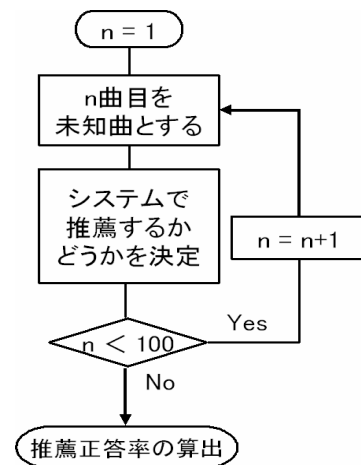


図 2 条件 1 での処理フロー

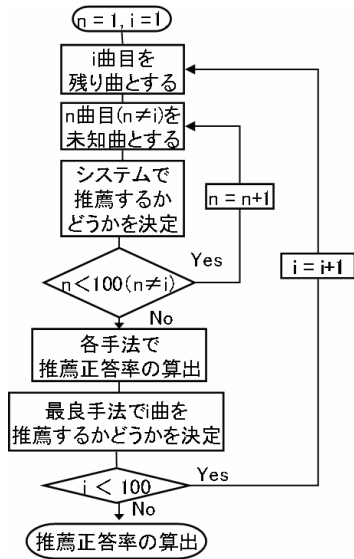


図3 条件 2,3 での処理フロー

4.3 結果と考察

各ユーザに対して、各条件で、判定が「推薦する」であった曲の内、ユーザの主観評価が 4 以上であった曲の割合である「推薦正答率」を用いてシステムの評価を行う。

時系列ウェーブレット変換の特徴量と時系列 MFCC の特徴量を推薦正答率で比較した結果を表 2 に示した。どちらの特徴量を推薦に使うのがよいかは、ユーザにより異なった。推薦正答率の平均値で比較すると、時系列ウェーブレット変換の特徴量の方がランダムに比べて 14.19%高く、かつ、時系列 MFCC を特徴量に用いる場合より高かった。これらのことから、時系列ウェーブレット変換の特徴量は楽曲推薦に有効と考えられる。なお、ユーザの主観評価が 4 以上であった割合(表 3)と推薦正答率(表 2)との間に明確な関係はみられなかった。

また、時系列ウェーブレット変換を利用した特徴量と時系列 MFCC 特徴量とリズム特徴量の三種類の特徴量を利用して楽曲推薦システムを構築した結果、約 67.53%とランダムに比べて平均 35.53 %高い正答率で推薦することができた(表 4, 条件 1)。98 曲を既知、1 曲を未知とし、その最良方法で残りの曲を推薦するかどうかを判定する方法では、55.8%と条件 1 よりも少し推薦正答率の平均値が低い結果となった(表 4, 条件 2)。しかし主特徴ベクトルを用いて計算される類似度 2 位の楽曲スコアも 4 以上のときに推薦するという方式をとると、67.16%と高い推薦正答率の平均値が得られた(表 4, 条件 3)。

そして、条件 1 では、一種類の特徴ベクトルだけを利用するのと比較して、本システムを用いることで、推薦正答率の平均値が 20%以上向上した(表 2, 4)。また、条件 2,3 では、一種類の特徴ベクトルを利用する方法が最頻度で選択されたのは 10 人中 1 人だけだった(表 6)。これらのことから、本システムの有効性を示せたと考えられる。また、時系列ウェーブレット変換を利用した特徴量が、条件 1 では 10 人中 8 人で、条件 2,3 では 10 人中 7 人で、主または副の特徴ベクトルとして選択された。このことから、時系列ウェーブレット変換を利用した特徴量は、本システムにおいて有用であると考えられる。

本システムで選ばれる最頻度の方法およびその特徴ベクトルはユーザに依存した(表 5,6)。この現象の詳細な解析を今後行う予定である。

今回は全ユーザが 10~42%の曲に 4 以上の主観評価を与えた。評価既知曲の中に 4 以上の主観評価の曲がなければ、未知曲に対して「推薦する」という判定はでない。本システムを実際に用いる際には、評価既知曲のうち 1 曲を未知曲として取り扱い、条件 2,3 と同様にして、各ユーザにとって最良の推薦方法を、74 種類(表 1)の方法の中から本システムが自動的に 1 つ選択する。そして、その方法で実際の未知楽曲を推薦するかを判定し、推薦した場合には、そ

表 1 本システムで自動選択の対象となる方法

No.	手法	特徴ベクトル		
		主	主以外	
1 2 3 4 5	一種類の 特徴量	ウェーブレット		
		MFCC		
		リズム a		
		リズム b		
		リズム c		
6 7 8 9 10 11 12 13 14 15 16 17 18 19	ウェーブレット		MFCC	
			リズム a	
			リズム b	
			リズム c	
	A	MFCC		ウェーブレット
				リズム a
				リズム b
				リズム c
	B	リズム a		ウェーブレット
C	リズム b		MFCC	
20 ~33 34 ~47 48 49 50 51 52 53 54 55 56	B	(No.6~19 と同じ組合せ)		
		C	(No.6~19 と同じ組合せ)	
			D	ウェーブレット
				MFCC, リズム b
		MFCC, リズム c		
	E	MFCC		ウェーブレット, リズム a
				ウェーブレット, リズム b
				ウェーブレット, リズム c
		F	リズム a	ウェーブレット, MFCC
リズム b				
リズム c				
57 ~65 66 ~74	E	(No.48~56 と同じ組合せ)		
	F	(No.48~56 と同じ組合せ)		

表2 条件1における各特徴ベクトルでの推薦正答率(%)とランダムとの差(%)

特徴ベクトルと評価 ユーザ	ウェーブレット		MFCC	
	推薦正答率	ランダムとの差	推薦正答率	ランダムとの差
1	43.9	3.9	65	25
2	53.3	24.3	35	6
3	51.7	17.7	58.1	24.1
4	47.8	5.8	51.1	9.1
5	44.8	15.8	20.7	-8.3
6	28.6	1.6	50	23
7	51.4	17.4	46.7	12.7
8	55.6	20.6	46.5	11.5
9	40.4	0.4	42.1	2.1
10	44.4	34.4	14.3	4.3
平均	46.19	14.19	42.95	10.95

表3 ユーザの評価が4以上であった楽曲の割合(%)

ユーザ	1	2	3	4	5	6	7	8	9	10
評価4以上(%)	40	29	34	42	29	27	34	35	40	10

表4 各条件における推薦正答率(%)とランダムとの差(%)

ユーザ	条件1		条件2		条件3	
	推薦正答率	ランダムとの差	推薦正答率	ランダムとの差	推薦正答率	ランダムとの差
1	80	40	55.6	15.6	83.3	43.3
2	57.1	28.1	40	11	28.6	-0.4
3	84.6	50.6	66.7	32.7	100	66
4	70	28	59.1	17.1	60	18
5	58.8	29.8	44.4	15.4	30.8	0.8
6	71.4	44.4	50	23	60	33
7	64.3	30.3	84.6	50.6	100	66
8	77.8	42.8	68.8	33.8	71.4	36.4
9	54.2	14.2	35.5	-4.5	37.5	-2.5
10	57.1	47.1	57.1	47.1	100	90
平均	67.53	35.53	55.8	24.18	67.16	35.06

表5 条件1で選択された方法と特徴ベクトル

ユーザ	手法	特徴ベクトル	
		主	主以外
1	B	MFCC	リズム b
2	B	ウェーブレット	リズム c
3	C	MFCC	リズム c
4	E	MFCC	ウェーブレット, リズム a
5	C	ウェーブレット	リズム a
6	E	MFCC	ウェーブレット, リズム a
7	E	MFCC	ウェーブレット, リズム b
8	E	ウェーブレット	MFCC, リズム b
9	B	リズム a	ウェーブレット
10	E	ウェーブレット	MFCC, リズム a

表6 条件2,3で最頻度で選択された方法と特徴ベクトル

ユーザ	手法	特徴ベクトル	
		主	主以外
1	B	MFCC	リズム c
2	一種類の特徴ベクトル	ウェーブレット	
3	B	MFCC	リズム a
4	B	MFCC	リズム b
5	D	ウェーブレット	MFCC, リズム c
6	D	ウェーブレット	MFCC, リズム a
7	E	リズム b	ウェーブレット, MFCC
8	F	MFCC	ウェーブレット, リズム b
9	B	リズム a	ウェーブレット
10	E	ウェーブレット	MFCC, リズム a

の楽曲の評価をユーザにってもらうことで評価既知曲を1曲増やし、本システムが最良の推薦方法を自動的に再選択する。このように、新たに推薦した曲に対するユーザ評価をシステムの性能改善に逐次利用できる。

5. まとめ

時系列 MFCC 特徴量と既報[7]で提案した時系列ウェーブレット変換を利用した特徴量とを、楽曲推薦という観点で比較し、後者の有効性を示した。そして、これら2種類の特徴量とリズム特徴量を用いた楽曲推薦システムを提案し、その有効性を示した。そして、時系列ウェーブレット変換を利用した特徴量の本システムにおける有用性を示した。本システムに協調フィルタリングを付加することを今後検討する。

参考文献

- [1] J. Breese, D. Heckerman, C. Kadie, "Empirical analysis of predictive algorithms for collaborative filtering", Proc. UBI, pp. 43-52, 1998.
- [2] B. Logan, "Music recommendation from song sets", Proc. of ISMIR, pp.425-428, 2004.
- [3] 帆足 啓一郎, 井ノ上 直己, "ユーザの音楽嗜好に基づく音楽情報検索手法", 情報処理学会研究報告, MUS[音楽情報科学研究会報告], vol.49, pp.79-84, 2003.
- [4] 奥乃 博, 北原 鉄朗, 吉井 和佳, "楽曲の特徴量抽出と検索技術", 電気学会誌, vol.127, no.7, pp.417-420, 2007.
- [5] G. Li, A. Khokhar, "Content-based indexing and retrieval of audio data using wavelets", Proc. of ICME 2000, vol.2, pp. 885-888, 2000.
- [6] T. Li, M. Ogihara, Q. Li, "A comparative study on content-based music genre classification", Proc. of ACM SIGIR 2003, pp.282-289, 2003.
- [7] 田仲 雅和, 吉富 康成, "ウェーブレット変換を用いたユーザの嗜好に基づく楽曲推薦システム", 第7回情報科学技術フォーラム一般講演論文集, vol.2, pp.245-248, 2008.
- [8] 村田 真一, 吉富 康成, 石井 博昭, "ウェーブレット変換を用いた音声電子透かし埋め込み位置の最適化", 日本オペレーションズ・リサーチ学会秋季研究発表会アブストラクト集, pp.210-211, 2007.
- [9] G. Tzanetakis, P. Cook, "Musical genre classification of audio signals", IEEE Trans. Speech Audio Process, vol.10, no.5, pp.293-302, 2002.
- [10] 後藤 真孝, 橋口 博樹, 西村 拓一, 岡 隆一, "RWC 研究用音楽データベース: 研究目的で利用可能な著作権処理済み楽曲・楽器音データベース", 情報処理学会論文誌, vol.45, no.3, pp.728-738, 2004.