2008－AVM－60 (1)
2008／3／6

楽曲推薦システムのための楽曲波形と歌詞情報を考慮した類似楽曲検索に関する一検討

舟澤 慎太郎† 北市 健太郎† 甲藤 二郎†

†早稲田大学理工学部コンピュータネットワーク工学科 〒169-8555 東京都新宿区大久保 3-4-1

E-mail: †shint@katto.comm.waseda.ac.jp, katto@katto.comm.waseda.ac.jp

あらまし

本稿では，楽曲の音響的特徴と歌詞情報を利用した類似楽曲検索手法を述べる．典型的な Content-based 楽曲検索において，楽曲の特徴はベクトルで表現され，それらの類似度を基に検索をする．この際に，楽曲の特徴ベクトルを構成する要素には，“ユーザが楽曲から受ける印象に影響を与える特徴量”を用いることで，対応できる検索要求の幅を広げ，検索精度の向上につながる．そこで本稿では，楽曲波形から得られる音響的特徴量に加えて，歌詞情報を特徴量として利用する．そして，ユーザによる評価実験を行い，この手法の有効性を評価した．

A Study on Similar Music Retrieval Using both Lyrics and Content for Music Recommendation System

Shintaro FUNASAWA† Kentaro KITAICHI† Jiro KATTO†

† Department of Computer Science, School of Science and Engineering, Waseda University

3-4-1 Okubo, Shinjuku-ku, Tokyo, 169-8555 Japan

E-mail: †shint@katto.comm.waseda.ac.jp, katto@katto.comm.waseda.ac.jp

Abstract

This paper presents similar music retrieval that uses acoustic features and lyrics features of musical pieces. In typical content-based music retrieval, the features of music pieces are expressed by vectors, and music pieces are retrieved based on their similarities. When generating the features vectors, using “features that influence impression that users receive from music” as elements improves retrieval accuracy. Therefore, in this paper, we use lyrics features in addition to acoustic features that are extracted from music signals. Furthermore, we estimate the efficacy of this music retrieval method by a subjective evaluation experiment.

1 はじめに

近年，ネットワーク技術の発展や音楽配信サービスの普及により，大量の楽曲データに手軽にアクセスできるようになった．また，MP3 や WMA，Ogg などのオーディオデータ圧縮技術の開発や，ハードウェア，メモリの小型大容量化などにより，ユーザは大量の楽曲データを持ち歩くことが可能となった．

こうした背景は，ユーザにとって楽曲データを選ぶ選択肢が増えるというメリットがある一方で，ユーザの好みに合った楽曲データを大量のデータの中から検索するのが困難になっている．そこで，自動でユーザの嗜好に合う楽曲を検索・提示してくれる，楽曲推薦システムの研究がなされている．

典型的な Content-based 楽曲推薦において，楽曲の特徴はベクトルで表現される．それらに対しなんらかの類似度尺度を用いて，楽曲間の類似度を算出し，ユーザが過去に高い評価をした楽曲データと類似度の高い楽曲を推薦曲として出力する．

このような手法では，最終的に類似度を基にして検索を行うため，ユーザの嗜好を適切に類似度に反映することが直接的に検索精度に影響を与える．故に，楽曲の特徴ベクトルを構成する要素として“ユーザの楽曲から受ける印象に影響を与える特徴量”を用いることが重要となる．こうすることで，より幅広い検索要求に対応でき，検索精度の向上につながる．

そこで本稿では，楽曲波形から算出される音響的特徴量を要素とした音響特徴ベクトルと共に，歌詞情報を基に生成した歌詞特徴ベクトル

を利用した類似楽曲検索手法を述べる．

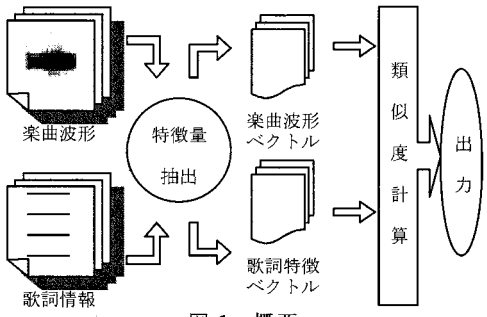

図 1　概要

2　Content-based 楽曲検索における歌詞情報の利用

参考論文[1][2]では，音響的特徴と歌詞情報を用いてアーティストの分類を行っている．この論文では歌詞情報を，名詞の TF*IDF 値，使用単語の長さや行の長さ，品詞の使用頻度などを用いて表現している．ただし，使用単語や行の長さ，品詞の種類などは，歌詞における文法的な特徴を表すもので，意味的な特徴は十分には表していない．故に，“ユーザの楽曲から受ける印象に影響を与える特徴量”とは考えがたい．よって本稿では，歌詞情報を表現するベクトルを，文法的特徴でなく，より意味的特徴を表現すると考えられる，名詞の TF*IDF 値のみを用いて構成する．

3　音響的特徴と歌詞情報を用いた類似楽曲検索

3.1　音響的特徴ベクトルの生成

楽曲の音響的特徴を表現するベクトルは，楽曲波形から抽出される以下の特徴量によって構成される．これらの特徴量は MATLAB の関数ライブラリ群 MIRtoolbox[4][5]を利用して抽出した．

- **MFCC (Mel-Frequency Cepstrum Coefficient)**

メル周波数ケプストラム係数とよばれる，対数スケール上のスペクトル包絡情報を表現する特徴量である．MFCC では包絡情報は低次の係数に表現されるので，最初から 13 次元までの値を用いる．

- クロマベクトル

スペクトルを 12 要素からなるベクトルに圧縮したもので，各次元は 1 オクターブ内の 12 の音高に対応しており，各音高のパワーを表す．

- スペクトル重心

楽曲の持つスペクトル分布における重心の値である．スペクトル重心は，スペクトルの形状を表現する．

- スペクトル変動

楽曲をフレーム分割し，現フレームと 1 つ前のフレームとのパワーの変化量である．連続する全フレーム間においてこの値を抽出し，その平均を用いる．スペクトル変動は，時間的なスペクトルの変化を表す．

- スペクトル平坦度

楽曲の持つスペクトルの算術平均と幾何平均との比によって計算される値で，スペクトルの分布が滑らかどうかを表現する．

- スペクトルロールオフ

楽曲のスペクトルにおいて，より低い周波数からみたときの全体に対する 85%のエネルギーを占めている周波数の値である．スペクトルの高周波成分に関する特徴量である．

- ローエナジー

楽曲信号全体の持つエネルギーの平均値を下回るエネルギーを持つフレームの割合である．振幅の分布を表す．

本稿では，これらの値を平均が 0，標準偏差が 1 になるように正規化したものを特徴量として音響的特徴ベクトルを生成する．

3.2　歌詞特徴ベクトルの生成

楽曲の歌詞特徴を表現するベクトルは，以下の手順で生成する．

1)　歌詞データから名詞を抽出

全楽曲の歌詞データから使われている名詞を抽出する．名詞の抽出には形態素解析器 MeCab[6]を利用した．こうして抽出したキーワードにおいて，TF(m,i) (Term Frequency)と DF(i) (Document Frequency)を求めておく．TF(m,i)とは楽曲 m におけるキーワード i の使用回数のことで，一方 DF(i)とは，キーワード i が使用されている楽曲の数である．

2)　各楽曲の各キーワードにおける TF*IDF 値を算出

次に DF(i)から IDF(i) (Inverse Document Frequency)を計算する．IDF(i)は以下の式で定義され，キーワード i が使われていない頻度を意味する．N は総楽曲数である．

$$IDF(i)=\log_{10}\left(\frac{N}{DF(i)}\right)$$

こうして，各楽曲の各キーワードにおいて，TF と IDF の積（TF*IDF）を算出する．

3)　楽曲を特徴付けるキーワードを選定

TF*IDF は，その名詞が楽曲を特徴付けるキーワードであるかどうかを表す指標である．この TF*IDF の高い名詞を楽曲を特徴付けるキーワードとして選定する．

しかし，TF*IDF の高い名詞でもそれ自体意味を持たない不要語（英数字，“こと”，“もの”など）や，ひらがなやカタカナなどの表記法の違いによって本来同一視しなければならない語がある．これらの語は直接指定して適切な処理を行っている．

4)　歌詞特徴ベクトルの生成

3)で選定したキーワードにおける TF*IDF 値

を並べたものを歌詞の特徴ベクトルにする．すなわち，歌詞特徴ベクトルの各次元は各キーワードに対応する．

下式は，楽曲 m における歌詞特徴ベクトルを示す．式中で，抽出されたキーワードは n 個となる．

$$lyrics\ (m) = \{tf\ (m, i_1) * idf\ (i_1), \ldots, tf\ (m, i_n) * idf\ (i_n)\}$$

3.3 類似度算出法

楽曲間の音響的類似度 $SIM_{CONTENT}$ は以下の式で表される．

$$SIM_{Content} = \alpha SIM_{MFCC} + \beta SIM_{Chroma} + \gamma SIM_{Others}$$

上式における SIM_{MFCC}，SIM_{Chroma}，SIM_{Others} とはそれぞれ MFCC，クロマベクトル，それ以外の特徴量の類似度を示す．α，β，γ はそれぞれの重み付け係数である．

MFCC とその他特徴量における類似度は，下式で示されるコサイン距離により算出される．

$$\cos(\vec{x}, \vec{y}) = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2} \sqrt{\sum_{i=1}^{n} y_i^2}}$$

コサイン距離は，ある特徴量空間における 2 要素間の類似度を計算するために有効であり，値が大きいほど類似しているといえる．また，クロマベクトルの類似度はユークリッド距離を [-1,1]に正規化した値に-1 を掛けた値を用いた．

一方，楽曲 M_1，M_2 間の歌詞類似度はコサイン類似度により算出される．

$$SIM_{Lyrics}(\vec{M}_1, \vec{M}_2) = \cos(\vec{M}_1, \vec{M}_2)$$

そして，音響的特徴類似度と δ の重みを付けた歌詞類似度を足し合わせて，楽曲間の類似度 SIM とする．

$$SIM = SIM_{Content} + \delta SIM_{Lyrics}$$

4 評価実験

上記の手法で類似度を算出し，それを指標とした類似楽曲検索においての評価実験を行った．

4.1 Web 上の情報を用いた評価実験

4.1.1 音響的特徴のみを考慮した類似楽曲検索

音響的特徴のみを考慮した類似楽曲検索において，検索指標となる音響的類似度を変化させ，その検索結果への影響を観測し，評価した．比較対象となる手法は以下の 8 通りである．

i) SIM_{MFCC} を基に検索
ii) SIM_{Chroma} を基に検索
iii) SIM_{Others} を基に検索
iv) SIM_{MFCC} と SIM_{Chroma} を基に検索
v) SIM_{MFCC} と SIM_{Others} を基に検索
vi) SIM_{Chroma} と SIM_{Others} を基に検索
vii) $SIM_{Content}$($\alpha=\beta=\gamma=1$)を基に検索
viii) ランダムに検索

実験に利用した楽曲は一般の J-POP200 曲である．各楽曲において上記 8 手法で検索し，類似楽曲を出力する．ここで，検索された楽曲が正解かどうかの判定は，音楽のソーシャルネットワーキングサービスである Last.fm[7]における“テイストの似たアーティスト”情報を利用した．この情報は同サイト上で，“ユーザのリスニング傾向を基に算出している”とされており，Last.fm のユーザ数は国内で 40 万人以上（全世界では 2000 万人以上）に達していることから，多数のユーザの嗜好を十分に反映していると見なすことができる．

この実験の結果を図 2 に示す．縦軸は正解率を表している．横軸の出力楽曲数とは，類似楽曲として出力する楽曲の数である．複数の楽曲が出力された場合，1 曲でも正解曲が含まれていれば正解とした．また，各手法において出力楽曲数が 1 曲，3 曲，5 曲のときの正解率の和を Score として評価した．各手法における Score を表 1 に示す．

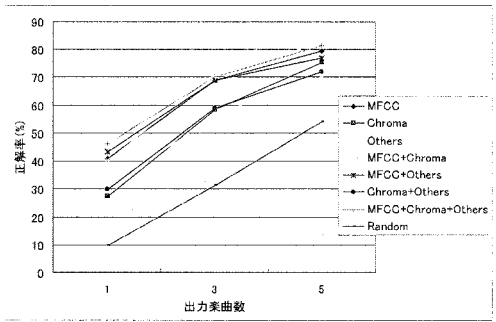

図 2 各手法における正解率

表 1 各手法における Score

	MFCC	Chroma	Others	MFCC+ Chroma
Score	1.895	1.615	1.575	1.94

	MFCC+ Others	Chroma+ Others	ALL	Random
Score	1.895	1.61	1.98	0.955

まず図 2 より，ランダムに検索した場合に比べて全ての手法で正解率が向上していることがわかる．また，表 1 の Score を比較すると，全ての特徴量を考慮した場合が最も優れている．しかし，MFCC とクロマベクトル以外の特徴量を追加しても精度にあまり変化が見られない．よってこれらの特徴量は，MFCC とクロマベクトルで十分に表現できていると推測できる．

4.1.2 音響的特徴と歌詞情報を考慮した類似楽曲検索

検索指標に音響的類似度，歌詞類似度，またそれらを組み合わせたものを用いた場合の検索結果への影響を観測し，評価した．比較対象となる手法は以下の 5 通りである．

ⅰ) $SIM_{Content}$ を基に検索
ⅱ) SIM_{Lyrics} を基に検索
ⅲ) $SIM_{Content}$ と SIM_{Lyrics} を足し合わせたものを基に検索
ⅳ) $SIM_{Content}$ と SIM_{Lyrics} を重み付けし足し合わせたものを基に検索
ⅴ) ランダムに検索

上記手法 ⅲ は各重み付け係数を $\alpha=\beta=\gamma=\delta=1$ とした場合である．また，手法 ⅳ においての重みは，それぞれの類似度を単独で検索指標に用いた場合の正解率を基にして算出した．実験に用いた楽曲データセット，評価法は 4.1.1 と同様である．

この実験の結果を図 3 に示す．また，4.4.1 と同様にして Score を算出し評価した．表 2 に各手法における Score を示す．

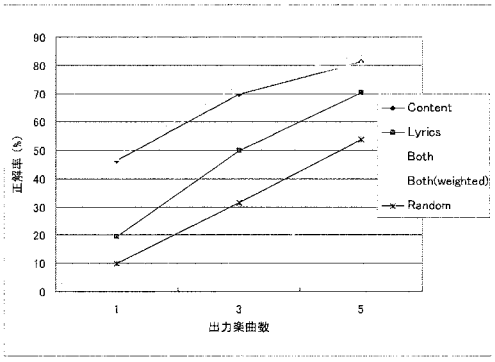

図 3　各手法における正解率

表 2　各手法における Score

	Content	Lyrics	Both	Both (weighted)	Random
Score	1.98	1.4	1.995	2.055	0.955

表 2 より，音響的類似度と歌詞類似度を単独で用いるより，組み合わせて用いたほうが正解率が良くなっていることがわかる．また，重み付けをすることにより多少ではあるが正解率が向上している．

しかし，音響的類似度と比較して歌詞類似度を用いた場合の正解率がかなり低くなっている．これは，ただ単純に歌詞情報を十分に抽出できていないか，正解に用いた“テイストの似たアーティスト”情報と歌詞との相関が音響的特徴に比べて低いためだと思われる．

4.2　被験者による評価実験

被験者に各手法で検索した類似楽曲について実際に評価してもらう実験を行った．

まず，被験者は自分の好きな楽曲を 1 曲選択し，それに類似する楽曲を各手法にて検索する．その結果出力された楽曲について以下の 3 つの質問に 5 段階で回答してもらい評価を行う．被験者は楽曲を視聴することができる．

質問 1 この楽曲は好きですか？
1.嫌い　2.好きではない　3.どちらでもない
4.どちらかと言えば好き　5.好き

質問 2 楽曲全体に関して，選択した楽曲と似ていると思いますか？
1.全く似ていない　2.あまり似ていない
3.どちらとも言えない　4.似ている
5.非常に似ている

質問 3 楽曲の情景に関して，選択した楽曲と似ていると思いますか？
1.全く似ていない　2.あまり似ていない
3.どちらとも言えない　4.似ている
5.非常に似ている

被験者数は 16 人で，比較手法は以下の 4 通りである．手法 ⅳ 中に記される SIM_{Tempo} とは，楽曲間の BPM の差を[-1,1]に正規化し，-1 を掛けた値である．これにより，楽曲間のテンポの差を考慮することができる．

ⅰ) $SIM_{Content}$ を基に検索
ⅱ) SIM_{Lyrics} を基に検索
ⅲ) $SIM_{Content}$ と SIM_{Lyrics} を足し合わせたものを基に検索
ⅳ) ⅲ に SIM_{Tempo} を足し合わせたものを基に検索

この実験の結果を図 4，5，6 に示す．各図中の出力楽曲数とは，各手法において類似度の高い上位何曲までのものに対する評価かを示す値である．

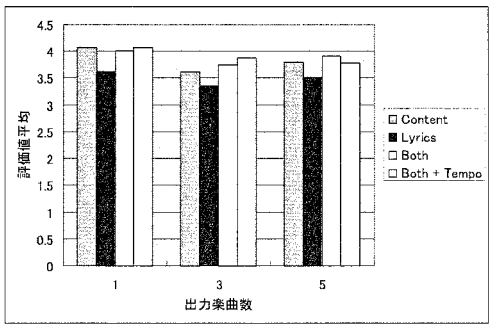

図 4　質問 1 に対する評価値平均

質問 1 に対する評価は，そのまま楽曲推薦の精度を表す．図 4 より，全体的に総合すると両

方の類似度を用いた場合の評価値が高くなっている．これは，音響的特徴と歌詞情報を共に考慮することの有効性を示している．また，歌詞類似度のみを用いた場合の評価値が全体を通して低くなっていることから，歌詞情報だけではユーザの嗜好を十分に表現できていないことがわかる．

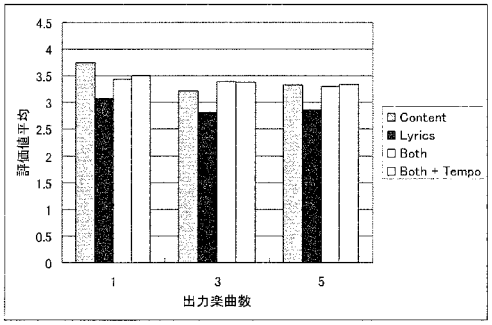

図 5　質問 2 に対する評価値平均

質問 2 に対する評価は，ユーザがある 2 つの楽曲に対し類似しているかどうかを判断するのに影響を与える特徴量を表す．図 5 より，音響的特徴を用いた場合に評価値が高い傾向にある．故に，主に音響的特徴が楽曲間の類似性を表現する要素となることがわかる．

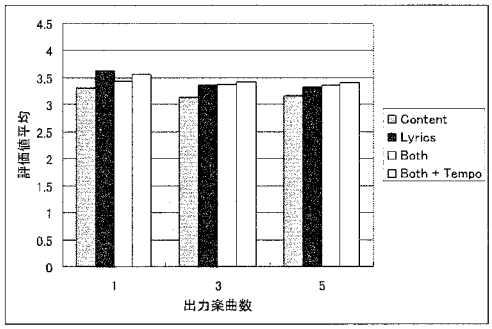

図 6　質問 3 に対する評価値平均

質問 3 に対する評価は，楽曲の情景を表現するのに影響を与える特徴量を表す．図 6 より，全体的に歌詞情報を用いた手法の方が評価値が高くなっている．これより，歌詞情報の方が楽曲の情景を表現するのに有効であると言える．

また，各手法において出力された上位 1，2，3 曲以内に好きな楽曲が含まれている被験者の数をカウントした．（表 3）具体的には，質問 1 に対して“5.好き”，または“4.どちらかと言えば好き”の評価値を与えた被験者の数である．

表 3 より，好きな楽曲が 1 位で出力された被験者は，音響的特徴と歌詞情報を用い，さらにテンポの差を考慮した手法を用いた場合に最も多くなっている．また，同手法において，全被験者について 3 位以内に好きな楽曲が出力されていることから，この手法の有効性が示されている．

表 3　出力楽曲に対し“好き”と評価した被験者数（被験者 16 人中）

	1 位以内	2 位以内	3 位以内
Content	11	13	15
Lyrics	9	13	14
Both	11	11	14
Both+Tempo	13	15	16

5　まとめ

本稿では，楽曲推薦システムのための類似楽曲検索手法について述べた．そして，評価実験により音響的特徴と歌詞情報を組み合わせて考慮することの有効性を確認した．

ユーザが楽曲を視聴する際に重視する特徴は人それぞれである．故に，個人の嗜好に適応させた推薦手法が必要となる．今回は，入力として与えた楽曲は 1 曲のみとし，また，音響的類似度と歌詞類似度を共に考慮する方法として，単純に足し合わせただけであったが，複数曲の入力に対してそれからユーザの嗜好を分析し，それぞれの嗜好に応じて 2 つの類似度の統合を行うことで，より良い推薦を実現できると思われる．

参　考　文　献

[1] Tao Li, Mitsunori Ogihara: ”Music Artist Style Identification by Semi-supervised Learning from both Lyrics and Content”, Proceedings of the 12th annual ACM international conference on Multimedia, 2004, pp.364-367

[2] Tao Li , Mitsunori Ogihara: “Toward Intelligent Music Information Retrieval”, IEEE Transactions on Multimedia, Vol.8, No.3, June 2006, pp.564-574

[3] George Tzanetakis, Perry Cook: ”Musical Genre Classification of Audio Signals”, IEEE Transactions on Speech and Audio Processing, Vol.10, No.5, July 2002, pp.293-302

[4] MIRtoolbox: http://users.jyu.fi/~lartillo/mirtoolbox/

[5] Olivier Lartillot: ”MIRtoolbox 1.0 User Guide”

[6] MeCab: http://mecab.sourceforge.net/

[7] Last.fm: http://www.lastfm.jp/