

# 印象に基づく楽曲検索のための個人適応手法の設計と評価

## Design and Evaluation of a User Adaptation Method for Impression-based Music-retrieval

熊本 忠彦  
Tadahiko Kumamoto

### 1. まえがき

大量にあるマルチメディアデータの中からユーザが持つ何らかの判断基準(嗜好や感性, 気分など)に合致するデータを見つけるための手段として, 印象に基づく検索方式が研究されている [1]-[9]. 本方式は, 特定のデータを探し出すという目的には向かないが, ユーザ自身予想しなかったような意外なデータや全く知らない新奇なデータを提示しうる発見的な検索手段と言える.

本方式においては (1) データから受ける印象がユーザによって異なる, (2) データの印象を表現する印象語についての認識がユーザによって異なる, といった2種類の個人差 [4] が考えられ, 何らかの個人適応機能を必要とする. 従来研究では, 印象語とデータの対応関係をユーザモデルと定義した上で (i) 学習用データをユーザに評価してもらい, 個人用のユーザモデルを事前に作成する [1][2] (ii) あらかじめ用意された基準ユーザモデルとユーザの印象語についての認識との差が最小となるよう基準ユーザモデルを修正する [3]-[5], といったことが行われており (2) の個人差に対しては様々な手法が提案されている. しかしながら (1) の個人差を対象に, ユーザの印象の受け方を複数のユーザモデルで記述し, 検索結果に対するユーザの評価によりユーザモデルの取捨選択を行う, というアプローチはこれまでなかった.

本論文では, 被験者 100 名が楽曲 80 曲を聴取し, その印象を数値化するという実験の結果に基づいて, 楽曲から受ける印象の類似性を分析し, 被験者を 20 グループに分類する. そして, 各グループに対し, ユーザモデル (楽曲の印象を数値化するための計算式, ならびに数値化した結果) を構築し, 適合フィードバックによりユーザモデルを取捨選択する手法を設計する. また, 既存の印象に基づく楽曲検索システム [9] に本手法を実装し, 性能評価実験 (30 名) によりその有効性を検証する.

### 2. 楽曲印象の多様性と類似性の分析

楽曲から受ける印象の多様性と類似性を調べるために, 以下のような印象評価実験を行った.

#### 2.1 印象評価実験による評価データの獲得

被験者は男性 39 名, 女性 61 名の計 100 名であり, その年齢構成は 20 歳未満 2 名, 20 代 45 名, 30 代 44 名, 40 代 8 名, 50 歳以上 1 名と幅広かった. 被験者には, 各楽曲 (計 80 曲) を 1 回もしくは 2 回聴取し, 表 1 に示された印象尺度 (楽曲印象を形容する印象語の対からなる評価尺度) のそれぞれに対し, 7 段階評

表 1: 印象尺度

番号	印象尺度を構成する印象語の対	
1	静かな	— 激しい
2	落ち着いた	— 忙しい
3	爽やかな	— 重苦しい
4	明るい	— 暗い
5	荘厳な	— 軽々しい
6	ゆったりとした	— 窮屈な
7	綺麗な	— 綺麗でない
8	楽しい	— 悲しい
9	気持ちが落ち着く	— 気持ちが高揚する
10	心が癒される	— 心が傷つく

価 (印象尺度 1 の場合「とても静かな (7 点), 静かな (6 点), 少し静かな (5 点), どちらとも言えない (4 点), 少し激しい (3 点), 激しい (2 点), とても激しい (1 点)」) もしくは「どちらでもない (*nil*)」の評価を行うことが求められ, その結果, 合計で 80,000 個 (80 曲  $\times$  100 名  $\times$  10 印象尺度) の評価データが得られた.

#### 2.2 楽曲印象の多様性に関する分析

まず, 楽曲から受ける印象の多様性を調べるために, 2.1 で得た評価データの各楽曲・各印象尺度における標準偏差を求めた. その結果を図 1 に示す.

標準偏差の分布は, 平均値 1.06, 中央値 1.04 であり, 半数以上が 1 より大きい値となっている. 被験者が評価できる点数が 1 点から 7 点であったことを考えると, 小さい値ではない. ある楽曲・ある印象尺度における評価データの分布を平均値が 4 点, 標準偏差が 1 の正規分布と仮定すると, 理論的には 100 名中約 68 名の被験者が 3 点から 5 点の評価を行い, 約 95 名が 2 点から 6 点の評価を行った計算になる. すなわち, 同じ楽曲を聴いても受ける印象は人それぞれであり, まったく逆の印象を受ける人も少なくないことを示している.

#### 2.3 楽曲印象の類似性に関する分析

2.2 で示したように, 楽曲から受ける印象は人それぞれであるが, 誰もがまったく異なる受け取り方をするというよりも, 印象の受け取り方が似ている人はいると考える方が自然である. そこで, 楽曲から受ける印象の類似性を調べるために, 代表的な階層的クラスタ分析手法の一つである「ユークリッド平方距離によるワード法」を用いて, 2.1 で得た評価データに対しクラスタ分析を行った. このとき, 被験者  $S_A$  の評価データ  $E_A$  と被験者  $S_B$  の評価データ  $E_B$  のユーク

独立行政法人情報通信研究機構 (NICT), けいはんな情報通信融合研究センター, メディアインタラクシオングループ, 〒619-0289 京都府「けいはんな学研都市」光台 3-5, kuma@nict.go.jp

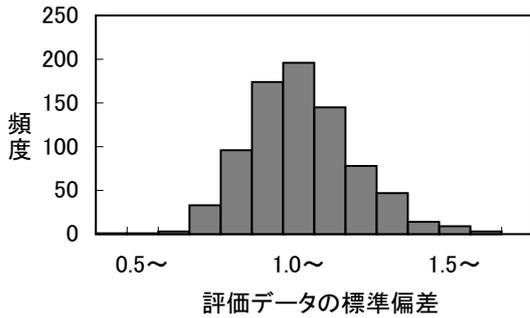


図 1: 各楽曲・各印象尺度における標準偏差の分布

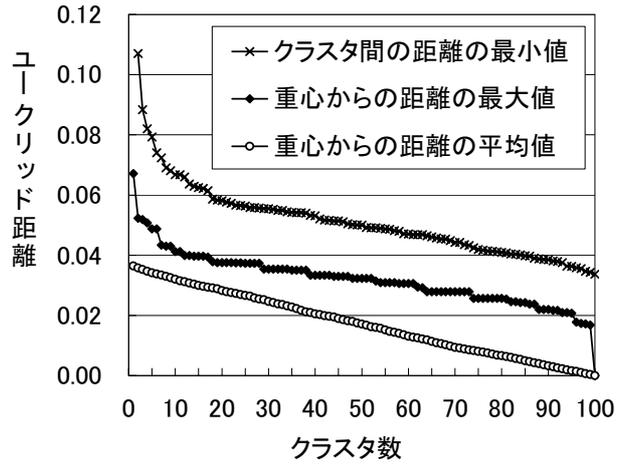


図 2: 各クラスタにおける重心からの距離の平均値と最大値, およびクラスタ間の距離の最小値

リッド平方距離  $D(E_A, E_B)^2$  を次のように定義した .

$$D(E_A, E_B)^2 = \sum_{m=1}^{80} \sum_{i=1}^{10} (x_{m,i}(E_A) - x_{m,i}(E_B))^2$$

但し, 被験者  $S_A$  が楽曲  $m$  に対して行った評価の印象尺度  $i$  における値を  $x_{m,i}(E_A)$  とする . なお,  $x_{m,i}(E_A) = nil$  のときは  $x_{m,i}(E_A) = 4$  として処理した .

ここで, クラスタ分析の過程を示すために, クラスタ分析のそれぞれの時点において, クラスタ  $C_k$  を形成する被験者  $S_j$  の評価データ  $E_j$  とそのクラスタの重心  $G_k$  との距離  $d(E_j, G_k)$ , ならびに任意の 2 つのクラスタの重心間の距離  $d(G_A, G_B)$  を求めた . 図 2 に距離  $d(E_j, G_k)$  の被験者 100 名に対する平均値と最大値, 及び距離  $d(G_A, G_B)$  の最小値を示す . 但し, 距離  $d(E_j, G_k)$  は, クラスタ  $C_k$  の重心  $G_k$  の楽曲  $m$ , 印象尺度  $i$  における値を  $x_{m,i}(G_k)$  と記述することにより,

$$d(E_j, G_k) = \sqrt{\sum_{m=1}^{80} \sum_{i=1}^{10} (x_{m,i}(E_j) - x_{m,i}(G_k))^2} / 800$$

と定義した . 距離  $d(G_A, G_B)$  も同様に

$$d(G_A, G_B) = \sqrt{\sum_{m=1}^{80} \sum_{i=1}^{10} (x_{m,i}(G_A) - x_{m,i}(G_B))^2} / 800$$

と定義した .

図 2 から, 重心からの距離の最大値ならびにクラスタ間の距離の最小値が最初の 10 クラスタ生成時に急激に減少し, 20 クラスタ以降はほぼ一律に減少しているのがわかる . これは, 他の被験者に比べ特異な評価を行った被験者 (群) が早い段階で新たなクラスタを形成したためと考えられる .

### 3. クラスタ分析に基づく複数ユーザモデルの構築

我々は, 文献 [8] において, 標準 MIDI ファイル形式の楽曲から抽出される音の高さ・強さ・長さ・音色に関する特徴量とその楽曲の印象を記述する 10 次元のベクトル (印象ベクトル) との対応関係を定式化するための手法を提案し, 2.1 で述べた評価データを用いて

表 2: 重回帰分析における自由度修正済み決定係数

印象尺度	平均値	最大値	最小値
1	0.785	0.866	0.607
2	0.802	0.869	0.698
3	0.680	0.791	0.592
4	0.679	0.780	0.550
5	0.689	0.772	0.595
6	0.764	0.847	0.649
7	0.665	0.742	0.541
8	0.675	0.794	0.608
9	0.748	0.851	0.569
10	0.699	0.846	0.580
全体	0.719	0.869	0.541

印象ベクトル生成式を設計した . すなわち, 楽曲 (80 曲) から抽出される特徴量を説明変数, 各楽曲に対し被験者 100 名が行った評価の印象尺度  $i$  における平均値を目的変数とする重回帰分析 (変数増加法) を印象尺度ごとに行い, その対応関係を重回帰式という形で定式化した . 今回は, クラスタ分析の結果求められたクラスタのそれぞれに対し, 同様の重回帰分析を行い, 各クラスタごとに印象ベクトル生成式 (=10 個の重回帰式) を設計した . 本論文では, この印象ベクトル生成式, ならびに印象ベクトル生成式によって楽曲から生成される印象ベクトルの集合をユーザモデルと定義し, 図 2 に示された結果と印象ベクトル生成式の設計にかかるコストのトレードオフから, 20 個のユーザモデルを構築することにした (実際には比較実験用として 10 個のユーザモデルも構築した) . なお, 重回帰分析における自由度修正済み決定係数は, 20 個のユーザモデル (=200 個の重回帰式) を構築した場合で, 表 2 のとおりであり, すべての場合において, 0.5 より大きく, 良好な結果が得られているのがわかる .

#### 4. 適合フィードバックによるユーザモデルの取捨選択

ユーザモデルの取捨選択は、検索結果(第1位候補曲)に対するユーザの評価を用いて、以下の手順で行う。

手順1) 入力された検索条件に対し、ユーザモデル  $M_n$  ( $n = 1, 2, \dots, 20$ ) の第1位候補曲  $m_n$  を求める。

手順2) 個人適応値  $p_n$  が  $nil$  (初期値であり、未評価であることを示す) もしくは閾値  $P_{high}$  以上であるユーザモデルの集合  $U_{high}$  が空でなければ、その集合の中で距離が最小となる楽曲  $m_{min}$  を求める。空の場合は、 $p_n$  が閾値  $P_{med}$  以上であるユーザモデルの集合  $U_{med}$  の中から  $m_{min}$  を求める。 $U_{med}$  も空の場合は、すべてのユーザモデルの中から  $m_{min}$  を求める。

手順3) 楽曲  $m_{min}$  を検索結果(第1位候補曲)としてユーザに提示する。

手順4) ユーザがその楽曲と入力印象との適合度(5点: 適合—4点—3点—2点—1点: 不適)を評価した場合のみ、以下の手順を適用し、評価しなかった場合は、手順1の待機状態となる。

手順5) 楽曲  $m_{min}$  を第1位候補曲としたすべてのユーザモデルにおいて、次式を用いて個人適応値  $p_n$  と評価済曲数  $k_n$  を更新する。但し、評価点数が5点から3点のときは、その点数をそのまま  $score$  とするが、2点もしくは1点のときは、ペナルティを加味し、それぞれ1点、-1点を  $score$  とする。

$p_n = nil$  (初期値) のとき

$$p_n = score$$

それ以外のとき

$$p_n = (p_n \times k_n + score) / (k_n + 1)$$

$$k_n = k_n + 1$$

手順6) 評価点数が5点でないとき、ユーザは「再検索」ボタンを押すことができる。「再検索」ボタンが押されたら、楽曲  $m_{min}$  を第1位候補曲としたすべてのユーザモデルを検索の対象外とした上で、手順2に戻る。但し、すべてのユーザモデルが検索対象外となったときは、「検索に失敗しました」と表示した上で、手順1の待機状態となる。

#### 5. 性能評価実験

提案手法を評価するために、既存の印象に基づく楽曲検索システム [9] をベースに3種類のシステムを構築し、以下のような性能評価実験を行った。

被験者は男性15名、女性15名の計30名であり、その年齢構成は20歳未満3名、20代23名、30代2名、50歳以上2名であった。被験者をまず各10名からなる3つのグループに分け、それぞれに異なるシステムを与えた。すなわち、システムA(10個のユーザモデルを用いて個人適応が行われる)、システムB(20個のユーザモデルを用いて個人適応が行われる)、システムC(個人適応は行われないが、「再検索」ボタンが押されたら、第N位候補曲を第N-1位候補曲に繰

#### 楽曲ファイル 印象ベクトル

ave_maria	(1.9 3.6 4.3 5.3 4.3 5.4 4.9 7.1 4.5 5.3)
eine_kleine1_str	(2.4 1.0 6.8 5.4 5.6 5.5 6.0 6.4 0.7 5.6)
gimnopedie1_pi	(3.8 4.3 5.0 2.4 5.7 5.1 5.1 1.6 4.6 2.0)
je_te_veux_pi	(4.8 5.4 5.9 4.9 7.6 5.1 5.3 5.7 5.6 4.6)
la_primavela1_str	(3.1 4.3 6.3 6.4 2.1 5.1 6.0 4.0 3.1 5.1)

(紙面の都合により、小数点第2位を四捨五入した)

図3: 生成された印象ベクトルの例

り上げて提示する)の3種類を用意した。一方、検索対象となる楽曲には、2.1の印象評価実験で用いた80曲に、さらに80曲を追加し(計160曲)、各ユーザモデルごとに対応する印象ベクトル生成式を用いて印象ベクトルを生成した。図3にユーザモデル  $M_1$  において生成された印象ベクトルの例を示す。また、個人適応値に対する閾値として、 $P_{high} = 4.0$ 、 $P_{med} = 3.0$ を設定した。

各被験者は、まず、本システムが受理できる印象語164語と程度語119語のリストを見ながら、システムへの入力文10文を作成し、以下の手順で検索を行った。

- (1) 第N文 ( $N = 1, 2, \dots, 10$ ) を入力し、検索する。
- (2) 検索結果がある場合は、その第1位候補曲を聴取し、入力印象との適合度(「適合している(5点)」、「少し適合している(4点)」、「どちらとも言えない(3点)」、「あまり適合していない(2点)」、「適合していない(1点)」)を評価する。ない場合は、手順(1)に進み、次の文を入力する。
- (3) 5点と評価した場合は、手順(1)に進み、次の文を入力する。5点以外の場合は、「再検索」ボタンを押し、手順(2)に進む。但し、「再検索」ボタンを押せるのは1つの文に対し、5回までとした。

以上の実験を終えた後、約30分の休憩後、各被験者は、同じ10文を用いて再度検索(上記手順(1)及び(2)のみ)を行い、それぞれの第1位候補曲に対し適合度評価を行った。この実験の結果を表3、表4に示す。

表3は各文に対する最初の適合度評価の結果\*を示し、表4は同じ10文を再度入力した際の適合度評価の結果を示している。但し、入力ミスや意味解析失敗による誤った検索結果に対する評価は除外した。また、同じ検索結果に対して異なる評価を与えた場合は、最初の評価を優先し、2回目の評価を最初の評価で置き換えた。

表3と表4を比べてみると、システムBの初回検索時の検索精度と2回目の検索精度には統計的に有意な差があるが(有意水準1%)、他の2つのシステムにはない(有意水準5%)ことがわかる。よって、提案手法(システムB)は有効であると言える。

\*システムA、BとシステムCでは平均値に開きがある。そこで、平均値の差の検定(有意水準5%)を行ってみたが、有意ではなく、被験者グループ間の揺れと考えられる。

表 3: 初回検索時の検索精度 (平均適合度)

システム	平均値	標準偏差	評価回数
A (10 モデル)	3.38	1.28	106
B (20 モデル)	3.39	1.36	99
C (適応なし)	3.69	1.09	97

表 4: 文を再度入力した際の検索精度 (平均適合度)

システム	平均値	標準偏差	評価回数
A (10 モデル)	3.73	1.24	92
B (20 モデル)	4.02	1.16	96
C (適応なし)	3.70	1.10	97

表 5: 同じ検索結果に対する 2 回目の評価と最初の評価の差

評価の差	-3	-2	-1	0	1	2	3
頻度	1	2	19	107	30	15	2
割合 (%)	0.6	1.1	10.8	60.8	17.0	8.5	1.1

ここで、同じ検索結果 (第 1 位候補曲) に対する評価の変化を表 5 にまとめる。表 5 は、楽曲から受ける印象の状況依存性、すなわち検索文脈や心的状態の変化により、同じ楽曲に対しても異なる印象を受ける場合があることを示唆しており、より高精度な個人適応を実現するためには、解決すべき課題と考える。

参考のために、被験者が作成した文の例を表 6 に示し、用いられた印象語、程度語の異なり数を表 7 に示す。

## 6. まとめ

同じ楽曲を聴取しても受ける印象は人によって異なる。そこで、本論文では、被験者 100 名に楽曲 80 曲の印象を評価してもらい、評価の仕方 (= 印象の受け方) が似た人どうしをクラスタ分析手法を用いてグルーピングするとともに、それぞれのグループに適したユーザモデル (楽曲の印象を数値化するための式、ならびに数値化した結果) を構築し、未知のユーザがどのグループに属するか、すなわちどのユーザモデルが適しているかを決定するための個人適応手法を提案した。そして、別の被験者 30 名による性能評価実験を通して、その有効性を検証した。

今後の課題として、従来研究と同様、印象語についての認識に対する個人差に対処する必要があると考えている。また、検索文脈や心的状態の変化に伴い印象の受け方も変化することから、実用システムを構築するためには、より柔軟な個人適応手法を考えていく必要がある。

## 参考文献

- [1] 栗田多喜夫, 加藤俊一, 福田郁美, 坂倉あゆみ: 印象語による絵画データベースの検索, 情報処理学

表 6: 被験者が作成した文 (一部)

めちゃめっちゃ迫力のある曲 / 清々しくて爽やかな曲 / 軽快でコミカルな曲 / スーパー感動する曲 / メリハリがあり情熱的な曲 / 重厚で荘厳で厳粛な曲 / しっかりとした叙情的な曲 / 結構落ち着いた曲 / 力強くダイナミックな曲 / 百パーセント心が癒される曲 / 美しく悲しい、透明な曲
--

表 7: 被験者が作成した文に用いられた印象語・程度語

システム	異なり数		印象語, 程度語の例 (頻度 5 以上を抜粋)
	印象語	程度語	
A	74 語	30 語	コミカルな / 心が癒される / 透明な / 明るい / とても
B	66 語	20 語	元気の出る / 迫力のある / 美しい / 古典的な / 力強い / 綺麗な / とても
C	63 語	47 語	明るい / 軽快な / 幻想的な / 情熱的な / 落ち着いた / 力強い / だいが / とても / 適度に

会論文誌, vol.33, no.11, pp.1373-1383, 1992.

- [2] 辻康博, 星守, 大森匡: 曲の局所パターン特徴量を用いた類似曲検索・感性語による検索, 信学技報, vol.SP96-124, pp.17-24, 1997.
- [3] 河辺和宏, 江澤義典, 平嶋宗, 豊田順一: 官能の差異を考慮したユーザモデル・チューニング法, 情処研報, vol. ヒューマンインタフェース 45-1, pp. 1-8, 1992.
- [4] 清木康, 金子昌史, 北川高嗣: 意味の数学モデルによる画像データベース探索方式とその学習機構, 信学論 (D-II), vol.J79-D-II, no.4, pp.509-519, 1996.
- [5] 木本晴夫, 感性語による画像検索とその精度評価, 情報処理学会論文誌, vol.40, no.3, pp.886-898, 1999.
- [6] 佐藤聡, 小川潤, 堀野義博, 北上始: 感情に基づく音楽作品検索システムの実現に向けての検討, 信学技報, vol.SP2000-137, pp.51-56, 2001.
- [7] 池添 剛, 梶川嘉延, 野村康雄, “音楽感性空間を用いた感性語による音楽データベース検索システム,” 情報処理学会論文誌, vol.42, no.12, pp.3201-3212, 2001.
- [8] 熊本忠彦, 太田公子: 印象に基づく楽曲検索システム: N グラム統計量の利用, 情報技術レターズ, vol.1, no.LD-6, pp.63-64, 2002.
- [9] 熊本忠彦, 太田公子: 印象に基づく楽曲検索: 自然言語インタフェースの実現, 情処学自然言語処理研報, vol.2003-NL-153, no.13, pp.97-104, 2003.