

印象に基づく楽曲検索のためのユーザモデリング手法

熊本 忠彦[†]

「楽しい」や「静かな」といった印象語を検索キーとする楽曲検索方式では(1)楽曲から受ける印象がユーザによって異なる(2)楽曲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、という2種類の個人差が問題となる。本論文では、前者に焦点を当て、ユーザの楽曲印象の受け方を記述するためのユーザモデリング手法、ならびに検索結果に対するユーザの評価に基づいて、より適合するユーザモデルを選択するための適合フィードバック手法を提案する。具体的には、印象評価実験(被験者100名、楽曲80曲)の結果に基づいて、楽曲から受ける印象の類似性をクラスタ分析し、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式、および数値化した結果)を構築し、適合フィードバックにより適当なユーザモデルを取捨選択する手法を設計する。また、以上の手法を既存の印象に基づく楽曲検索システムに実装し、性能評価実験(30名)を行うことにより、その有効性を検証する。

A User Modeling Method for Impression-based Music Retrieval

TADAHIKO KUMAMOTO[†]

We have developed a scheme for music retrieval that adapts to the user's impressions of the musical pieces. First, we conducted impression-estimation experiments in which 100 subjects gave their impression of 80 musical pieces, and then, using a clustering method, we classified the 100 subjects into 20 groups based on the results. Next, we created a user model for each group consisting of formulas for numerically expressing the impressions and a set of vectors calculated using the formulas. We then developed a procedure for identifying the most suitable model for an unidentified user. Testing of the models and procedure in an existing impression-based music-retrieval system demonstrated the effectiveness of the proposed scheme.

1. ま え が き

大量にあるマルチメディアコンテンツの中から特定のコンテンツを探し出すための手段としては、書誌情報(タイトルや製作者名など)による検索が一般的であり、一部では内容情報(歌声やハミング、スケッチなど)に基づく検索^{1)~6)}も可能となっている。これらの検索手段は、探し出したいコンテンツを特定するだけの具体的な情報を入力できる場合にはよいが、そうでない場合、たとえば、漠然とした要求しかない場合や検索対象に関する知識が乏しい場合など欲しいコンテンツを特定できない場合には不向きといえる。一方、ユーザが持つ何らかの判断基準(嗜好や感情、気分など)に合致するコンテンツ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^{7)~15)}。印象という主観的な情報を検索キーとするこれらの方

式は、特定のコンテンツを探し出すという目的には向かないが、ユーザの知らないコンテンツを提示しうる発見的な検索手段といえる。

印象に基づく検索方式においては(1)コンテンツから受ける印象がユーザによって異なる(2)コンテンツ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、という2種類の個人差⁹⁾が考えられ、何らかの個人適応機能が必要とされる。先行研究では、印象語とコンテンツの対応関係をユーザモデルと定義したうえで(i)学習用データを各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^{8),10)}(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^{7),9),11)}、といったことが行われており、タイプ(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、タイプ(1)の個人差を対象とする研究はこれまで見当たらなかった。

そこで本論文では、タイプ(1)の個人差に焦点を当て、ユーザの印象の受け方を複数のユーザモデルで記

[†] 独立行政法人情報通信研究機構・知識創成コミュニケーション研究センター・自然言語グループ
Computational Linguistics Group, National Institute of Information and Communications Technology

述し、その中から適切なユーザモデルを選択するという新たな個人適応方式を提案する．具体的には、ユーザの印象の受け方を記述するためのユーザモデリング手法を提案するとともに、検索結果に対するユーザの評価に基づいて、より適合するユーザモデルを選択するための適合フィードバック手法を示す．なお、検索対象としては、標準 MIDI ファイル形式の楽曲を採用する．

本論文では、まず、被験者 100 人が楽曲 80 曲を聴取し、各楽曲の印象を数値化するという印象評価実験を行う．この実験の結果に基づいて、楽曲から受ける印象の類似性をクラスタ分析し、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）を構築するとともに、適合フィードバックに基づいてユーザモデルを取捨選択するための手法を設計する．また、既存の印象に基づく楽曲検索システム¹⁵⁾に本手法を実装し、別の被験者 30 名を対象とする性能評価実験を行うことにより、その有効性を検証する．

ここで、本論文の以下の構成を示す．まず、2 章でユーザの楽曲印象の受け方を調べるための評価データを収集する．次に 3 章で評価データを用いてクラスタ分析を行い、印象の受け方が類似している被験者同士をグルーピングするとともに、各グループに対し、ユーザモデルを構築する．4 章で検索結果に対するユーザの評価結果に基づいて、ユーザモデルの取捨選択を行う適合フィードバック手法を設計する．そして、5 章で被験者（30 名）による性能評価実験を行い、提案手法の有効性を示す．最後に 6 章で本論文のまとめと今後の課題について述べる．

2. 楽曲印象の受け方を表すデータの収集

ユーザの楽曲印象の受け方を調べるために、以下のような印象評価実験を行い、評価データを収集した．

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 名であった．被験者は、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 段階評価（たとえば、印象尺度 1 の場合は「とても静かな（7 点）、静かな（6 点）、少し静かな（5 点）、どちらともいえない（4 点）、少し激しい（3 点）、激しい（2 点）、とても激しい（1 点）」の 7 段階）もしくは「どちらで

表 1 印象尺度

Table 1 Ten impression scales designed for impression-based music retrieval.

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

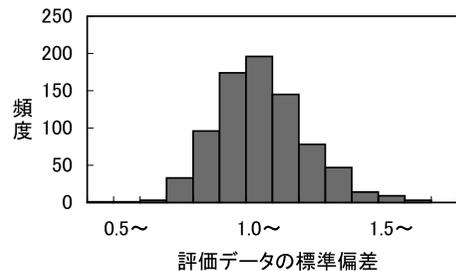


図 1 各楽曲・各印象尺度における標準偏差の分布

Fig. 1 Standard deviation of data in each impression scale of each musical piece.

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 × 100 名 × 10 印象尺度）の評価データが得られた．

なお、楽曲印象を評価する印象尺度には、すでにタイプ (2) の個人差が含まれており、その解釈には少なからず多義性があると考えられるが、楽曲印象を評価する語を限定し、評価尺度を統一することにより、タイプ (2) の個人差は格段に減少しているものと考えられる．また、印象尺度の解釈に多義性があっても、多くの楽曲に対して同じような評価を行った人同士をグルーピングしているので、その多義性は平均化され、タイプ (2) の個人差の影響は小さくなっているものと考えられる．

2.2 楽曲印象の多様性に関する分析

楽曲から受ける印象の多様性を調べるために、2.1 節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．被験者の評価できる点数が 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。なお、標準偏差の分布において、最大値は 1.65、最小値は 0.44 であった。

3. クラスタ分析に基づく複数ユーザモデルの構築

3.1 楽曲印象の受け方に関するクラスタ分析

2.2 節で示したように、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がいると考える方が自然である。そこで、代表的な階層的クラスタ分析手法の 1 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 節で得た評価データに対しクラスタ分析を行い、印象の受け方が類似している被験者同士をグルーピングした。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。

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

ただし、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、および距離 $d(G_A, G_B)$ の最小値を示す。ただし、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、

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

と定義された。距離 $d(G_A, G_B)$ も同様に、

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

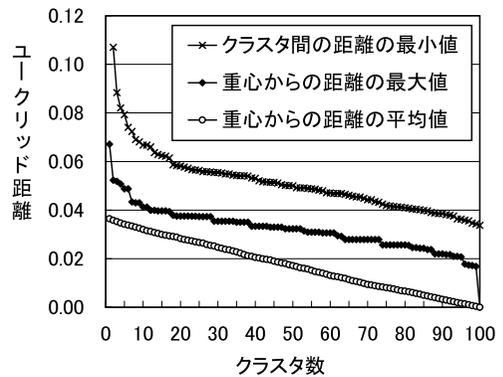


図 2 クラスタ分析の過程

Fig. 2 Process of clustering analysis.

と定義された。

図 2 から、生成されたクラスタ数が 1~10 のとき、重心からの距離の最大値もクラスタ間の距離の最小値も急激に変化し、クラスタ数が 20~100 のときは、ほぼ一律に緩やかに変化しているのが分かる。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。実際、各クラスタの個体数は、10 クラスタ生成時のときで、1, 1, 7, 26, 4, 1, 27, 27, 5, 1 (生成順)、20 クラスタ生成時のときで、1, 1, 1, 27, 1, 1, 12, 4, 1, 5, 2, 1, 9, 2, 4, 8, 16, 2, 1, 1 (生成順)であり、個体数の少ないものが多かった。

3.2 複数ユーザモデルの構築

筆者らは、文献 14) 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 N グラム特徴量 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ている。すなわち、楽曲(計 80 曲)から抽出される N グラム特徴量 ($N = 1, 2, \dots, 5$) を説明変数、各楽曲に対し被験者(計 100 名)が行った評価の印象尺度 i ($i = 1, 2, \dots, 10$) 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)⁶⁾ 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ている。このとき、 N グラム特徴量の N の組み合わせ方(たとえば bigram と trigram のみを利用)を 5 通り、 N グラム特徴量に対する重みの与え方を 3 通り用意したので、結局、印象尺度ごとに 15 回の重回帰分析を行う必要があった。以上の結果、印象尺度ごとに設計された計 10 個の重回帰式が印象ベ

音(音符)の時間的推移を記述する N グラム形式の特徴量であり、音の高さ・強さ・長さに関しては $N = 1, 2, 3, 4, 5$ の N グラム特徴量が生成され、音色に関しては $N = 1$ の unigram 特徴量が生成される。

クトル生成式であり、各重回帰式によって求められる値が印象ベクトルの要素となる。

本論文では、本手法¹⁴⁾をクラスタ分析の結果求められたクラスタのそれぞれに適用し、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。このとき、図2に示されたクラスタ分析の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、クラスタ数として20クラスタを採用した。すなわち、被験者100名が20クラスタに分類された時点の各クラスタにおいて、そのクラスタに属する被験者が行った評価の平均値を目的変数、楽曲から抽出される特徴量を説明変数とする重回帰分析(変数増加法)を行い、印象ベクトル生成式(10個の重回帰式)を設計した。また、性能評価実験時の比較対象として、被験者100名が10クラスタに分類された時点の各クラスタに対しても、印象ベクトル生成式を設計した。本論文では、このような印象ベクトル生成式とその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集合をユーザモデルと呼んでいる。

ここで、参考のために、ユーザモデル M_1 (20クラスタ生成時)の印象尺度1の場合の重回帰式(説明変数、偏回帰係数、定数項)を表2に示す。ユーザモデル M_1 の印象尺度1の場合、説明変数は連続する2音からなる bigram 特徴量であり、各 bigram 特徴量の相対出現頻度に重み w を掛け加えたものがそれぞれの説明変数に値として代入される。表2において、第1要素の記号 h, v, d は特徴量の種類(音の高さ、強さ、長さ)を表し、数値は第1音から得られる特徴量の値を示している。ハイフンは同一チャンネルにおいて複数の音が同時に発音されたことを示している。第2要素の記号 u, e, d は、第2音から得られる特徴量の値が、第1音のそれに比べ、大きい(u)、同じ(e)

あるクラスタ・ある印象尺度に対応する重回帰式を設計するためには、前述したように、15回の重回帰分析を行う必要がある。したがって、全体では(15回×10印象尺度×クラスタ数)回の重回帰分析を行うことになる。加えて、多重共線性の問題を回避するために変数増加法を採用したため、1回の重回帰分析の過程において、適切な説明変数の取捨選択を行う必要がある。この取捨選択は、通常、3~10回行われたので、全体としては、かなりの時間と労力を費やすこととなる。したがって、クラスタ数をなるべく少なくしたいという欲求が生じる。その一方で、クラスタ数が多ければ多いほど、多様なユーザモデルを構築できるので、より高い精度を得られる可能性がある。音の高さ、強さ、長さは、それぞれノートナンバ値、オンベロシティ値、ノートオンメッセージからノートオフメッセージが到着するまでの時間(ミリ秒換算)に対応しているが、各特徴量がとりうる値の大きさの違いを考慮して、音の長さや強さに関する特徴量に対しては、抽象化を行い、値の変化に対する感度が鈍くなるよう設計している。

表2 ユーザモデル M_1 の印象尺度1の場合の重回帰式
Table 2 Multiple regression equation for impression scale No.1 in user model M_1 .

説明変数	偏回帰係数	説明変数	偏回帰係数
v5 d	28.0	h44 u	22.2
v6-4-2 e	-466.7	h58-55 e	-102.4
v9-8 d	-132.2	h60 d	-55.8
v10 e	2.7	h61-56 d	-54.6
v12 e	-1.0	h63-55 xs	-339.3
v12 u	-11.4	h69 u	-44.2
d1 u	4.5	h74-57 d	66.9
d8 e	5.6	h89 u	-72.7
d62 d	86.9	定数項	3.8

あるいは小さい(d)ことを表しており、記号 xs は第2音が無音であることを表している。

なお、重回帰分析の精度を表す自由度修正済み決定係数¹⁶⁾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3のとおり、10個のユーザモデル(100個の重回帰式)を構築した場合で表4のとおりであり、すべての分析結果において、0.5より大きく、良好な結果が得られているのが分かる。

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、検索結果中の第1位候補曲に対するユーザの評価(5点:適合~1点:不適)を用いて、以下の手順で行われる。

手順1) 楽曲の印象(検索条件)が入力されたら、それぞれのユーザモデル M_n ($n = 1, 2, \dots, 20$)において、距離が最小となる第1位候補曲 m_n を求める。

手順2) ユーザモデル M_n に対するユーザの個人適応値を p_n とするとき、 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が1つ以上あれば、その中から距離が最小となる楽曲 m_{min} を求める。なければ、個人適応値 p_n が閾値 P_{med} 以上であるユーザモデルを求め、その中から距離が最小となる楽曲 m_{min} を求める。そのようなユーザモデルもない場合は、すべてのユーザモデルの中から

サンプル数と説明変数の数との差が小さい場合、決定係数は大きくなる傾向がある。この不具合を修正したのが自由度修正済み決定係数であり、残差平方和を S_e 、偏差平方和を S_{yy} 、サンプル数を n 、説明変数の数を q とするとき、次の式で計算される。

$$R^{2'} = 1 - \frac{S_e / (n - q - 1)}{S_{yy} / (n - 1)}$$

表 3 重回帰分析における自由度修正済み決定係数 (20 ユーザモデル構築時)

Table 3 Coefficients of determination adjusted for the degrees of freedom in multiple regression analysis for creating 20 user models.

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

表 4 重回帰分析における自由度修正済み決定係数 (10 ユーザモデル構築時)

Table 4 Coefficients of determination adjusted for the degrees of freedom in multiple regression analysis for creating 10 user models.

印象尺度	平均値	最大値	最小値
1	0.793	0.867	0.596
2	0.797	0.920	0.681
3	0.672	0.743	0.609
4	0.676	0.729	0.609
5	0.709	0.789	0.613
6	0.779	0.856	0.676
7	0.675	0.735	0.603
8	0.666	0.733	0.619
9	0.773	0.850	0.640
10	0.698	0.791	0.606
全体	0.724	0.920	0.596

距離が最小となる楽曲 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 (検索条件) との適合度 (5 点: 適合—4 点—3 点: 中間—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。ただし, 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 が未評価のとき,

$$p_n = score$$

$$k_n = 1$$

それ以外の場合,

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

$$k_n = k_n + 1$$

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うえで, 手順 2 に戻る。ただし, 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」と表示したうえで, 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¹⁵⁾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, 印象評価実験の被験者 100 名とは異なる 30 名 (20 歳未満 3 名, 20 代 23 名, 30 代 2 名, 50 歳以上 2 名) が採用された。被験者をまず各 10 名からなる 3 つのグループ (男女 5 名ずつ) 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 (提案システムであり, 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 (個人適応は行われませんが, 「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 節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), 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。表 5 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。なお, ベースとなった楽曲検索システムは自然言語インタフェースを有していたので, システム A, システム B, システム C への印象入力とは文形式で行

各システムを利用した被験者の年齢構成は, システム A: 20 代 9 名, 30 代 1 名, システム B: 10 代 2 名, 20 代 5 名, 30 代 1 名, 50 代 2 名, システム C: 10 代 1 名, 20 代 9 名であった。

個人適応値が $P_{high} = 4.0$ 以上のユーザモデルは, 平均して「良い」評価の検索結果を提示するユーザモデルであると考えられるので, このようなユーザモデルを本論文では「高適応」と位置づける。同様に, 平均して「悪くない」評価の検索結果を提示する $P_{med} = 3.0$ 以上のユーザモデルを「中適応」, 平均して「悪い」評価の検索結果を提示する $P_{med} = 3.0$ 未満のユーザモデルを「低適応」と位置づける。

表 5 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Table 5 Impression vectors generated from several musical pieces using user model M_1 .

アヴェ・マリア	(1.9 3.6 4.3 5.3 4.3 5.4 4.9 7.1 4.5 5.3)
アイネ・クライネ・ナハトムジーク	(2.4 1.0 6.8 5.4 5.6 5.5 6.0 6.4 0.7 5.6)
ジムノペディ	(3.8 4.3 5.0 2.4 5.7 5.1 5.1 1.6 4.6 2.0)
ジュ・トゥ・ヴ	(4.8 5.4 5.9 4.9 7.6 5.1 5.3 5.7 5.6 4.6)
ラ・ブリマヴェーラ	(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 位で四捨五入されている)

表 6 システムが受理可能な印象語の例

Table 6 Examples of impression words that can be accepted by our system.

静かな、激しい、落ち着いた、忙しい、爽やかな、重苦しい、明るい、暗い、荘厳な、軽々しい、ゆったりとした、窮屈な、綺麗な、楽しい、悲しい、気持ちが落ち着く、気持ちが高揚する、心が癒される、心が傷つく、うるさい、かっこいい、コミカルな、しっとりとした、ダイナミックな、ほっとする、ほのぼのとしている、メリハリのある、リラックスできる、哀れな、愛しい、穏やかな、懐かしい、感動する、緩やかな、嬉しい、気まぐれな、気高い、気持ちが安らぐ、気持ちが暗くなる、気持ちが和む、強い、軽い、軽やかな、軽快な、元気が出る、厳かな、厳肅な、幻想的な、古典的な、刺激的な、耳障りな、寂しい、主張のある、柔らかな、重厚な、叙情的な、情熱的な、心が休まる、心が豊かになる、心地よい、崇高な、清々しい、静かに流れるような、壮大な、爽快な、単純な、弾んだ、断固とした、沈んだ、透き通った、透明な、迫力のある、美しい、浮かれた、猛烈な、優しい、優雅な、雄大な、陽気な、流れるような、力強い、恋しい
--

われることになる。

各被験者は、まず、本システムが受理できる印象語 164 語 (表 6 参照) 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、1 文ずつ順番に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 候補曲を聴取し、入力した印象との適合度 (「適合している (5 点)」、「少し適合している (4 点)」、「どちらともいえない (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。検索結果がない場合は、手順 (1) に戻り、次の文を入力する。
- (3) 5 点と評価した場合は、手順 (1) に戻り、次の文を入力する。5 点以外の場合は、「再検索」ボタン

印象語 164 語は肯定形 82 語とその否定形 82 語からなるが、紙面の都合により、表 6 には肯定形のみを示す。詳しくは文献 17) の程度語一覧表を参照されたい。

表 7 初回検索時と 2 回目検索時の検索精度

Table 7 Retrieval accuracies in the first and second retrieval using ten input sentences.

		初回検索時	2 回目検索時
システム A (10 ユーザモデル)	μ	3.38	3.73
	σ	1.28	1.24
	N	106	92
システム B (20 ユーザモデル)	μ	3.39	4.02
	σ	1.36	1.16
	N	99	96
システム C (個人適応機能なし)	μ	3.69	3.70
	σ	1.09	1.10
	N	97	97

を押し、手順 (2) に戻る。ただし、「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。

以上の実験を終了後、各被験者は、約 30 分の休憩を取り、その後、再び同じ 10 文を用いて検索 (上記手順 (1) および (2) のみ) を行った。初回検索時および 2 回目検索時の検索結果 (第 1 候補曲) に対する適合度評価の結果 を表 7 にまとめる。表 7 において、 μ は適合度の平均値、 σ は標準偏差、 N は適合度評価の回数を示している。また、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するとともに、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。

自由度修正済み決定係数は 10 ユーザモデル構築時のほうが良い結果 (表 3, 表 4 参照) を示していたが、検索結果に対する適合度評価の改善という点では提案手法 (システム B) の方が有効であった。すなわち、初回検索時と 2 回目検索時の適合度の平均値を比べてみると、システム B (提案システム) には有意水準 1% で統計的に有意な差があるが (Z 検定¹⁸⁾: $Z = 3.46$)、他の 2 つのシステムには有意水準 5% でも有意な差がなかった (システム A: $Z = 1.958$, システム C: $Z = 0.07$)。

さて、システム B (提案システム) を用いた被験者 10 名を対象に、各ユーザモデルにおける個人適応値の推移を調べてみたところ、推移パターンとして、ユーザモデル (全 20 個) の半分以上が高適応 (4.0 以上) となったパターン A (2 名)、半分以上が中適応 (3.0 以上 4.0 未満) となったパターン B (2 名)、半分以上が低適応 (3.0 未満) となったパターン C (2 名)、高適応・中適応・低適応のユーザモデル数がほぼ等しくなったパターン D (4 名) の 4 パターンが観測された。それぞれの推移パターンの例を図 3, 図 4, 図 5,

システム A, システム 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。

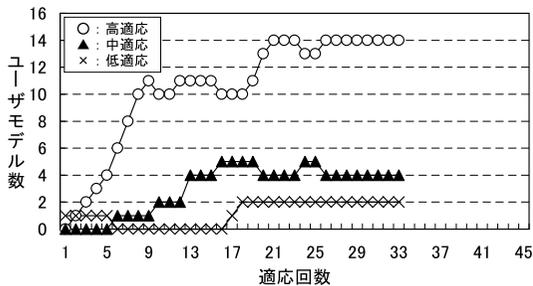


図 3 ユーザモデル個人適応値の推移パターン A

Fig. 3 Change pattern "A" of fitness values in user models.

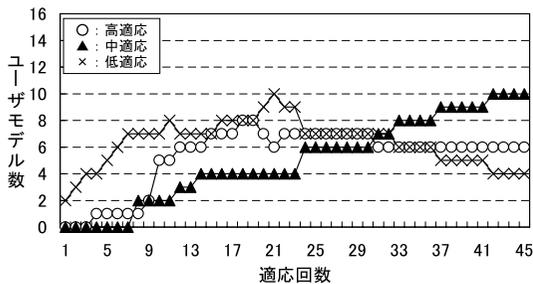


図 4 ユーザモデル個人適応値の推移パターン B

Fig. 4 Change pattern "B" of fitness values in user models.

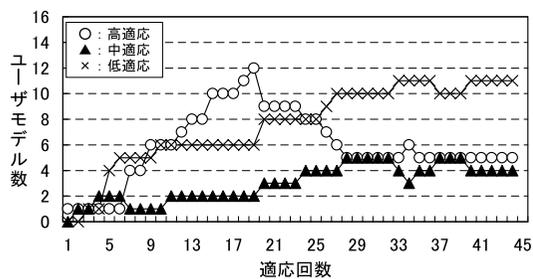


図 5 ユーザモデル個人適応値の推移パターン C

Fig. 5 Change pattern "C" of fitness values in user models.

図 6 に示す。なお、各図は、適合度評価の回数（適応回数）が増えるにつれて、高適応・中適応・低適応のユーザモデルの数がどう変化するかを示している。

クラスタリングの考え方、すなわち距離が大きいものを異なるクラスタに、距離が小さいものを同じクラスタに分類するという考え方からすると、あるユーザに適合するユーザモデルは、個人適応が進むにつれて、1 つもしくはごく少数のユーザモデルに絞られていくものと予想されたが、実際にはそうっていない。推移パターン A では高適応なユーザモデルが多数を占めているし、他の推移パターンでもまだ 5~7 このユーザモデルが高適応と位置づけられている。その原因が

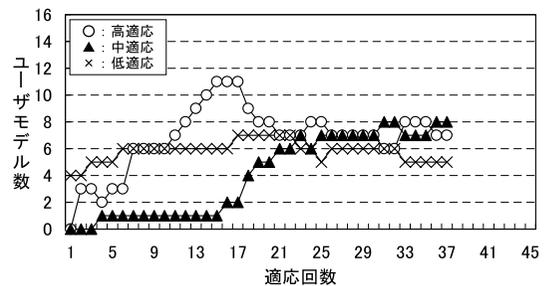


図 6 ユーザモデル個人適応値の推移パターン D

Fig. 6 Change pattern "D" of fitness values in user models.

どこにあるのか、単に適応回数が少ないからなのか、検索対象となる楽曲数が少なかったためなのか、それとも 20 このユーザモデルでは新規ユーザに対し適したユーザモデルを見つけられないのか、といったことを明らかにしていく必要がある。今後の課題とする。

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) が類似した人同士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方式を提案した。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、先行研究と同様、タイプ (2) の個人差である印象語についての認識に対する個人差に対しても、何らかの個人適応方式を考えていくことがあげられる。今回の印象評価実験において、楽曲の印象を評価する際に印象語の対を用いていることから、収集された評価データにはタイプ (2) の個人差による誤差がすでに含まれていると考えられる。したがって、先行研究で提案されているようなタイプ (2) に対する個人適応方式と今回の提案方式をうまく統合できれば、お互いの相乗効果で、より高精度な印象に基づく楽曲検索を実現できるものと考えられる。また、検索文脈 (聴取する楽曲の連続の仕方) や心的状態の変化にともない印象の受け方も変化すると考えられることから、検索文脈あるいは心的状態に依存して、受ける印象がどう変わるのかといったことを定性的かつ定量的に調査し、その対処法を探っていく必要がある。

参 考 文 献

- 1) 椋木雅之, 美濃導彦, 池田克夫: 対象物スケッチによる風景画像検索とインデックスの自動生成, 信学論 (D-II), Vol.J79-D-II, No.6, pp.1025-1033 (1996).
- 2) Blackburn, S.G. and DeRoure, D.C.: A Tool for Content Based Navigation of Music, *Proc. 6th ACM International Multimedia Conference*, Bristol, UK, pp.361-368 (1998).
- 3) Cascia, M.L., Sethi, S. and Sclaroff, S.: Combining Textual and Visual Cues for Content-based Image Retrieval on the World Wide Web, *Proc. IEEE Workshop on Content-based Access of Image and Video Libraries*, Santa Barbara, USA, pp.24-28 (1998).
- 4) 園田智也, 後藤真孝, 村岡洋一: WWW 上での歌声による曲検索システム, 信学論 (D-II), Vol.J82-D-II, No.4, pp.721-731 (1999).
- 5) 橋口博樹, 西村拓一, 張 建新, 滝田順子, 岡 隆一: モデル依存傾斜制限型の連続 DP を用いた鼻歌入力による楽曲信号のスポッティング検索, 信学論 (D-II), Vol.J84-D-II, No.12, pp.2479-2488 (2001).
- 6) 小杉尚子, 小島 明, 片岡良治, 串間和彦: 大規模音楽データベースのハミング検索システム, 情報処理学会論文誌, Vol.43, No.2, pp.287-298 (1999).
- 7) 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, 情報処理学会研究報告, Vol. ヒューマンインタフェース 45-1, pp.1-8 (1992).
- 8) 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, 情報処理学会論文誌, Vol.33, No.11, pp.1373-1383 (1992).
- 9) 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).
- 10) 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).
- 11) 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).
- 12) Sato, A., Ogawa, J. and Kitakami, H.: An Impression-based Retrieval System of Music Collection, *Proc. 4th International Conference on Knowledge-Based Intelligent Engineering Systems and Allied Technologies*, Brighton, UK, pp.856-859 (2000).
- 13) 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).
- 14) 熊本忠彦, 太田公子: 印象に基づく楽曲検索システムの設計・構築・公開, 人工知能学会論文誌, Vol.21, No.3, pp.310-318 (2006).
- 15) Kumamoto, T.: Design and Implementation of Natural Language Interface for Impression-based Music-retrieval Systems, *Proc. International Conference on Knowledge-Based Intelligent Information and Engineering Systems*, Wellington, New Zealand, pp.139-147, LNAI3214, Springer (2004).
- 16) 菅 民郎: 多変量統計分析, 現代数学社, 京都 (2000).
- 17) 熊本忠彦: 程度語の序列化と自然言語感性検索への応用, 情報処理学会自然言語処理研究会研究報告, Vol.2004, No.108, pp.77-82 (2004).
- 18) ホーエル, P.G. (著), 浅井 晃, 村上正康 (訳): 初等統計学, pp.172-176, 培風館, 東京 (1989).
(平成 17 年 12 月 20 日受付)
(平成 18 年 4 月 10 日採録)

(担当編集委員 森本 康彦)



熊本 忠彦 (正会員)

昭和 63 年筑波大学第三学群情報学類卒業。平成 2 年筑波大学大学院理工学研究科修士課程修了。同年郵政省通信総合研究所(現, 独立行政法人情報通信研究機構)入所。現在に至る。近年は, 印象マイニング, 印象検索, 印象表出の研究に従事。平成 8 年博士(工学)(筑波大学)。FIT2004 論文賞受賞。電子情報通信学会, 人工知能学会, 言語処理学会, 日本データベース学会各会員。