

印象に基づく楽曲検索システムにおける個人適応について

熊本 忠彦 太田 公子

独立行政法人 通信総合研究所 けいはんな情報通信融合研究センター

表 1: 印象尺度

印象尺度	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

1. まえがき

大量にある楽曲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、印象に基づく楽曲検索方式が研究されている[1]-[3]。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ような新奇なデータを提示しうる発見的な検索手段と言える。

本方式は、印象という主観的な情報を検索条件として用いるため、何らかの個人適応機能を必要とする。すなわち、楽曲から受ける印象はユーザによって異なり、一つではないので、入力された検索条件が同じでも、ユーザの印象の受け方に応じて検索結果を適宜変える必要がある。そこで、本稿では、被験者100人に楽曲80曲の印象を評価してもらい、楽曲から受ける印象の多様性と類似性を分析するとともに、その結果に基づいて印象に基づく楽曲検索のための個人適応手法を提案する。

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人、女性61人の計100人であり、その年齢構成は20歳未満2人、20代45人、30代44人、40代8人、50歳以上1人であった。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（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た。

以上の実験の結果、80,000個（80曲×100人×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、標準偏差が1の正規分布と仮定すると、理論的には約68人の被験者が3点から5点の評価を行い、約95人が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

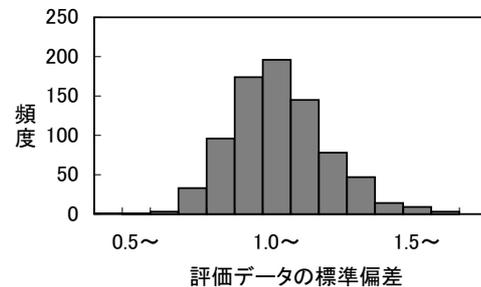


図 1: 各楽曲・各印象尺度における評価データの標準偏差の分布

タに対しクラスタ分析を行った。このとき、被験者 S_A の評価データと被験者 S_B の評価データのユークリッド平方距離 $D(S_A, S_B)^2$ は次のように定義された。

$$D(S_A, S_B)^2 = \sum_{m=1}^{80} \sum_{i=1}^{10} (x_{m,i}(A) - x_{m,i}(B))^2$$

但し、被験者 S 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 $x_{m,i}(S)$ とする。なお、 $x_{m,i}(S) = nil$ のときは $x_{m,i}(S) = 4$ として処理した。

結果を図2と図3に示す。図2は、クラスタ分析のそれぞれの時点において、クラスタ C を形成する被験者 S の評価データとそのクラスタの重心 G_C との距離 $d(S, G_C)$ を求め、被験者100人に対する平均値と最大値を算出した結果である。但し、見易さの観点から、距離 $d(S, G_C)$ の値が各楽曲・各印象尺度における誤差（重心からのずれ）の平均値を表すように、

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

と定義した。 $x_{m,i}(G_C)$ は、クラスタ C の重心 G_C の楽曲 m 、印象尺度 i における値である。 $x_{m,i}(S) = nil$ のときは $x_{m,i}(S) = 4$ として計算した。一方、図3は、クラスタ分析のそれぞれの時点における任意の2つのクラスタの重心間の距離 $d(G_{C_1}, G_{C_2})$ を求め、その最小値を算出した結果を示している。

図2と図3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

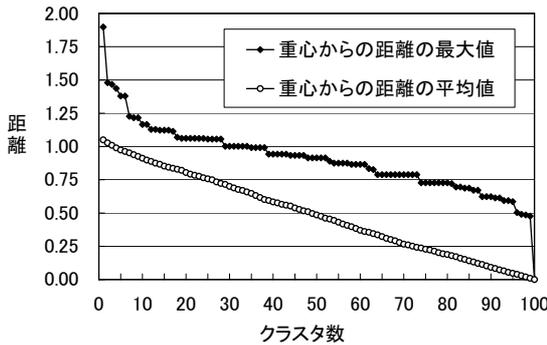


図 2: 各クラスタにおける重心からの距離の平均値と最大値

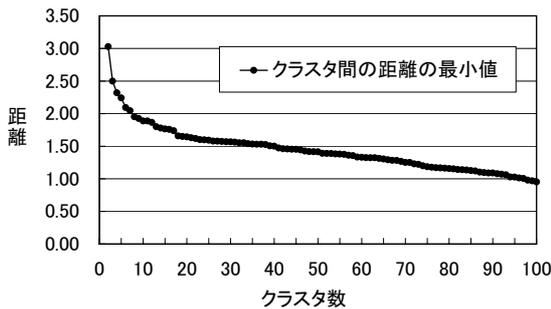


図 3: クラスタ間の距離の最小値

し、その後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。なお、10 クラスタ生成時における各クラスタの個体数は、1, 1, 7, 26, 4, 1, 27, 27, 5, 1 (生成順)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

我々は、文献 [3] において、標準 MIDI ファイル形式の楽曲データ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データの印象を記述する 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印象ベクトル生成式を設計した。すなわち、楽曲データ(80 曲)から抽出される特徴量を説明変数、各楽曲データに対し被験者 100 人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 10 クラスタ/20 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。本稿では、それぞれの印象ベクトル生成式によって楽曲データ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、10 クラスタの場合で 10 個のユーザモデルを、20 クラスタの場合で 20 個のユーザモデルを構築した。なお、重回帰分析における自由度修正済み決定係数は、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選択

ユーザモデルの選択は、検索結果(第 1 位候補曲)に対するユーザの評価を用いて、以下のようにして行われる。

手順 1) 入力された検索条件に対し、ユーザモデル M_n ($n = 1, 2, \dots, 10$ or 20 : クラスタ番号) 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

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

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、その中から距離が最小となるユーザモデル M_{min} を求める。空の場合は、 p_n が閾値 P_{medium} 以上であるユーザモデルの集合 U_{medium} の中から M_{min} を求める。 U_{medium} 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) ユーザモデル M_{min} に対応する楽曲 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) ユーザがその楽曲に対し 5 段階評価(5 点: 適合—4 点—3 点—2 点—1 点: 不適)を行った場合のみ、以下の手順が適用され、行わなかった場合は、手順 1 の待機状態となる。なお、この評価が 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 $score = 1$, $score = -1$ とする。

手順 5) 楽曲 m_{min} を第 1 位としたすべてのユーザモデルにおいて、個人適応値 p_n と評価済曲数 k_n を更新する。

$$p_n = score \quad (p_n = nil : \text{初期値のとき})$$

$$p_n = (p_n \times k_n + score) / (k_n + 1) \quad (\text{それ以外のとき})$$

$$k_n = k_n + 1$$

手順 6) 5 段階評価の点数が 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. 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、本稿では、被験者 100 人に楽曲 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(楽曲データから受ける印象を記述するベクトルの集合)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。

参考文献

- [1] 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.
- [2] 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.
- [3] 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.