

印象に基づく楽曲検索システム：Nグラム統計量の利用

A Music-Retrieval System based on User's Impressions Input: Use of N-gram Statistics

LD-6

熊本 忠彦 太田 公子
Tadahiko Kumamoto Kimiko Ohta

表 1: 印象尺度 1 のための楽曲印象値計算式

楽曲特徴量	係数	楽曲特徴量	係数	定数項
s	-3.835	v4 xs	17.226	5.075
42	-3.775	v7-5-4 d	97.405	
43	-5.844	v10 e	0.559	
46	1.334	v11 u	-5.663	
48	-1.306	v12 u	-4.869	
h44 d	-20.908	d3 u	-7.272	
h62-58 u	59.918	d3 e	-1.877	
h73-64 e	-61.359	d4 e	-2.671	
h78-66 u	-83.845	d51 d	83.808	
h81-78 u	99.110	d61 d	21.226	

1. まえがき

音楽に詳しくない人でも簡単に利用できる楽曲検索システムとして、10 対の印象尺度（楽曲印象を形容する語句の対からなる評価尺度）[熊本 02a]の中から 1 対以上を選択し、7 段階評価することにより動作するシステムを提案する。

本システムの最大の特徴は、楽曲から抽出される物理的特徴量（楽曲特徴量と呼ぶ）として、音の高さや強さ、長さ、音色に関する N グラム統計量を採用している点にあり、これにより楽曲のダイナミックな変化を表現することが可能となっている。先行研究では、楽曲特徴量として音の高さや強さ、リズムやテンポ、調性（短調/長調）といった音楽構成要素に対する平均や分散、時間的割合などの静的な特徴量を用いていることが多い[佐藤 01, 池添 01]。一部では、連続する 3 音の音の高さや長さの推移をパターン化した、時間的推移を考慮した特徴量も試されているが、連続する音の数が一定 (= 3) であり、限定的な推移しか取り扱えなかった[辻 97, 佐藤 01]。

本論文では、提案システムの構成を示すとともに、被験者 25 名による主観評価実験の結果について述べる。なお、本システムが取り扱える楽曲形式は標準 MIDI ファイルであり、楽曲ジャンルとしてはクラシック系を対象としている。

2. システムの構成

本システムは主に、指定された楽曲の印象値（印象尺度毎に 1 個の実数値）を算出し、既定のファイルに登録する「楽曲印象値自動付与部」と、ユーザ入力と楽曲印象値の距離を計算する「楽曲検索部」から構成される。以下、この 2 つの処理部について述べる。

2.1 楽曲印象値自動付与部

まず、楽曲印象値の計算方法であるが、筆者らは、楽曲から抽出される物理的特徴量を説明変数、印象尺度 m ($m = 1, 2, \dots, 10$) における楽曲印象値（被験者 100 名が 80 曲を聴取し、付与した印象値の平均値）を目的変数とする重回帰分析（変数増加法）[菅 00]を行い、楽曲印象値計算式（重回帰式）と楽曲特徴量（重回帰式を構成する説明変数）を決定した[熊本 02b]。その一例を表 1 に示す。

表 1 において、「s, 42, ..., 48」は音色に関する特徴量であり、「s」は無音を表し、他の数値は GM 規格に基づく音色番号に対応している。第 1 要素に「h, v, d」のいずれかを付された特徴量は、それぞれ連続した N 個（印象尺度 1 の場合は $N = 2$ ）の音における音の高さ、強さ、長さの時間的推移を表す特徴量であり、それぞれノートナンバー値、オンベロシテイ値、発音開始時刻から消音時刻までの時間（ミリ秒）に対応している。「-」でつながれた音は同時に発音された音を表しており、例えば「h62-58」は音の高さ 62 の音と 58 の音が同時に発音されたことを示している。また、各特徴量の第 1 音 x_1 と第 2 音 x_2 にはそれぞれ表 2, 表 3 のルールが適用されている。表 3 における直前音 x_{i-1} と対象音 x_i の比較は、同時発音された音が対象の場合はその最大値と最小値を用いて行われ、単音が対象の場合は最大値 = 最小値として行われる。

なお、各楽曲特徴量の値は、その特徴量の重み付き出現頻度をその特徴量と同じ種類、同じ N グラムである特徴量の重み付き出現頻度で割ることにより求められる。但し、重み付けの方

表 2: 第 1 音 x_1 の抽象化ルール

音の高さ・音色の場合は、何もしない。
音の強さの場合は、 $x_1/10$ の商に 1 を足した値で置換する。但し、 $x_1 = s$ (無音) のときは何もしない。
音の長さの場合は、 $x_1/50$ の商に 1 を足した値で置換する。但し、 $x_1 = 0$ (無音) のときは何もしない。

表 3: 第 2 音以降 x_i ($i = 2, \dots, N$) の抽象化ルール

直前音 x_{i-1}	対象音 x_i	記号
s	*	sx
*	s	xs
(*...*)	(大きい...大きい)	u
(*...*)	(大きい...同じ)	u
(*...*)	(大きい...小さい)	ud
(*...*)	(同じ...大きい)	u
(*...*)	(同じ...同じ)	e
(*...*)	(同じ...小さい)	d
(*...*)	(小さい...大きい)	du
(*...*)	(小さい...同じ)	d
(*...*)	(小さい...小さい)	d

記号「*」は任意の有音（s 以外）を表している。

法には表 4 に示す 3 タイプがあり、印象尺度によって使い分けられている。楽曲特徴量として用いられる N グラムも同様であり、まとめると表 5 のようになっている。なお、音色に関する特徴量には、すべての印象尺度において unigram のみが用いられている。

2.2 楽曲検索部

さて、ユーザが入力できる印象は、表 6 に示した 10 対の印象尺度に対する 7 段階評価（1 点, 2 点, ..., 7 点）もしくは無得点の組み合わせ ($8^{10} - 1$ 通り) であり、印象尺度 i ($i = 1, 2, \dots, 10$) におけるユーザの入力を x_i 、楽曲 m の印象尺度 j ($j = 1, 2, \dots, 10$) における楽曲印象値を $v_{m,j}$ とすると、ユーザ入力と楽曲 m の距離 d_m は次式で計算される。

$$d_m = \sum_{k=1}^{10} \frac{(x_k - v_{m,k})^2}{\sigma_k^2} / \sum_{n=1}^{10} l_n, \quad \sigma_k^2 = \sum_{m=1}^{80} \sigma_{m,k}^2 / 80$$

但し、 x_k の値が無得点のときは、 $x_k = v_{m,k}$, $l_k = 0$, そうで

表 4: 3 種類の重み付け方法

タイプ	特徴量の種類	重み
w_1	全種類	1
w_2	音の長さ 上記以外	1 \sum 音の長さ
w_3	音の長さ 音の強さ 上記以外	1 \sum 音の長さ \sum (音の強さ ² × 音の長さ)

表 5: 各印象尺度において用いられる N グラムと重みタイプ

印象尺度	N グラムの組み合わせ	重みタイプ
1	bigram	w_1
2	bigram, trigram	w_1
3	unigram	w_2
4	bigram, trigram, 4-gram	w_3
5	unigram	w_3
6	bigram, trigram, 4-gram	w_1
7	bigram	w_1
8	bigram, trigram, 4-gram	w_2
9	bigram, trigram	w_1
10	bigram	w_3

表 6: 筆者らが提案している印象尺度 (10 対)

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 - 綺麗なでない / 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

ないときは、 $l_k = 1$ とする。また、 $\sigma_{m,k}^2$ は、重回帰分析で用いた被験者データ (80 曲 × 100 人分) の楽曲 m 、印象尺度 k における分散である。

楽曲検索部は、以上の式を用いて、ユーザ入力と各楽曲との距離を計算し、距離が小さい楽曲を最大 20 曲提示する。

3. 主観評価実験

被験者 25 名 (男性 12 名, 女性 13 名) による主観評価実験を行った。被験者は検索条件を自由に入力でき、その結果に対し、適合度評価 (5 段階評価で 5 点満点) を行った。その結果を表 7 に示す。

表 7 は検索順位毎の適合度の平均値であり、学習データは重回帰分析で用いた 80 曲を検索対象とした場合の結果を、未知データは異なる 80 曲を検索対象とした場合の結果を示している。

表 7 から、学習データに対する性能は十分といえるが、未知データに対しては改善の余地が見受けられる。その方策としては、以下の 3 点が考えられる。

(1) 各個人の音楽感性 (楽曲印象の受け方) に適応する。

2.2 で定義した印象尺度 k における分散の平均値 $\bar{\sigma}_k^2$ という概念は、数学的には厳密ではないが、被験者が受けた楽曲印象における個人差の程度を示す指標と考えられ、表 8 のような値となっている。いずれの印象尺度においても 1 前後であることから、個人差は比較的大きいと言え、楽曲検索において個人適応が必要であることを示している。

(2) 楽曲印象値計算式をニューラルネットワークに変更する。

先行研究において、重回帰式を用いた場合とニューラルネットワーク (NN) を用いた場合とで検索精度が比較されているが、総じて NN の方が高い検索精度を得ている [池添 01]。

表 7: 楽曲検索・聴取による適合度評価実験の結果

	1 位	2 位	3 位
学習データ	4.3	4.1	3.9
未知データ	3.8	3.8	3.7

表 8: 各印象尺度における楽曲分散の平均値

印象尺度	σ_k^2 の値	印象尺度	σ_k^2 の値
1	1.0623	6	1.2165
2	1.1212	7	1.2549
3	1.1532	8	0.9221
4	0.9792	9	1.3917
5	1.1760	10	1.2188

(3) N グラムだけでなく静的な特徴量も併用する。

N グラムを用いた楽曲検索研究は今までなかったもので、今回は敢えて N グラムだけを用いた。今後は、従来から用いられている静的な特徴量との組み合わせも考えていく必要がある。

なお、被験者に対して実施したアンケートにおいて、システムの使いやすさ (100 点満点) に対しては、86.8 点を得ており、画面・操作手順等の設計に関しては合格点と考えられる。

4. まとめ

本論文では、筆者らが開発した印象尺度ベースの楽曲検索システムの構成と主観評価実験の結果について述べた。

筆者らは、このようなシステムのユーザを音楽に詳しくない人と仮定している。そのため、主観評価実験だけでなく楽曲印象値計算式の設計においても、音楽経験の乏しい人を被験者として優先的に採用した。したがって、システム同士の性能を比較するためには、共通の音楽データベース [後藤 02] を利用するというだけでなく、同じ被験者による評価実験が不可欠と考えられる。この目的のため、本システムを公開することを予定している。

筆者らが最終的に目指すシステムは、検索したい楽曲の印象を自然言語で入力すれば、その人の音楽感性 (楽曲印象の受け取り方) や言語感覚 (印象を表現する言葉の選び方) を考慮した上で、そのような印象を持つ楽曲を出力するというものであり、今後の課題である。

参考文献

- [後藤 02] 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: クラシック音楽データベースとジャズ音楽データベース, 情処研報, 2002-MUS-44, 5, pp. 25-32 (2002).
- [池添 01] 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, 情処学論, 42, 12, pp. 3201-3212 (2001).
- [管 00] 菅 民郎: 多変量統計分析, 現代数学社, 京都 (2000).
- [熊本 02a] 熊本忠彦, 太田公子: 印象に基づく楽曲検索: 検索ニーズに合った印象尺度の設計, 情処研報, 2001-NL-147, 6, pp. 35-40 (2002).
- [熊本 02b] 熊本忠彦, 太田公子: 印象に基づく楽曲検索: 楽曲印象値の自動付与, 情処研報, 2002-DBS-127, 12, pp. 89-96 (2002).
- [佐藤 01] 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, SP2000-137, pp. 51-56 (2001).
- [辻 97] 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, SP96-124, pp. 17-24 (1997).