

単語共起頻度データベースを使用した 任意の言葉の印象に合った楽曲検索

A study on impression estimation using arbitrary words
for music retrieval system based on database of word co-occurrence frequency

頭川 愛†
Ai Zukawa

酒向 慎司†
Shinji Sako

北村 正†
Tadashi Kitamura

1. はじめに

近年の楽曲検索において、キーワードを用いた検索以外にも多様な手法が提案されている。その一例として感性検索が挙げられ、ユーザが求める楽曲の印象を入力とすることから、具体的な曲名などがわからないときにも有効であり、未知の音楽を発見できる効能も期待できる。

このような感性に基づいた楽曲検索の研究はこれまでもいくつか提案されているが、一定の印象語から選択する方法など、検索システムの入力方法が限定的なものが多かった。しかし、音楽を言葉で言い表す際には様々な語が用いられ、複数の語や文章のような形で表現されることも多いことから、音楽を形容するための語は種類や数に制約のない形が望ましいと考えた。

そこで、限られた印象語と楽曲の対応付けを明らかにした Hevner の研究 [1] を利用し、あらゆる言葉を Hevner の印象語群で表すことで任意の言葉の入力を可能とした楽曲検索を提案する。各軸を印象語群と対応させた空間を設定し、印象が明らかである言葉を代表語としていくつか配置することで、それらを基準として入力語の位置を決定する。その点に最も近い楽曲が入力語にふさわしいものであるといえる。

先行研究では、ある言葉と類似している語として共起確率の上位 100 語を使用した。代表語が少ないため入力語が適切に印象空間に写像されないという問題が生じた。これは、より下位にある言葉も類似語として用いることで解決されると思われる。本研究では、類似語を増やすことで検索できる語数を多くし、かつ共起確率をどれほどの語数用いると最適であるかを実験によって調べた。

2. 印象空間による楽曲検索

2.1 印象空間の設定

心理学者 Hevner は、楽曲と印象の関係を説明するモデルとして、感性を表す 66 語を 8 つの印象語群 C_1, C_2, \dots, C_8 (図 1) に分け、それらと個々の楽曲の特徴を表す楽曲構成要素 (調性、テンポ、音高、リズム、和声、旋律) との関係を明らかにした。ここでは、それぞれの印象語群を独立した軸とする 8 次元空間を印象空間 S とする。

2.2 楽曲の写像

Hevner の手法では楽曲構成要素 $\mathbf{v} = (v_1, v_2, \dots, v_6)^t$ と、楽曲の印象に相当する印象語群の位置 $\mathbf{c} = (c_1, c_2, \dots, c_8)^t$ との関係を表すための相関行列 T (表 1) を定めている。本研究もこれに倣い、式 (1) によって楽曲を印象空間へ写像する。なお、楽曲構成要素 \mathbf{v} の算出は芳村の研究 [2] を参考とする。

$$\mathbf{c}^t \cdot \mathbf{T} = \mathbf{v}^t \quad (1)$$

†名古屋工業大学大学院工学研究科, Graduate School of Engineering, Nagoya Institute of Technology

	C_8	C_1	C_2		C_3
	力強さ	神聖さ	悲しさ		憧れ
	勇ましさ	高貴さ	陰気		しなやかさ
	壮大さ	畏怖	暗さ		柔らかさ
C_7	重々しさ	威厳	憂鬱さ		夢見る
面白さ					
興奮さ					
激しさ					
動揺					
	C_6	C_5	C_4		
	楽しさ	滑稽さ	平穩さ		
	陽気さ	活発さ	のんびり		
	嬉しさ	繊細さ	静かさ		
	華やかさ	優美さ	ゆっくり		

図 1: Hevner の印象語群

表 1: 印象語群と楽曲構成要素の相関行列 T

	v_1 :調性	v_2 :テンポ	v_3 :音高	v_4 :リズム	v_5 :和声	v_6 :旋律
C_1	0.047	-0.121	-0.114	0.290	0.040	0.160
C_2	-0.142	-0.104	-0.218	0.048	-0.094	0.000
C_3	-0.238	-0.139	0.069	-0.125	0.541	0.000
C_4	0.035	-0.173	0.092	-0.032	0.135	0.120
C_5	0.250	0.052	0.183	0.129	0.162	-0.120
C_6	0.285	0.173	0.069	-0.161	0.216	0.000
C_7	0.000	0.182	-0.103	0.032	-0.189	-0.280
C_8	0.000	0.052	-0.149	0.161	-0.108	-0.320

3. 言葉の空間写像

印象語群に含まれる言葉は代表的なものだけでなく、未知の言葉の印象空間上の位置を決める必要がある。ここでは、印象語群に関係する語を媒介にし、それらの語の間にある類似性を利用することで、未知の語の印象空間上の位置を推定する手法を提案する。

3.1 代表語の配置

印象語群に含まれている印象語に類似した語を代表語と呼ぶ。意味が類似している言葉同士は同じ文書内で使われることが多いと仮定し、単語間の共起確率を類似関係を表す尺度として用いる。2 つの言葉 w_1, w_2 の共起確率 $P(w_1, w_2)$ は、 $F(w_1)$ を w_1 の出現頻度、 $F(w_2)$ を w_2 の出現頻度、 $F(w_1, w_2)$ を w_1, w_2 の共起頻度としたとき、式 (1) で定義される。

$$P(w_1, w_2) = \frac{F(w_1, w_2)}{F(w_1) + F(w_2)} \quad (2)$$

各印象語群につき 4 語を印象語とし、それらとの共起確率の上位いくつかを代表語として S の軸上に配置する。

3.2 入力語の写像及び検索方法

入力語 w に対して、代表語 r_1, r_2, \dots, r_n の n 個と共起するとき、 w の各軸の座標 $c_i(w)$ は、式 (3) で算出する。

$$c_i(w) = \frac{\sum_{j=1}^n P(w, r_j) \cdot P(C_i, r_j)}{\sum_{j=1}^n P(w, r_j)} \quad (i = 1, 2, \dots, 8) \quad (3)$$

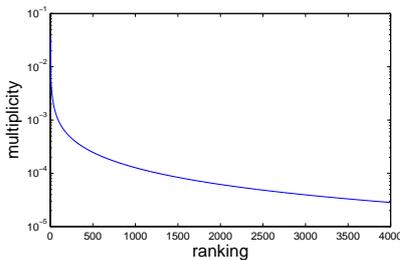


図 2: 4000 位までの共起確率の推移の平均

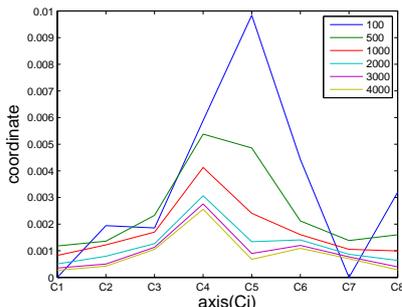


図 3: 「明るい」の座標

印象空間 S において、入力語と全ての楽曲のユークリッド距離を測定し、最も距離の近い楽曲が入力語と最も印象が合っているとして検索結果となる。

4. 予備実験

本手法には、2つの言葉の共起確率を示すコーパスとして、ALAGIN 言語資源・音声資源が配布する単語共起頻度データベースを使用する。これは、約1億ページのウェブ文書に含まれる語のうち、477458語の共起確率を計算したものである。先行研究では上位100語のデータを使用した。より多くの言葉を用いることで言葉が適切に写像されると考えられる。どれほどの言葉を使用すべきか検討するために、2つの予備実験を行った。

4.1 共起確率の推移

初めに、共起確率はどのように変化していくのかを調べた。4000位までの共起確率の平均推移を図2に示す。この結果より共起確率は上位のみが大きく、下位は同等の小さい値を示すことがわかる。

4.2 空間写像の検討

また、比較的印象が明確である言葉について、印象空間上のどのような位置に写像されるかを調べた。その一例を図3に示す。横軸が C_1, \dots, C_8 に対応している軸を表し、縦軸はそれぞれの値を表している。この結果より500語を超えると軸どうしの値の差が減り、言葉の特徴を表すことができなくなると考えられる。

5. 実験

4章の結果より、類似語の使用語数は100語から500語の間が適切であると思われる。その範囲で検索できる語数と主観評価実験により最適な使用語数を見つける。

5.1 検索可能語数の変化

本手法では、代表語を多いほど検索できる言葉が増えると期待される。増加の傾向を見るために、100語から500語の検索可能語数を調べた。結果を表2に示す。使

表 2: 検索可能語数

使用語数	100	200	300	400	500
検索可能語数	228342	402817	458295	472227	475457

表 3: 主観評価実験の結果

使用語数	100	200	300	400	500
ドラマティック	3.6	1.5	2.9	3.2	3.2
沈む	2.4	2.9	2.9	2.9	3.5
華美	1.7	3.8	3.8	3.8	1.7
元気	1.9	2.8	3.3	1.9	1.7
全体平均	2.4	2.8	3.2	3.0	2.5

用語数が100語と200語では約2倍に増え、徐々に増加率は減るが300語では96%を占めている。

5.2 主観評価実験

5.2.1 実験方式および実験条件

入力語に対し、検索された楽曲がどれほど印象が合っているかを5段階で評価する。ただし、全く合っていない場合を1、非常に合っている場合を5とする。類似語の使用語数を100語から500語まで変化させ評価を比較した。なお、同じ楽曲が出力された場合は一度のみ評価する。

楽曲はRWCデータベース[3]内のクラシック音楽61曲のMIDIデータを使用する。被験者は10名、入力語とした言葉は「ドラマティック」「沈む」「華美」「元気」の4語である。

5.2.2 実験結果および考察

被験者10名の評価平均を表3に示す。先行研究の使用語数である100語のときに比べ、全体平均が0.8ほど向上したことがわかった。しかし、入力語によって大きく差があり、「ドラマティック」では100語使用したとき、「沈む」では500語使用したときに最も評価が高くなった。全ての結果を見ると、他の語数に比べ300語使用したときに評価が安定しており、全体平均も最高であることから300語が最適語数であると言える。しかし、評価は全体的にかなり低くなってしまった。

6. むすび

本研究では、任意の言葉を入力とした楽曲検索の実現に向け、言葉を複数の代表語で近似的に表現するという手法に対し、言葉の類似関係として使用する共起語の最適語数を明らかにした。主観評価実験の結果、上位300語を使用したときに最も評価が高くなり、またこのとき約45万語という任意に近い語数を検索可能であるということが示された。しかし、評価は高いとは言えず、さらなる手法の改善が必要である。今後の課題として、より多くの入力語での実験結果を用いた本手法の有効性の検証や、Hevnerの研究以外の先行研究の使用などが挙げられる。

参考文献

- [1] Hevner, K., "Experimental studies of the elements of expression in music", *American Journal of psychology*, Vol.48, pp.246-268, 1936.
- [2] 芳村 亮 他, "任意の言葉の印象と音楽心理学に基づく楽曲自動生成方式", DEWS2007 論文集, 電子情報通信学会, 2007.
- [3] RWC 研究用音楽データベース, <http://staff.aist.go.jp/m.goto/RWC-MDB/index-j.html>