6R-6

印象空間を用いた任意の言葉による楽曲検索：言葉の写像方法の改善と評価方法の再構築

頭川 愛 †，酒向 慎司 †，北村 正 †
† 名古屋工業大学大学院工学研究科

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に基づいた検索が挙げられ，ユーザが求める楽曲の印象を入力することから，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．

従来の感性検索の研究では，一定の感性語から選択したりSD 法を用いたりする入力方法が主流であった．[1] また近年では，印象を表すのに空間を用いる研究がなされている．[2] 印象を空間上の座標で表すことで，より曖昧な表現が可能となる．しかし，ユーザ自身がイメージした印象を座標に置き換えることは困難であることから，空間における座標と言葉を結び付けると更に利便性が向上すると考え，印象を表す空間を用いることで任意の言葉から楽曲を検索する手法を提案した．[3]

14 の感性語対を用いて印象空間を生成し，その空間に感性語の関係の深い言葉を配置して代表語と名付けた．そして，それらの代表語を用いて任意の言葉の座標を決定した．しかしながら，従来は膨大な言葉の中の関係性を用いていたため，関係性の値がとても小さくなってしまうという問題があり，代表語の数についても正確な調査を行っていなかった．また，有効性の検証として主に小規模な主観評価実験しか行っておらず，更なる詳細な評価が必要である．

本研究では，代表語の写像方法を見直し，代表語数による検索結果の違いを明らかにした．また，楽曲の写像と言葉の写像それぞれについての評価実験を行い，本手法の有効性を検証する．

2 印象空間の生成

先行研究 [4] で実施した聴取実験の結果を利用する．RWC 研究用音楽データベース [5] のクラシック音楽を評価対象とし，歌声が含まれていない 44 曲から与える印象が一定と考えられる 15 秒間を 1 サンプルとして 100 サンプル使用している．被験者 119 名の男女に全てのサンプルを聴かせ，それぞれに対して 14 の感性語対について SD 法の 7 段階で印象評価をさせた．

聴取実験で得られた評価平均を因子分析し，印象空間 S を生成する．正当な評価がされていないデータを除き，更に一般的な印象空間にするため平均との差が標準偏差以上のものを省いた．因子分析の結果得られた印象空間 S を図 1 に示す．感性語対の対になる言葉は，原点を中心とした点対称の位置に配置する．第 1 因子軸は”明るい”，”陽気な”などが大きい値を示していることから「明るさ」，第 2 因子軸は”激しい”，”慌しい”などが大きい値を示していることから「激しさ」を表していると言える．

3 楽曲の写像

楽曲データから印象空間 S 上の座標を決定するため，楽曲の印象と関係していると考えられる特徴量と聴取実験の結果から写像式を生成する．

聴取実験の結果を重回帰分析することで，特徴量から楽曲の印象空間 S の第 1 軸，第 2 軸それぞれの値を決定する写像式の係数を算出する．目的変数は，因子負荷量が最も高い感性語対から第 1 因子を「明るい–暗い」の評価平均，第 2 因子を「穏やかな–激しい」の評価平均とする．ここで，評価平均は 1 から 7 の値であるが印象空間 S の範囲は $(-1, 1)$ とするため，評価平均 x を式 (1) を用いて正規化する．

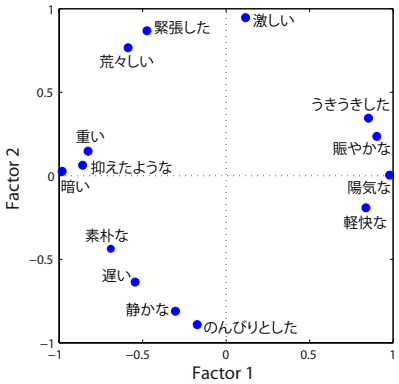

図 1: 印象空間 S

表 1: 楽曲特徴量

音量
テンポ
major と minor の音量差
スペクトルの変化度合
弱音の割合
主音の明瞭度
調性の重心の変化度合
中心周波数
85%を占める低音域の割合
不協和音の多さ
MFCC13 次元
ゼロ交差数

$$x' = \frac{x-4}{3} \tag{1}$$

また，印象によって関係している特徴量は異なると考えられるため，軸ごとに説明変数に用いる特徴量を変数増加法を用いて決定する．変数の候補は，フレームごとに抽出した表 1 に示す特徴量の平均と標準偏差とする．抽出には MIRToolbox[6] を使用した．

分散比 F が 2 未満になるまで変数を追加していった結果，第 1 軸について 16 の特徴量を用いた写像式，第 2 軸について 17 の特徴量を用いた写像式が生成された．

4 言葉の写像

4.1 任意の言葉の写像方法

任意の言葉を個々に表現することは難しいが，印象を表す代表的な言葉との関係性で表すことができると考えた．代表的な言葉は印象空間 S 上に配置された 14 対の感性語がふさわしいが，高々28 語との関係性で任意の言葉を表現するのは困難である．従って，感性語と関係の深い言葉も印象空間 S 上に配置する．ここで，言葉の関係性として，同じ文内で使われる言葉どうしは関係が深いと仮定し，共起確率を使用

Music retrieval system for any words using impression space: Improvement of mapping words and reconstruction of evaluation
Ai Zukawa †, Shinji Sako †, Tadashi Kitamura †, † Graduate School of Engineering, Nagoya Institute of Technology

Copyright ©2014 Information Processing Society of Japan.
All Rights Reserved.

する．約 1 億の Web 文書から算出した ALAGIN[7] の単語共起頻度データベースを利用し，2 つの言葉 w_1, w_2 の共起確率 $C(w_1, w_2)$ 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 (2) で定義される．

$$C(w_1, w_2) = \frac{F(w_1, w_2)}{F(w_1) + F(w_2)} \tag{2}$$

言語の特性から，形容詞や形容動詞よりも名詞のほうが関係の共起しやすいと考えられるため,「軽い」を「軽さ」など全ての感性語を名詞に変換して共起確率を調べ, 大きいものから順に使用する．以後，これらの言葉を代表語と呼ぶ．この代表語から入力語の座標を決定する．

4.2 入力語の座標決定と検索方法

入力語 w に対して共起確率が高い言葉を調べ，それらの中から代表語を探索する．代表語が $r_1, r_2, ..., r_N$ の N 個あった場合，w の座標 (x_{1w}, x_{2w}) は式 (3) のように算出する．

$$\left(\frac{\sum_{i=1}^{N} C(w, r_i) * x_{1r}}{\sum_{i=1}^{N} C(w, r_i)}, \frac{\sum_{i=1}^{N} C(w, r_i) * x_{2r}}{\sum_{i=1}^{N} C(w, r_i)} \right) \tag{3}$$

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求め，距離の近い楽曲ほど印象が似ていると判断し，検索結果として出力する．

4.3 代表語の設定方法の改善

代表語の写像方法について，従来では共起確率の値をそのまま用いていたが，データベース内で扱う言葉全体での関係性であるため値が小さすぎてしまうという問題があった．そのため，代表語と全ての感性語との共起確率を用いることを検討する．A_n と共起確率が高い代表語 r の座標を以下のように決定する．

$$(x_{1r}, x_{2r}) = \left(\frac{C(r, A_n)}{\sum_{i=1}^{28} C(r, A_i)} \cdot x_{1A_n}, \frac{C(r, A_n)}{\sum_{i=1}^{28} C(r, A_i)} \cdot x_{2A_n} \right)$$

複数の感性語で出現した言葉は，それらの中心とする．ここで，印象空間 S の原点に近い代表語はあまり印象を持っていないと考えられるため，原点との距離が小さい代表語を除去することを検討する．本研究では，原点との距離が 0.1 未満の代表語，0.2 未満の代表語を除去した場合をそれぞれ調査する．

また，先行研究 [3] の結果から関係の深い言葉として共起確率の上位 300 語を使用していたが，本手法と条件が大きく異なることから，語数を検討する必要がある．本研究では 300 語以上は必要がないと考え，関係の深い言葉として 100 語，200 語，300 語を使用した場合をそれぞれ調査する．

5 評価実験

5.1 楽曲写像の評価

楽曲が正しく写像されているかを検証するため，印象空間生成で用いた聴取実験の評価結果との比較を行う．第 1 軸, 第 2 軸それぞれにおいて因子負荷量 λ を用いて，評価平均 x を式 (4) のように変換し，楽曲の座標値との差の平均を算出する．

「暗い」,「激しい」といった写像式に用いている感性語だけでなく，第 1 軸では「陽気な」「賑やかな」，第 2 軸では「のんびりとした」「緊張した」など因子負荷量が大きい感性語の評価と楽曲との差が小さいことから，本研究で生成した写像式は有効であることが示された．

$$x' = \frac{x - 4}{3} \cdot \lambda \tag{4}$$

表 2: 楽曲の座標と評価平均の差

	第 1 軸		第 2 軸	
感性語	λ	平均差	λ	平均差
重い	-0.8273	0.1748	0.1481	0.2546
暗い	-0.9822	0.1355	0.0264	0.2515
うきうきした	0.8521	0.1694	0.3455	0.1960
静かな	-0.3034	0.2206	-0.8103	0.1238
激しい	0.1163	0.2420	0.9469	0.0890
陽気な	0.9801	0.1201	0.0036	0.2493
抑えたような	-0.8600	0.1690	0.0640	0.2497
遅い	-0.5445	0.2002	-0.6360	0.1878
荒々しい	-0.5872	0.2323	0.7667	0.2029
賑やかな	0.9023	0.1396	0.2367	0.2224
のんびりとした	-0.1734	0.2405	-0.8916	0.1398
軽快な	0.8368	0.1723	-0.1910	0.2561
緊張した	-0.4747	0.2384	0.8680	0.1953
素朴な	-0.6919	0.1788	-0.4361	0.2185

5.2 言葉写像の評価

言葉が正しく写像されるかを検証するため，類義語とされる言葉どうしがどのような位置に配置されるかを検証した．Weblio 類語辞典 [8] を用いて感性語の類義語を 3 語ずつ調べ印象空間に写像し，感性語が写像された言葉との距離との差を算出する．関係の深い語として用いる語数と，原点からの距離 d による代表語の除去範囲を変化させた結果を表 3 に示す．

関係語数が多く，また代表語の除去を行わないときが最も差が小さくなることがわかった．また，最も距離が近い感性語が正しい語も同じ条件のときが最も多かった．ただし，あまりいい結果とは言えず更に調整が必要である．

表 3: 感性語と類義語の差

関係語数	使用範囲	平均差
100 語	全て	0.4759
	$d \geq$0.9	0.4846
	$d \geq$0.8	0.4932
200 語	全て	0.4181
	$d \geq$0.9	0.4348
	$d \geq$0.8	0.4499
300 語	全て	0.4095
	$d \geq$0.9	0.4290
	$d \geq$0.8	0.4494

6 むすび

本研究では，印象空間を用いた任意の言葉を用いる楽曲検索の実現に向け，言葉の写像に用いる代表語の配置方法を見直し，さらに代表語数を変化させて性能を評価することで最適な条件を決定した．また，1 つの感性語対のみを使用して生成した写像式は，他の感性語対においてもほぼ同等の結果が得られることがわかった．今後の展望として，更なる代表語の調整と大規模な主観評価実験の実施を考えている．

参考文献

[1] 熊本忠彦 他，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，21 巻 3 号 K.
[2] Yi-Hsuan Yang, et al., A Regression Approach to Music Emotion Recognition, IEEE Transactions on Audio, Speech, and Language Processing.
[3] 頭川愛 他，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，FIT2012.
[4] 岩月靖典 他，利用者のプロフィールを用いた個人性を考慮した楽曲の印象推定，HCG シンポジウム 2012.
[5] RWC 研究用音楽データベース, http://staff.aist.go.jp/m.goto/RWC-MDB/index-j.html
[6] Lartillot O, et al., MIR in Matlab(II):A Toolbox for Musical Feature Extraction from Audio, In Proc. ISMIR2009.
[7] ALAGIN 言語資源・音声資源サイト, http://alaginrc.nict.go.jp/
[8] Weblio 類語辞典，http://thesaurus.weblio.jp/

Copyright ©2014 Information Processing Society of Japan.
All Rights Reserved.