

程度語の序列化と自然言語感性検索への応用

熊本 忠彦†

† 独立行政法人情報通信研究機構 けいはんな情報通信融合研究センター
〒 619-0289 京都府「けいはんな学研都市」光台 3-5
E-mail: †kuma@nict.go.jp

あらまし 「楽しい」や「明るい」のような印象情報を入力とする感性検索においては、印象の微妙な違いを表現するための「比較的」や「適度に」といった程度語も入力したい。しかしながら、従来の感性検索に関する研究では、程度語の扱いは限定的であり、多くても2~3語であった。そこで、本稿では、感性検索において利用される程度語119語を類語辞典から抽出し、その意味的役割を数値化する。具体的には、程度語の意味的役割を、修飾している印象語の意味を強める/弱めることと捉え、その強弱関係を被験者実験により決定する。

キーワード 程度副詞, マルチメディア検索, 感性検索, 自然言語インタフェース

Relative Ranking of Degree Modifiers and Its Application to Impression-Based Content-Retrieval by Natural Language Input

Tadahiko KUMAMOTO†

† Keihanna Human Info-Communication Research Center,
National Institute of Information and Communications Technology
Hikaridai 3-5, Kansai Science City, Kyoto 619-0289, Japan
E-mail: †kuma@nict.go.jp

Abstract When we use impression-based multimedia retrieval systems where impressions such as “happy” and “bright” are input, we want to use degree modifiers such “adequately” and “comparatively” in order to more accurately express the impression we want to input. In conventional impression-based retrieval systems, however, at most two or three typical degree modifiers can be accepted. In this article, therefore, we extract 119 degree modifiers from the thesaurus, and numerically express the semantic role of each degree modifier. The semantic role of a degree modifier is defined as strengthening or weakening the meanings of the impression word which has been modified by the degree modifier, and all the degree modifiers are arranged according to their strength.

Key words Adverbs representing a degree, multimedia retrieval, kansei retrieval, natural language interface

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ある特定のコンテンツを探す場合には、書誌情報（題名や製作者名、提供会社名など）や内容情報（歌詞やメロディ、スケッチや色使いなど）による検索[1]~[3]が有効であるが、漠然とした要求しかなく、具体的な情報を入力できないような場合には、不向きと言える。ある特定のコンテンツを探し出すというよりも、自分自身が持つ何らかの判断基準（嗜好や感性、気分など）に合致するものを見つけ出すという場合には、印象情報に基づく検索が有効と考えられ、いくつかのグループによって研究されている[4]~[12]。

印象情報の入力方法としては、システムが提示する(1)複数個の印象語の中から1つ以上を選択する[4],[5],[9]、(2)複数個

の印象語の中から1つ以上を選択し、それぞれを多段階評価（例えば5段階評価）する[8]、(3)複数個の印象尺度（印象語の対となる評価尺度）の中から1つ以上を選択し、それぞれを多段階評価する[10],[11]、といったGUI(Graphical User Interface)ベースの方法が提案されているが、程度語の利用という観点から見てみると、多段階評価に対応した、ごく少数の程度語（せいぜい1~2語）が利用されているにすぎず、十分とは言えない。一方、より自由で制約の少ない印象入力方法として、印象語を含む文を入力とするNLI(Natural Language Interface)ベースの方法[6],[7],[12]も提案されているが、程度語の利用に関してはほとんど述べられていない。しかしながら、印象の微妙な違い（例えば「比較的楽しい」や「やや落ち着いた感じ」など）を入力したいという要求は、当然考えられることから、程度語の意味的役割を処理するための何らかの機構が必要とされる。

程度語の意味的役割に関しては、ファジー集合論の分野で議論が盛んであり、程度語は、語の意味を表現するファジー集合を変換する演算子もしくは関数として扱われている [13]。しかしながら、その研究対象は、典型的な、ごく少数の程度語（「とても」や「少し」など）に限られている上、その演算子/関数を個別に設計しようというスタンスであり、多数の程度語を统一的に扱おうというものではない。

我々の目的は、程度語の意味的役割を記述するための尺度を提案することではなく、検索対象となるマルチメディアコンテンツから抽出される物理的特徴量と程度語によって修飾された印象語とを対応付けることである。我々は、すでに標準 MIDI ファイル形式の楽曲データから抽出される物理的特徴量と 10 個の印象尺度（印象語の対からなり、7 段階評価が可能な尺度）[14]との対応付けを行っており [11]、さらに、印象尺度と印象語 164 語との対応付けも行っている [12]。そこで、本稿では、程度語の意味的役割を、修飾している印象語の意味を強める/弱めることと捉え、その強弱関係を被験者実験により決定する。具体的には、類語辞典 [15] の「程度」の項目から感性検索時に利用される程度語 119 語を抽出し、それぞれの程度語の意味的役割を、検索対象から抽出される物理的特徴量と対応付けられた印象尺度を用いて評価することにより、順序付ける。

2. 楽曲印象のベクトル記述

本章では、我々が提案している、(1) 楽曲の印象を表現するための記述形式 (10 次元のベクトル形式であり、記述されたものを印象ベクトルと呼ぶ) [11]、(2) 標準 MIDI ファイル形式の楽曲データから印象ベクトルを生成するための手法 [11]、(3) 印象語から印象ベクトルを生成するための手法 [12]、の概略を示す。

2.1 印象ベクトルの記述形式

印象ベクトルは、10 次元のベクトルであり、各要素は表 1 に示した 10 個の印象尺度 [14] に対応している。すなわち、各要素の意味は対応する印象尺度を構成する印象語の意味によって定義付けられ、その値は印象尺度の評定スケール (7 段階評価値) に準拠する。例えば、楽曲の楽しさ/悲しさに関する印象を表すには、印象尺度 8 に点数を与えればよく、その点数が「とても楽しい (7 点) — 楽しい (6 点) — 少し楽しい (5 点) — どちらとも言えない (4 点) — 少し悲しい (3 点) — 悲しい (2 点) — とても悲しい (1 点)」という評定スケールのどのあたりに位置しているかによって、印象の程度 (度合い) が示される。このとき、他の印象尺度に *don't care* 項を意味する記号 *nil* を割り当てることにより、印象ベクトルは、

$$(\text{nil nil nil nil nil nil nil } x \text{ nil nil})$$

という 10 次元のベクトル形式で記述される。但し、記号 x は印象の程度を表す実数値であり、0 点~8 点の値をとる。

2.2 楽曲データから生成される印象ベクトル

我々は、楽曲データから抽出される物理的特徴量 (音の高さ、長さ、強さ、音色に関する時系列データ) を説明変数、印象尺度 m ($m = 1, 2, \dots, 10$) における評価値 (被験者 100 名がクラシック 80 曲を聴取し、印象尺度 m に与えた 7 段階評価値の平均値) を目的変数とする重回帰分析 (変数増加法) を行い、楽曲データからその印象を記述する印象ベクトルを生成する手法^[注1]

[注1]: 印象ベクトルの第 m 要素の値が、対応する重回帰式 (重回帰式を構成する説明変数とその係数、ならびに定数項) を用いて求められる。値が 0 点以下の場合は 0 点に、8 点以上の場合は 8 点に丸められる。

表 1 印象語の対からなる印象尺度 [14]

印象尺度	印象尺度を構成する印象語
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

を提案した [11]。参考のために、本手法を用いて楽曲データから生成された印象ベクトルの例を表 2 に示す。

2.3 印象語から生成される印象ベクトル

入力された印象語は、表 3 に示されたような印象語解釈ルール [12] を用いて、印象ベクトルに変換される。

この印象語解釈ルールは以下のようにして作成された。まず、被験者 100 人 (男性 50 人、女性 50 人) に印象語 (全部で 144 語^[注2]) と印象尺度 (10 個) を紙ベースで提示し、その印象語で表現される曲 (例えば「優しい曲」) はどのようなイメージの曲か、ということをもつ以上の印象尺度に対する 7 段階評価という形で回答してもらった。その結果、印象語と各印象尺度における 7 段階評価値との対応関係が得られたが、このデータの中には被験者による変動が大きいものや、不適切な印象尺度と考えた被験者が多いものもあった。そこで、それぞれの印象語毎に、データの平均値 \bar{x} と標準偏差 σ を求め、区間 $(\bar{x}-1, \bar{x}+1)$ の間に含まれている被験者の数 (被験者による得点分布を正規分布と仮定した場合の理論上の数) を計算した。そして、この数が被験者 100 人の過半数に相当する 51 人以上であるデータだけを採用することにし、この条件を満たさなかったデータは除去した (実際には記号 *nil* を付与し、*don't care* 項として扱った)。その結果得られた対応表が印象語解釈ルールである。

3. 程度語の序列化

3.1 程度語に関する検討

類語辞典 [15] の「程度」の項目から感性検索時に利用される程度語 119 語 (表 4 参照) を決定した。検索対象を楽曲に限った場合、不適切な語もあったが、あえて広く決定した。

さて、程度語の意味的役割は、その係り先である印象語が持つ意味を強めたり、弱めたりすることと言える。したがって、程度語を受理できるようにするためには、印象語から生成された印象ベクトルの大きさを拡大または縮小するための仕組みが必要とされる。印象尺度の評定スケール上では、7 段階の評価が可能であり、「少し」や「とても」によって修飾された印象語も評価することができる。このとき、印象語そのものに対しては 6 点、「とても」によって修飾された印象語には 7 点、「少し」の場合は 5 点という点数が与えられていることから、程度語によるベクトルの大きさの拡大・縮小はこの評定スケールに準じたものである必要がある。すなわち、ある印象語のある要素が 6 点であるとき、その印象語が「とても (少し)」によって修飾されているなら、最終的に生成される印象ベクトルの当該要素は 7 点 (5 点) であることが求められる。

[注2]: 我々の印象尺度で使われている 20 語を除いた。

表2 楽曲データから生成された印象ベクトルの例

楽曲データ名	1	2	3	4	5	6	7	8	9	10
annie_laurie_pi	4.94	5.16	4.26	4.28	4.30	4.96	5.57	3.63	5.01	5.36
appassionata3_pi	1.70	1.90	3.02	3.16	5.40	3.04	4.69	3.26	2.50	3.92
arabesque1_pi	3.86	4.07	4.09	3.70	4.40	4.24	5.14	3.89	4.19	4.50
ave_maria	6.10	5.93	3.91	3.29	4.59	5.15	5.43	3.44	5.05	5.24
eine_kleine1_str	2.49	3.80	4.45	6.05	4.89	4.20	5.20	5.83	3.76	4.53
fuge_g_or	3.67	4.10	2.59	2.41	6.00	4.03	5.44	2.88	3.90	4.23
marche_tsc_pi	2.80	2.73	4.48	5.72	4.19	3.72	4.87	5.55	2.91	4.46

表3 印象語解釈ルール (一部)

印象語	1	2	3	4	5	6	7	8	9	10
哀れな	nil	nil	2.48	2.13	nil	nil	nil	1.75	nil	nil
懐かしい	5.01	5.40	nil	4.47	nil	5.37	nil	nil	nil	5.72
古典的な	5.42	5.56	nil	3.47	5.57	nil	5.51	nil	5.06	5.09
心地よい	5.30	5.61	5.52	5.28	nil	5.66	6.01	5.21	6.03	6.13
優しい	5.49	5.79	5.62	5.27	nil	5.62	6.01	5.10	5.85	6.16
力強い	2.13	nil	nil	nil	nil	nil	nil	nil	2.38	nil

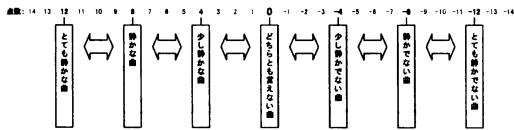


図1 点数の基準表 (29段階評価)

3.2 被験者実験に基づく程度語の序列化

印象語から生成された印象ベクトルは、その印象語を修飾している程度語に応じて、定数倍される。すなわち、ある印象語から生成された印象ベクトルを $\vec{v} = [v_i] (i = 1, 2, \dots, 10)$ 、その印象語を修飾している程度語の定数を a とすると、程度語によって修飾された印象語の印象ベクトルは、

$$\vec{v}' = [a(v_i - 4) + 4] \quad (i = 1, 2, \dots, 10) \quad (1)$$

と表される。この操作は、 a が1より大きければ拡大であり、小さければ縮小を意味する。また、 v_i の値が4より大きければ、その印象尺度 i の左側にある印象語の意味に対する操作であり、小さければ、右側にある印象語の意味に対する操作となる。4点(「どちらとも言えない」)のときは、 a の値に関係なく、元の値が保持され、「どちらとも言えない」のままとなる。

それぞれの程度語の定数を求めるために、以下のような被験者実験を行った: まず、被験者100人(男性50人、女性50人)^(注3)に程度語(全部で119語)によって修飾された印象語(実際には印象尺度1の左側の印象語「静かな」を用いた)と図1のような点数の基準表を紙ベースで提示し、その程度語と印象語の組み合わせによって表現される曲(例えば「わりと静かな曲」)が点数の基準表に照らし合わせてどのあたりに位置するか、ということに回答してもらった。この点数の基準表は、「どちらとも言えない」を原点(0点)とする数直線であり、「とても静かな曲」より静かな曲と「とても静かでない曲」より静かでない曲を想定することにより、29段階の評価が可能となっている。

以上の実験の結果、それぞれの程度語に対し、図2~図8に示したようなヒストグラムが得られたが、ヒストグラムの形状

(注3): 印象語解釈ルール作成のための実験に参加した被験者とは重複しない。

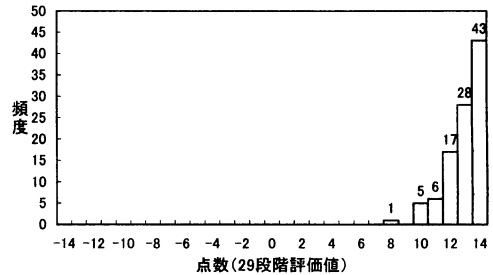


図2 「極端に (静かな)」の評価結果

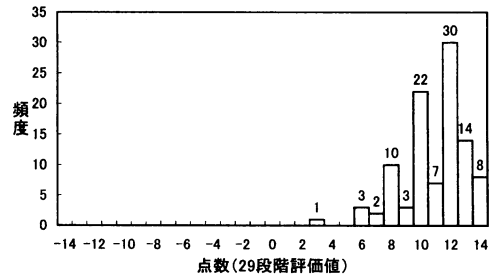


図3 「本当に (静かな)」の評価結果

表5 程度語定数表 (一部)

最高に (静かな)	1.75	比較的 (静かな)	0.54
本当に (静かな)	1.57	少し (静かな)	0.5
とても (静かな)	1.5	やや (静かな)	0.47
割と (静かな)	0.63	かすかに (静かな)	0.18

は多様であり、正規分布と言えないものも多い。そこで、単純に平均を取るのではなく、図9に示したアルゴリズムを用いて、各程度語の代表値を求めることにした。そして、それぞれの代表値を印象尺度の評定スケール(7段階評価値)に換算し、式(1)を用いて逆算することにより、各程度語の定数 a を決定した。ここで、各程度語の代表値を表4にまとめるとともに、程度語定数の例を表5に示す。なお、当アルゴリズムにおいて、移動平均を用いたのは、程度語の多義性や個人差に由来する評価結果のばらつきに対し、最も多くの人に支持された解釈を選び出すためであり、被験者数が過半数を超える最小範囲内のデータだけで平均を求めたのは、少なくとも過半数の人が支持するような解釈に対し、その平均的な解釈を決定するためである。

4. 性能評価実験による検証

我々は、程度語によって修飾された印象語から印象ベクトルを生成する手法を楽曲検索システム[12]に実装し、被験者40名

表4 図9の手法により得られた各程度語の代表値（程度語の順序関係）

どはずれて～	14	極めて～	13.39	大層～	12.5	結構～	7.66	まあまあ～	4.21
この上なく～	14	ものすごく～	13.38	えらく～	12.5	中々～	7.65	少し～	4
最高に～	14	百パーセント～	13.26	とっても～	12.4	平凡に～	6.98	～め	3.88
とんでもなく～	14	一等～	13.11	すごく～	12.39	平々凡々に～	6.82	やや～	3.76
とてつもなく～	14	至極～	13.1	最低に～	12.32	並に～	6.78	いささか～	3.71
最上に～	14	あまりに～	13.03	うんと～	12.09	月並みに～	6.76	幾らか～	3.22
桁外れて～	14	めっちゃめっちゃ～	12.95	相当～	12.06	適度に～	6.53	少々～	3.16
一番～	14	超～	12.93	非常に～	12	程合いに～	6.44	幾分～	3.12
最も～	14	断然～	12.91	大変～	12	中ぐらいに～	6.41	ちょっと～	3.08
ウルトラ～	13.67	スーパー～	12.91	とても～	12	頃合に～	6.3	ほのかに～	2.83
猛烈に～	13.63	真に～	12.9	馬鹿に～	11.77	ほどほどに～	6.14	こころもち～	2.09
桁違いに～	13.63	格段に～	12.89	ひときわ～	11.71	お手頃に～	6	ちょっぴり～	2.07
極度に～	13.61	めっちゃくちゃ～	12.86	やけに～	11.57	適当に～	5.98	中途半端に～	2
極端に～	13.61	著しく～	12.83	かなり～	11.12	まずまず～	5.81	少なくとも～	1.8
べら棒に～	13.59	ひどく～	12.81	いやに～	10.98	なるべく～	5.23	わずかに～	1.8
～すぎる	13.57	すこぶる～	12.79	大いに～	10.86	なるたけ～	5.15	かすかに～	1.46
圧倒的に～	13.56	十二分に～	12.75	ずいぶん～	10.75	そこそこ～	5.09	さほど～でない	-3.81
むちゃくちゃ～	13.56	高度に～	12.72	ひとしお～	10.57	割と～	5.04	さして～でない	-4.11
段違いに～	13.55	めっぽう～	12.65	存分に～	10.24	割合～	4.87	大して～でない	-4.32
どえらく～	13.53	段ちに～	12.6	実に～	10.19	割り方～	4.78	それほど～でない	-5.07
底抜けに～	13.53	まったく～	12.58	全然～	9.68	まあ～	4.71	あんまり～でない	-5.97
過度に～	13.52	本当に～	12.58	十分～	8.54	わりかし～	4.67	あまり～でない	-6
飛び切り～	13.5	めっちゃ～	12.53	だいぶ～	8.51	割に～	4.57	よほど～でない	-10.34
はなはだしく～	13.47	はなはだ～	12.53	普通に～	8	比較的～	4.28		

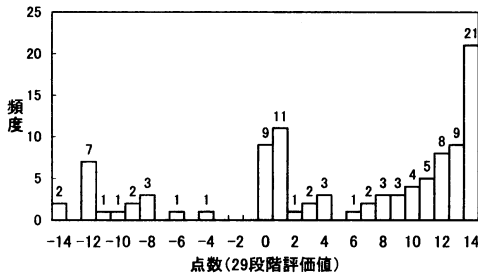


図4 「最低に（静かな）」の評価結果

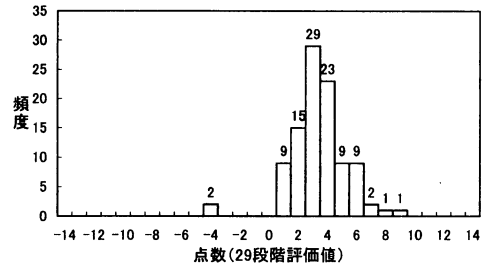


図6 「幾分（静かな）」の評価結果

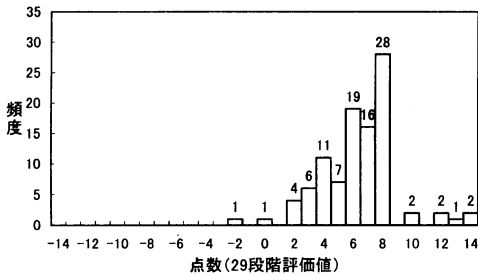


図5 「適度に（静かな）」の評価結果

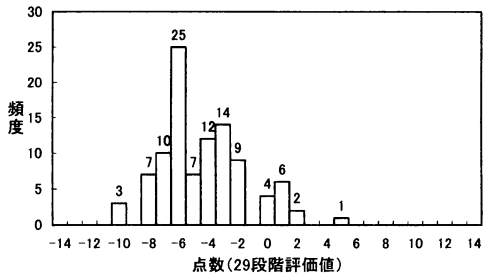


図7 「それほど（静か）でない」の評価結果

(男性 20 名, 女性 20 名) による性能評価実験を行った。このとき、作成された程度語定数表の妥当性を検証するために、程度語補正係数 b を導入し、式 (1) を、

$$\bar{v} = [ba(v_i - 4) + 4] \quad (i = 1, 2, \dots, 10) \quad (2)$$

と再定義した。

このような程度語補正係数を導入したのは、(i) 程度語定数表を作成するための実験では、29 段階の評価が可能であったが、

楽曲データ（から抽出される物理的特徴量）と印象尺度とを対応付けするための実験では、7 段階評価であったため、同じ程度語（「とても」や「少し」）でも意味の粒度が異なる、(ii) 楽曲データと印象尺度との対応付けは、楽曲を聴取しながらの評価であり、楽曲の印象と言葉（程度語+印象語）が直接的に対応付けられたのに対し、程度語定数表の作成では、言葉の意味を印象尺度で意味づけており、程度語どうしの相対的な意味関係が求められたにすぎない、といったことから、程度語どうしの

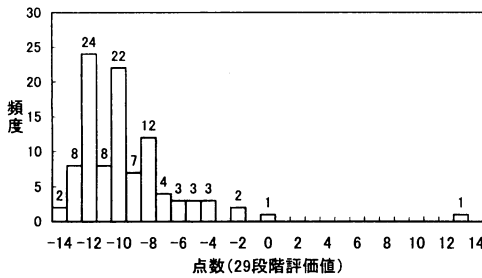


図8 「よほど（静か）でない」の評価結果

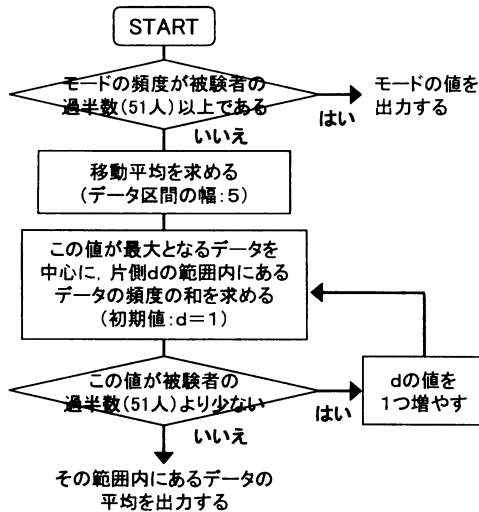


図9 ヒストグラムの代表値を算出するためのアルゴリズム

表6 ユーザが入力した印象語と程度語の組み合わせ

結構心地よくてかすかに寂しい曲, 割とかっこいい曲, 実にほっとする曲, えらく主張のある曲, 少々コミカルな曲, うんと通き通った曲, かなり暗い曲, 極めてうるさい曲, とても単純な曲
--

相対的な意味関係を保ちつつ、程度語定数によって示される程度語の意味的役割と楽曲印象との対応付けの妥当性を調べるためである。

さて、実験の手順であるが、各被験者は、まず(1)入力可能な印象語(164語)と程度語(119語)のリストを見ながら、システムへの入力となる文を作った。次に(2)この文をシステムa ($b = 0.60$)、システムb ($b = 0.88$)、システムc ($b = 1.00$)のそれぞれに入力し、楽曲検索を行った。そして(3)それぞれの第1位候補楽曲を聴取し、入力した印象との適合度(5点満点)を5段階で評価した。但し、同じ楽曲が第1位候補として提示された場合は、最初の1回だけ楽曲聴取・適合度評価を行い、2回目以降は楽曲を聴取せずに、同じ適合度を与えることとした。以上の手順を約10回繰り返した。

表6にユーザが作った文の例を示し、表7に各システムにおける適合度の平均値と標準偏差を示す。

表7は、いずれのシステムもほぼ同等の性能(有意水準5%で適合度に統計的な有意差はなかった)であったことを示しているが、検索対象となった楽曲が160曲と少なかったため、多少印象ベクトルが異なっても、同じ楽曲が検索されることが多少

表7 各システムの性能評価の結果

システム	平均値	標準偏差	データ数
a (0.60)	3.76	1.24	394
b (0.88)	3.87	1.20	393
c (1.00)	3.80	1.26	393

観測された。そこで、同じ検索結果(第1位のみ)が提示されたデータをノイズとして除去した上で、2つのシステムどうし(aとb, aとc, bとc)を比較してみた。しかしながら、いずれの組み合わせにおいても統計的な有意差はなく(有意水準5%)、同程度の検索性能であった。

もともと、同じ楽曲を聴いても受ける印象が異なる、受けた印象が同じでも言葉による表現の仕方が異なる、といった個人差が存在するわけだが、程度語によって表現される意味の違いがこのような個人差よりも小さかったため、検索性能に差が生じなかったものと考えられる。

5. 程度語の類型化

前章で述べたように、程度語によって表現される意味の違いは、ユーザどうしの個人差に比べ、小さいものと思われる。このような状況下で、それぞれの程度語の微妙な違いを詳細に記述しても、感性検索というタスクでは役に立たない(異なる検索結果は得られない)。そこで、前章で得られた、それぞれの程度語に対する評価データ(被験者100人分の評価値)を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ウォード法」を用いて、クラスタリングすることにより、程度語の類型化を試みる。このとき、程度語 W_A の評価データ E_A と程度語 W_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。

$$D(E_A, E_B)^2 = \sum_{s=1}^{100} (x_s(E_A) - x_s(E_B))^2$$

但し、程度語 W_A に対する被験者 s の点数(29段階評価値)を $x_s(E_A)$ とする。

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 を形成する程度語 W_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 $d(E_j, G_k)$ を求めた。図10に距離 $d(E_j, G_k)$ の最大値と平均値を示す。但し、距離 $d(E_j, G_k)$ は、クラスタ C_k の重心 G_k における第 s 要素の値を $x_s(G_k)$ と記述することにより、

$$d(E_j, G_k) = \sqrt{\sum_{s=1}^{100} (x_s(E_j) - x_s(G_k))^2} / 100$$

と定義される。

図10を見てみると、クラスタ数1~7のとき、重心からの距離の最大値は急激に小さくなっており、その後は比較的緩やかに減少しているように見える。したがって、クラスタ数7のときに特徴的な程度語グループが出揃ったものと考えられる。そこで、クラスタ数7の場合の各クラスタの構成とそれぞれのクラスタ内における重心からの距離の最大値を表8に示す。また、各クラスタから程度語1語を選び、その評価データのヒストグラムを図2~図8に示す。

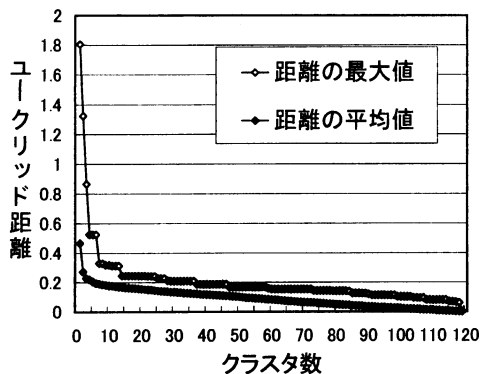


図 10 程度語のクラスタリング

表 8 クラスタ数 7 の場合の各クラスタの構成とクラスタ内における重心からの距離の最大値

クラスタを構成する程度語	最大値
あまりに~/うんと~/えらく~/この上なく~/すごく~/とつても~/とてつもなく~/とても~/とんでもなく~/どえらく~/どはずれて~/はなはだしく~/ひどく~/べら棒に~/むちゃくちゃ~/めちゃくちゃ~/めちゃめちゃ~/めっちゃ~/ものすごく~/ウルトラ~/スーパー~/一番~/圧倒的に~/大変~/底抜けに~/最も~/最上に~/最高に~/格段に~/桁外れて~/桁違いに~/極めて~/極度に~/極端に~/段違いに~/猛烈に~/百パーセント~/至極~/著しく~/超~/過度に~/非常に~/飛び切り~/馬鹿に~/~すぎる	0.25
いやに~/かなり~/すこぶる~/ずいぶん~/だいたい~/はなはだ~/ひとときわ~/ひとしお~/まったく~/めっぽう~/やけに~/一等等~/全然~/十二分に~/十分~/大いに~/大層~/存分に~/実に~/断然~/本当に~/段ちに~/相当~/真に~/高度に~/最低に~/	0.33
いささか~/お手頃に~/そこそこ~/なるだけ~/なるべく~/ほどほどに~/まあまあ~/まあ~/まぎます~/わかりかし~/並に~/中々~/中くらいに~/割と~/割に~/割り方~/割合~/平々凡々に~/平凡に~/普通に~/月並みに~/比較的~/程合いに~/結構~/過度に~/適当に~/頃合に~/	0
いささか~/お手頃に~/そこそこ~/なるだけ~/なるべく~/ほどほどに~/まあまあ~/まあ~/まぎます~/わかりかし~/並に~/中々~/中くらいに~/割と~/割に~/割り方~/割合~/平々凡々に~/平凡に~/普通に~/月並みに~/比較的~/程合いに~/結構~/過度に~/適当に~/頃合に~/	0.33
かすかに~/こころもち~/ちよつと~/ちよつぱり~/ほのかに~/やや~/わずかに~/中途半端に~/少々~/少し~/少なくとも~/幾らか~/幾分~/~め	0.23
あまり~/でない/あんまり~/でない/さして~/でない/さほど~/でない/それほど~/でない/大して~/でない	0.24
よほど~/でない	0

表 4 を見てみると、代表値が「静かな (8 点)」より大きいものが全程度語の約 60% を占めており、印象語の意味を強める働きをする程度語が豊富なことがわかる。次に、表 4 と表 8 を比べてみると、「静かな (8 点)」より大きい代表値を持つ程度語は、3 つのクラスタに分けられているが、「最低に~」が要素数 1 のクラスタを形成しているため、実質的には 2 つのクラスタに分けられたことになる。したがって、その意味的役割においては、大して差がないことがわかる。なお、「最低に~」は、図 4 に示したように、被験者が付与した点数が広く分散しており、意味の曖昧性によって異なるクラスタを形成したものと考えられる。

6. まとめ

本稿では、印象情報に基づくマルチメディアコンテンツ検索への応用を前提に、程度語の意味的役割を、修飾している印象語の意味を強める/弱めることと定義し、類語辞典 [15] から抽出された程度語 119 語の強弱関係を被験者実験により決定した。このとき、程度語によって修飾された印象語を、検索対象から抽出される物理的特徴量と対応付けられている印象尺度 (実際には 7 段階評価の印象尺度を 29 段階評価に拡張した「点数の基準表」を用いた) で評価することにより、程度語が印象語に与える影響を調べた。また、その過程において得られた実験データを用いて、程度語どうしの意味的類似性を調べ、類型化を行った。

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, Jun. 1996.
- [2] 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,” 情処学論, vol.43, no.2, pp.287-298, Feb. 1999.
- [3] 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, Apr. 1999.
- [4] 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,” 情処学論, vol.33, no.11, pp.1373-1383, Nov. 1992.
- [5] 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, Mar. 1997.
- [6] Shouji Harada, Yukihiko Itoh, and Hiromasa Nakatani, “Interactive image retrieval by natural language,” Optical Engineering, Vol. 36 No. 12, pp.3281-3287, Dec. 1997.
- [7] 原田将治, 伊東幸宏, 中谷広正, “感性語句を含む自然言語文による画像検索のための形状特徴空間の構築,” 情処学論, vol.40, no.5, pp.2356-2366, May 1999.
- [8] 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, Feb. 2001.
- [9] 大前寛子, 石橋直樹, 清木 康, 安西祐一郎, “複数の楽器を用いた楽曲を対象とするメタデータ自動生成方式とその実現,” 情処学データベースシステム研報, vol.2001-DBS-125, no.84, pp.145-152, Jul. 2001.
- [10] 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, vol.42, no.12, pp.3201-3212, Dec. 2001.
- [11] 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, Sept. 2002.
- [12] Tadahiko Kumamoto, “Design and implementation of natural language interface for impression-based music-retrieval systems,” Proc. International Conference on Knowledge-Based Intelligent Information and Engineering Systems, KES2004, LNAI3214, Springer, pp.139-147, Wellington, New Zealand, Sept. 2004.
- [13] L. A. Zadeh, “Outline of a new approach to the analysis of complex systems and decision processes,” IEEE Trans. on Systems, Man, and Cybernetics, vol.SMC3, no.1, pp.28-44, Jan. 1973.
- [14] 熊本忠彦, 太田公子, “印象に基づく検索のための印象語選定法の提案,” 情処学論, vol.44, no.7, pp.1808-1811, July 2003.
- [15] 大野晋, 浜西正人, “類語国語辞典,” 角川書店, 東京, 1985.