

# 関連語を用いた感性語の意味のモデル化

山崎 勢太<sup>†</sup>杉本 徹<sup>‡</sup>芝浦工業大学大学院 理工学研究科<sup>†</sup>芝浦工業大学 工学部<sup>‡</sup>

## 1. 研究の背景と目的

近年、単語・語彙の意味情報の蓄積は機械翻訳システムや検索システムの性能向上に大きく貢献している。しかし、このような単語の意味や概念に関する研究は名詞や動詞に関するものが多い。本研究では、人の情動や美的感情、印象などを表現する単語を感性語と呼び、感性語の意味のモデル化の方法を確立することを目的とする。感性語と因果関係を持つ語をその感性語の関連要素とし、それぞれの語の関連度を求めることでモデル化を行う。

## 2. 研究の概要

感性語と因果関係を持つ語をコーパスから抽出し、関連要素の候補とする。それらの単語の関連度を求め、関連度が高い単語を関連要素とする。本研究の手法で求めた関連要素に対して、それらが感性語と因果関係を持つ語として適当であるか、また、EC サイトのレビュー文中の関連要素から求めた商品の感性値が人間の感覚にあっていないかを評価する。今回は「かっこいい」、「かわいい」、「おしゃれ」の3単語を研究対象の感性語とする。

## 3. 関連要素

### 3.1 関連要素の条件

関連要素とは、感性語と因果関係をもつ単語の中で、感性語との関連が深い形容動詞語幹の名詞または形容詞の単語である。因果関係の有無の判断には助詞を用いる。「近未来的でかっこいい」や「渋くてかっこいい」のように、感性語とともに助詞の「て」または「で」が使われており、コーパス上で「て」または「で」を伴って一定の回数以上出現する場合に、その単語は感性語と因果関係をもつと判断する。それらの単語の関連度を算出し、関連度が上位の単語を関連要素とする。

## 3.2 関連要素の抽出

大規模日本語 n-gram データ[1]の 3-gram のデータを用いて、感性語と因果関係をもつ語を抽出する。3-gram なので、各項目は「クール/で/かっこいい」のような  $w_1/w_2/w_3$  の形をしている。 $w_3$ が感性語、 $w_2$ が「て」または「で」であり、その項の出現回数が100回以上の場合に  $w_1$  を抽出する。 $w_1$  を MeCab[2]で解析し、品詞が形容詞または形容動詞語幹の名詞である場合に  $w_1$  を関連要素の候補とした。「かっこいい」、「おしゃれ」から抽出された関連要素の候補はそれぞれ約250語、「かわいい」から抽出された候補は約500語であった。これらの関連要素の候補について、それぞれの感性語との関連度を算出し、上位20語を関連要素とした。

## 4. 関連度

関連度の尺度として WeedsPrec[3]を改良した尺度である balPrec[3]を用いる。

### 4.1 WeedsPrec

WeedsPrec とは、単語  $u$  が単語  $v$  の下位概念である度合いを表す尺度である。 $v$  が  $u$  を完全に包含している場合に最大値の1.0となる。計算式を以下に示す。

$$\text{WeedsPrec}(u \leq v) = \frac{\sum_{w \in U \cap V} f(u, w)}{\sum_{w \in U} f(u, w)}$$

本研究では  $u$  が関連要素、 $v$  が感性語である。 $U$  は大規模日本語 n-gram データの 3-gram において  $u$  から2単語以内に出現した単語の集合であり、 $V$  は同様に 3-gram において  $v$  から2単語以内に出現した単語の集合である。 $f(u, w)$  は共起のしやすさを表す尺度であり、以下の式で定義される。 $PMI(u, w)$  は相互情報量を表す。

$$f(u, w) = \begin{cases} PMI(u, w) & PMI(u, w) \geq 3 \\ 0 & PMI(u, w) < 3 \end{cases}$$

$$PMI(u, w) = \log \frac{P(u, w)}{P(u)P(w)}$$

$P(u, w)$  は 3-gram において、 $u$  と  $w$  が共起する確率であり、 $P(u)$  と  $P(w)$  は  $u$ 、 $w$  がそれぞれ 1-gram に出現する確率である。

Semantic Modeling of Sensibility Words Based on Related Words  
<sup>†</sup>Seita Yamazaki Graduate School of Engineering and Science, Shibaura Institute of Technology

<sup>‡</sup>Toru Sugimoto College of Engineering, Shibaura Institute of Technology

#### 4.2 balPrec

balPrec は、LIN ( $u, v$ ) [3] と WeedsPrec ( $u \leq v$ ) の幾何平均をとることで、単語  $u$  の出現回数が少ない場合に精度が低くなる WeedsPrec の問題に対処している。WeedsPrec は方向性をもった尺度であるのに対し、LIN は対称的な尺度である。

$$\text{balPrec}(u \leq v) = \sqrt{\text{LIN}(u, v) \cdot \text{WeedsPrec}(u \leq v)}$$

$$\text{LIN}(u, v) = \frac{\sum_{w \in U \cap V} [f(u, w) + f(v, w)]}{\sum_{w \in U} f(u, w) + \sum_{w \in V} f(v, w)}$$

関連度の尺度として balPrec を用いることにより得られた関連要素の上位 5 単語を表 1 に示す。

表 1: 感性語と関連要素

かっこいい	かわいい	おしゃれ
ボーイッシュ	おしゃま	小奇麗
スポーティ	ボーイッシュ	カジュアル
ごつい	まん丸	スポーティ
男前	丸っこい	シック
精悍	チャーミング	エレガント

#### 5. 評価実験

本研究の手法で求めた関連要素に対して、それらが感性語と因果関係を持つ語として適当であるか、また、EC サイトのレビュー文中の関連要素から求めた商品の感性値が人間の感覚にあっていないかをアンケートにより評価する。

##### 5.1 感性語と関連要素の関係に関する実験

###### 5.1.1 実験方法

本研究で求めた「かっこいい」、「かわいい」、「おしゃれ」の関連要素それぞれ 20 語について、①「感性語ならば関連要素である」といえるか、②「関連要素ならば感性語である」といえるかどうかの 2 つの質問に 5 段階評価で回答してもらった。②の回答の平均値が①の平均値より大きく、②の平均値が 3 より大きい場合に、感性語と関連要素は因果関係をもつと判断する。また、このアンケートで求めた理想的な関連要素に関して、関連度の値の大きさによる評価も行った。なお被験者は本学の学生 10 名である。

###### 5.1.2 実験結果

感性語と因果関係をもつと判断された関連要素の割合は、「かっこいい」においては 80%、「かわいい」においては 100%、「おしゃれ」においては 60%であった。また、アンケートにおける①の回答の平均値が大きかった関連要素の上位 10 語の関連度の平均と全 20 語の関連度の平均を比較したところ、「かっこいい」、「かわいい」、「おしゃれ」の 3 単語とも上位 10 語の平均の方が値が大きかった。

#### 5.2 関連要素から得られる感性値を用いた実験

EC サイトは Amazon を利用する。実験の対象の商品のジャンルは「スマホケース」、「腕時計」の 2 種類である。それぞれのジャンルから、さらに 10 種類の商品を選び実験を行った。被験者は本学の学生 10 名である。

##### 5.2.1 実験方法

商品が「かっこいい」かどうか、「かわいい」かどうか、「おしゃれ」かどうか、それぞれ 5 段階評価で回答してもらい、その結果を基に、本研究で求めた商品の感性値とレビュー文中の感性語の出現率の値を比較した。

##### 5.2.2 感性値と出現率

感性語ごとに、レビュー文中の関連要素の関連度を加算し、それをレビュー文中の名詞、形容詞の総出現回数で割ったものを感性値とする。また、感性値の計算において、関連要素を感性語のみとした場合をその感性語の出現率とする。ただし、関連要素が形容詞または形容動詞だった場合に、「かわいくない」や「クールじゃない」のように、関連要素の後に助動詞「ない」が接続している場合は関連度を 0 とする。

##### 5.2.3 実験結果

アンケート結果と感性値、出現率との相関係数を求めた。「腕時計」の結果を表 2 に示す。

表 2: 腕時計におけるアンケートとの相関係数

感性語	アンケート結果と感性値との相関係数	アンケート結果と出現率との相関係数
かっこいい	0.10	0.44
かわいい	0.76	0.68
おしゃれ	0.70	0.13

#### 6. まとめと今後の展望

本研究では、感性語の関連要素を調べ、感性語との関連度を求めることで感性語の意味のモデル化を行った。感性語との因果関係をもった関連要素を高い確率で求める事ができた。レビュー文を使った実験では、関連要素を用いず、感性語のみで感性値を求めたほうが良い結果になる場合があった。商品の種類によって関連要素があまり出現せず、適切な値が求められなかったと考えられる。今後の課題として、感性語により評価される物の種類によって異なる関連要素を求める方法を検討する必要がある。

##### 参考文献

- [1] 工藤拓, 賀沢秀人: Web 日本語 N グラム第 1 版, 言語資源協会, 2007
- [2] MeCab: <http://mecab.sourceforge.net/>
- [3] Lili Koterman et al., "Directional distributional similarity for lexical inference", Natural Language Engineering, vol.16, No.4, 2010