

感情表現による書評情報の可視化手法の提案

Visualization of the book review by emotional words

松浦有容[†] 渥美幸雄[‡]
Arimasa Matsuura[†] Yukio Atsumi[‡]

[†] 専修大学大学院 経営学研究科
[‡] 専修大学 経営学部

[†] Graduate School of Business Administration, Senshu University

[‡] School of Business Administration, Senshu University

要旨

書評に含まれる感情表現を基にして、その書評が対象とする書籍の雰囲気可視化する手法を提案すると共に、提案手法を検証する為のシステムを実装した。書評という質的データを量的データとして扱うことで書籍の雰囲気を数値化し、その分布を散布図とレーダー・チャートとして示すことで可視化を行う。本研究では、書評の分析に感情の分類モデルを適用し、書籍の持つ雰囲気を4つに分類する。書評中の感情表現から、それぞれの雰囲気を示す語句を抽出し、その集計値より書籍が持つ雰囲気を推測する。散布図により、書評から推測される書籍の雰囲気の分布を示す。レーダー・チャートにより、雰囲気の推測の基となった書評中の感情表現の分布を示す。

1. はじめに

書評は、書籍についての感想や意見を述べたものであり、書籍の雰囲気を、読者に伝える。

本研究では、文芸書を対象とした書評について、それに含まれる感情表現から、書籍の雰囲気可視化する手法を提案する。書評という質的データの分析を、それに含まれる語句の頻度という計測可能な量的データとして扱うことで実現する。購読する文芸書を選ぶ時に利用することを目指す。

2. 先行研究について

製品やサービスについてのレビューから意見や評価を抽出することは、マーケティング分野での応用が期待され、主として、自然言語研究の立場からの研究例がある。

乾らは、自然言語研究の立場から、レビュー等の評価情報からの評判・感性抽出に関する研究を報告している[1]。

書評もレビューの一つと考えられる。自然言語研究の応用例として、書評中の語句より、書籍推薦のキーワードを抽出することを試みた研究として、以下がある。

原田らは、主として児童書を対象にして、書籍から感じる感情や雰囲気等の感性パラメータより書籍を検索するシステムの研究を行っている[2]。

佐々木らは、ミステリーを対象にして、推薦に必要なカテゴリーを分析し、作品を表す特徴の自動抽出を行う枠組みを提案している[3]。

本研究では、書評中の感情表現に、感情の分類モデルを適用することで、書籍の雰囲気可視化することを試みた。

3. 感情表現と書籍の雰囲気

3.1. 感情の分類

感情の分類は、主に心理学の分野で研究されている[4]。本研究では、可視化の観点から、ラッセルの円環モデルを参考にした感情の分類モデルを書評の分析に適用する。

ラッセルは、“全ての感情は、「快-不快」と「覚醒-眠気」を軸とする平面に円環上に並んでいる”、とする円環モデルを提唱した(図1)。

一方、中村芳樹は、環境の快適性の判断にラッセルの円環モデルを適用した[5]。各象限に、「おもしろい」、「いらいら」、「リラックス」、「つまらない」という環境状態を当てはめた。

本研究では、書籍の雰囲気への定義に、中村芳樹による環境状態の分類を引用する。

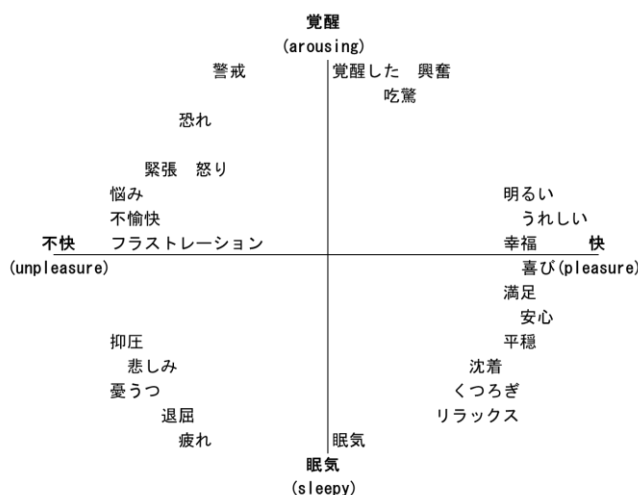


図1 ラッセルの円環モデル

3.2. 感情の表現

中村明は、近現代の日本の作家 197 人の作品 806 編から、感情表現の用例を収録した『感情表現辞典』を編集した[6]。日本語表現の観点から、2093 語の感情表現を 10 の区分に分類している。

(喜、怒、哀、怖、恥、好、厭、昂、安、驚)

本研究では、『感情表現辞典』を、書籍の雰囲気への推測する為の辞書として使用する。

3.3. 感情表現と書籍の雰囲気への対応

ラッセルの円環モデルで取り上げられている感情表現の多くは『感情表現辞典』にも収録されており、お互いの感情の分類には、親和性がある。本研究では、『感情表現辞典』に収録されている感情表現に対して、ラッセルの円環モデルを適用する。

書籍の雰囲気、「円環モデル」での区分、『感情表現辞典』での区分の対応を表1に示す。

表1 感情表現と書籍の雰囲気への対応

書籍の雰囲気への区分	「円環モデル」の区分	『感情表現辞典』の区分
おもしろい	快	覚醒 喜、昂、驚
いらいら	不快	覚醒 怖、怒、厭
リラックス	快	眠気 安
つまらない	不快	眠気 哀

4. 感情表現による書籍の雰囲気への可視化の方法

4.1. 書籍の雰囲気への推測

4.1.1. 書籍の雰囲気への推測の方法

本研究では、書評中の感情表現の出現頻度から求めた点をラッセルの円環モデル上にプロットし、その点が属する象限により、書籍の雰囲気への推測する。

個々の点は、一つの書評の意見を推測したものだが、複数の書評から求めた点を同時に示すことで、書籍が持つ雰囲気への全体的な傾向が推測できる。

以下に、書籍の雰囲気を推測する手順を示す。

- 1) 書評を品詞単位に分解する。
- 2) 分解した語句より、書籍の雰囲気の区分毎に感情表現を集計する。
- 3) 感情表現の集計値から、4.1.2項に示す方法で平面上の点を求める。
- 4) 求めた点の位置が、書籍の雰囲気を示す。

4.1.2. 平面上の点の求め方

感情表現の集計値から、以下の手順により、 $-1 \leq (x, y) \leq +1$ を範囲とする平面上の点を求める。

- 1) 書評の長さの影響をなくす為、各感情表現の集計値を抽出した感情表現の総数で割り、正規化する。
- 2) 正規化した集計値を各象限に割り当てたデータ行列を考える。(快/覚醒が+、不快/眠気が-)。

$$X_{i1}=(x_{i1}, x_{i1}), X_{i2}=(-x_{i2}, x_{i2}), X_{i3}=(x_{i3}, -x_{i3}), X_{i4}=(-x_{i4}, -x_{i4}) \quad 1 \leq i \leq n(\text{書評数}) \dots\dots\dots (式1)$$

- 3) データ行列の総和を求める。

$$Y_{ij}=X_{i1}+X_{i2}+X_{i3}+X_{i4}=(x_{i1}-x_{i2}+x_{i3}-x_{i4}, x_{i1}+x_{i2}-x_{i3}-x_{i4}) \quad 1 \leq j \leq 2(\text{次元数}) \dots\dots\dots (式2)$$

- 4) 式2より求めた行列の各行をXY座標の組とする。
- 5) 求めたXY座標の示す点が、書籍の雰囲気を示す。

4.1.3. 書籍の雰囲気の判定

書籍の雰囲気の判定基準を表2に示す。

表2 書籍の雰囲気の判定基準

	「快-不快」	不快	快
「覚醒-眠気」		-1.0 < 0.0	0.0 ≤ +1.0
覚醒	0.0 ≤ +1.0	いらいら	おもしろい
眠気	-1.0 < 0.0	つまらない	リラックス

4.2. 書籍の雰囲気の可視化

上記の手順で求めた数値から、散布図とレーダー・チャートを作成することで書籍の雰囲気を可視化する。散布図は、書評から推測した書籍の雰囲気を平面上にプロットしたものである。その散らばり具合から、書籍の雰囲気の大まかな傾向を示す。レーダー・チャートは、書評から抽出した感情表現の数を区分毎に線で結んだものである。雰囲気の推測の基となった書評中の感情表現の詳細な分布を示す。

5. 提案手法の実装

5.1. 提案手法の検証システム

提案手法を検証するシステムの実装要件を以下に示す。

- ・ 書評中の感情表現の抽出。
- ・ 書評中の感情表現による書籍の雰囲気の数値化。
- ・ 書籍の雰囲気を示す散布図とレーダー・チャートの作画。

5.2. 感情表現の抽出の実装

日本語には活用のある語があり、単純な文字列比較では、感情表現は抽出できない。又、一つの感情表現が複数の語句から成り立つ場合もある。例えば、「面白くない」という感情表現は、「面白く」と「ない」の2語に分解され、「面白く」は「面白い」の活用形である。しかし、「面白い」は、1語の基本形からなる感情表現でもある。この様に、感情表現の抽出には、品詞の活用と、その感情表現が何語から構成されているかを考慮する必要がある。本研究では、品詞解析を行い、活用の基本形を感情表現の構成数だけ比較する。品詞解析には茶筌を利用する[7]。『感情表現辞典』を基に、感情分類区分、感情表現語、活用の基本形、感情表現の構成数、からなる検証システム用の感情表現辞書を作成する。

5.3. 書籍の雰囲気の数値化の実装

5.3.1. 感情表現の集計

茶釜を利用して書評を品詞解析し、分解された語句を感情表現辞書と比較し、一致した語句を書籍の雰囲気の区分毎に集計する。

5.3.2. 書評の数値化

感情表現の集計値から、書評が示す書籍の雰囲気を数値化し、平面上の点を求める。

5.4. 散布図とレーダー・チャートの作画の実装

書籍の雰囲気を示す散布図(図2)とレーダー・チャート(図3)を作画する。実装はPerlで行い、作画用のライブラリにはGDを利用する。

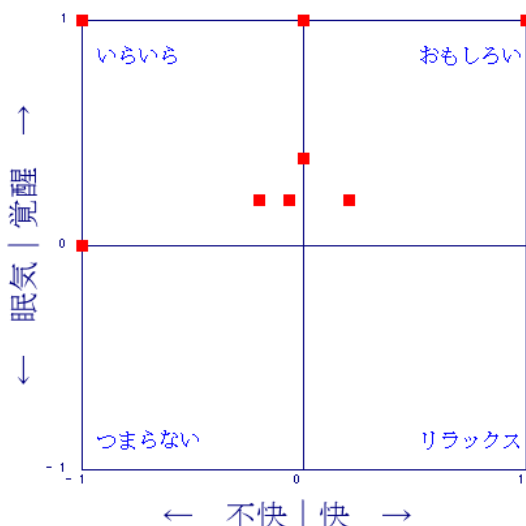


図2 散布図の例

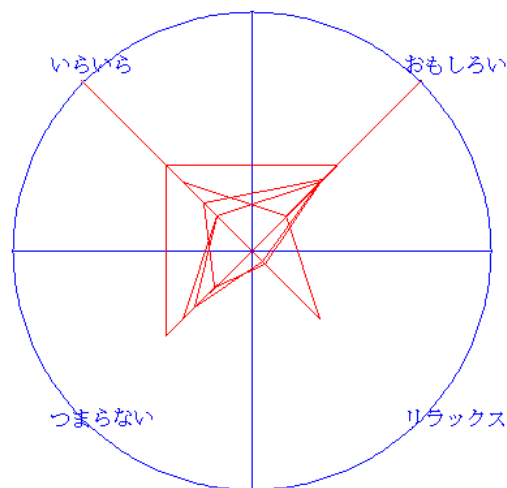


図3 レーダー・チャートの例

6. おわりに

本研究では、書評中の感情表現から、書籍の雰囲気を可視化する手法を提案し、検証用のシステムを実装した。今後は、本システムを利用して書評の分析を行い、手法の有効性を確認する予定である。

7. 参考文献

- (1) 乾孝司、奥村学、「テキストを対象とした評価情報の分析に関する研究動向」、自然言語処理、Vol. 13 No. 3、2006
- (2) 原田隆史、「感性パラメータを用いた類似する小説の提示」、情報知識学会誌、Vol. 21 No. 2、2011
- (3) 佐々木若菜、関洋平、青野雅樹、「Web 書評を対象としたカテゴリー分析と読み手が受けた印象や感情の自動抽出」、言語処理学会 第13回年次大会論文集、2007
- (4) 濱治世、鈴木直人、濱保久、『感情心理学への招待』、サイエンス社、2001
- (5) 中村芳樹、「快適な居住空間の設計法」、
<http://www.enveng.titech.ac.jp/sogokamoku/sogoVol12.pdf>、2011/11/26 現在
- (6) 中村明、『感情表現辞典』、東京堂出版、1993
- (7) chasen legacy -- an old morphological analyzer、<http://chasen-legacy.sourceforge.jp/>、2011/11/26 現在