

# 感情表現による書評情報の可視化手法の提案と実装

## 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

Graduate School of Business Administration, Senshu University

### 概要

書評に含まれる感情表現から、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。書評という質的データを量的データとして扱うことにより書籍の雰囲気可視化し、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。書評の分析には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、書評中の感情表現の分布を示す。提案手法を検証するためのシステムを実装し、文芸書を対象とした書評を分析し、提案手法の娯楽小説に対する適合度を確認した。

### 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。書評は新聞や雑誌に掲載され、読者が購読する書籍を選ぶ際の判断基準となる。書評が対象とする書籍は多岐に及ぶが、対象書籍の内容を一般読者に紹介するといった性格上、執筆者は、学者や作家のような専門家である場合が多い。書評の読者は、専門家の意見を参考にしつつ、自分にとっての有効性や満足度を考えて、書籍の購読を検討する。

書評は新聞や雑誌といったマスメディアにあっては伝統的なコンテンツであるが、供給側と需要側の関係は単方向だったといえる。書評を需要する一般の読者が自分の意見を述べる機会は、書籍折り込みの読者カードに感想を書いて返信するくらいしかなかった。ところが、書評を巡る状況は、ネットワーク化の進展により変化した。一

般の読者がインターネットという場に自分の書評を発表することができるようになったからである。

個人がネットワークに参加する機会は、インターネットが普及する以前ではパソコン通信が主流であった。大手プロバイダ内には文学系のコミュニティがあって、参加者が自分の読んだ書籍の書評を投稿していた。参加者が限定されたコミュニティ内ではあるが、そこでは書籍についての議論が行われていた。パソコン通信という基盤は、インターネットに置き変わったが、書評は有力なコンテンツであり、書評を掲載するサイトが多数ある。

本稿では、文芸書を対象としたインターネット上の書評について、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。書評は質的データといえるが、そこ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、書

籍の雰囲気の推測可視化を実現する。購読する文芸書を選ぶときに判断材料を与えることを目指す。

## 2. 研究の着眼点

### 2.1 インターネット上の書評検索の問題点

近年、何かを調べるときには、インターネットの検索エンジンを利用することが多い。書評も検索エンジンを利用して検索することができる。表 2.1 に Google と Yahoo! を利用したインターネット上の書評の検索例を示す。膨大な数のサイトが検索されることが分かる。Google などの商用検索サービスでは、スポンサー企業のサイトが上位にランクされる。したがって、検索キーワードにタイトルや著者名を与えると検索結果の上位には通販サイトがランクされる。個人サイト上の書評は、一般に検索結果の下位に埋もれてしまうことが多い。検索キーワードを工夫(「書評 OR 感想」を追加)することでそれらの問題はある程度は改善できるが、検索されるサイトの数は依然として膨大である。内容の判断は目

視に頼らざるを得ないため多くの書評が見落とされる。

### 2.2 書評の可視化

書評は、かつては、マスメディアによって供給される専門家による示唆に富んだものであったが、インターネット上の書評は一般読者が発信するもので、玉石混合といわざるを得ない。他方、インターネット上の書評は、一般的な製品でいえばユーザーによるレビューと考えることができる。そこから読者の意見を抽出することは可能と思われる。

従来型の新聞や雑誌の書評は専門家によるもので、高品質だが量は限定的である。それに対して、インターネット上の書評はだれでも投稿できるので、量が多いが質の面で問題がある。質と量の問題はインターネット上の多くのコンテンツで指摘できるが、書評も同じ問題を抱えている。書評の質は投稿者に依存する問題であるので改善できないが、量の問題はインターネット上の書評を整理して閲覧の指針を与えること

表 2.1 インターネット上の書評の検索例(2011/12/28 現在)

検索キーワード	検案件数	
	Google	Yahoo!
阿部和重 グランド・フィナーレ	47700	47700
阿部和重 グランド・フィナーレ 書評 OR 感想	13200	13100
角田光代 対岸の彼女	180000	180000
角田光代 対岸の彼女 書評 OR 感想	32300	32300
古川日出男 LOVE	108000	108000
古川日出男 LOVE 書評 OR 感想	75900	75900
吉田修一 パレード	162000	257000
吉田修一 パレード 書評 OR 感想	53600	53600
青木淳悟 四十日と四十夜のメルヘン	47600	47600
青木淳悟 四十日と四十夜のメルヘン 書評 OR 感想	13500	13600

で利用者の利便を改善できると思われる。Amazon.co.jpなどで書籍の評価を星の数の大小で示すのも指針の例である。「書評コミュニティ一本が好き！」[01]では書籍の特徴を「ビジネス-ホビー・実用」軸と「理科系-文科系」軸で評価する仕掛けを提供している(図2.1)。書評の投稿者は自分が感じた書籍の特徴を平面上にプロットすることができる。書評の投稿者による評価が画面の上で一望できるという利点がある。

インターネット上の書評には潜在的な需要があつて、十分に供給されてもいるが、質と量の問題から必ずしも有効利用されていない。Amazon.co.jpなどでの星の数や「書評コミュニティ一本が好き！」での2軸による評価などの指針も与えられているが、書評そのものから導き出された指針ではないので書評の質を担保するまでには至らず、検索の利便を与えるものではない。しかし、原田らの報告[02]によれば、イン

ターネット上の書評は従来型の新聞や雑誌の書評に比べて多彩な内容を含んでいる。同一書籍についてのインターネット上の書評を集めてそこから何らかの評価を抽出できれば、それは、読者の意見の総体と考えることができる。

そこで、インターネット上の書評を整理し、検索の利便の向上を実現するための方法として書評の可視化を考えた。意見の総体を図に示せれば、一部の信用性に欠けるかも知れない書評を気にせずに全体の意見が把握可能となる。全体と細部が同時に概観できる可視化の方法を検討する。可視化により検索の利便が向上し、インターネット上の書評を閲覧する煩雑さからも解放されることが期待できる。

本稿では、文芸書を対象としたインターネット上の書評について、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。文芸書は専門書やビジネ

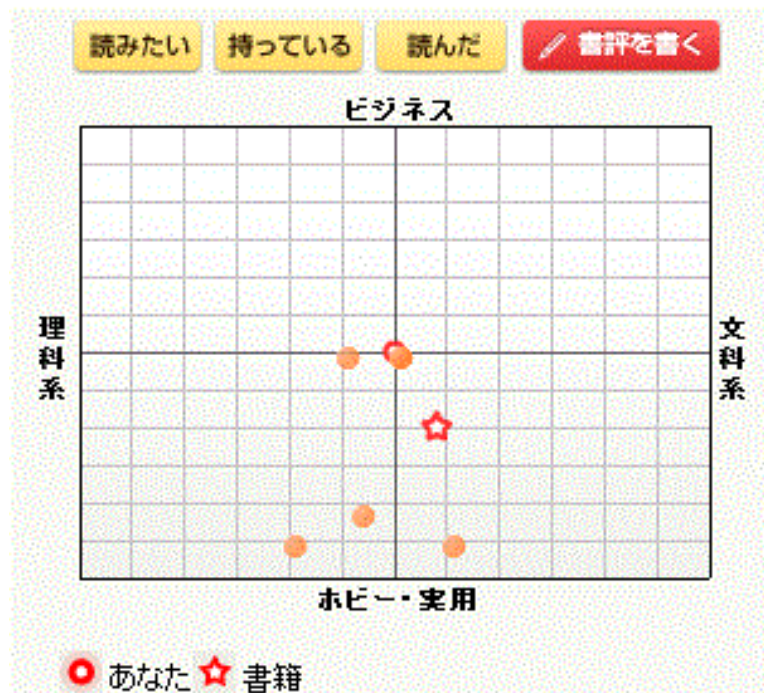


図2.1 「書評コミュニティ一本が好き！」での書籍の評価

ス書とは異なり、事実としての有効性よりも、喜怒哀楽といった読者の感情に訴える要素が高い。そのような書籍の書評では事実を客観的に述べるのではなくて、書評者が感じたことが語られる。したがって、書評中の感情表現から書籍の雰囲気や推測することができる考えた。書評そのものから書籍の雰囲気や推測するので、共通の土台で書評を比較することになる。読者が購読する文芸書を選ぶ時に判断材料を与えるという観点から、意義があると考え。

### 2.3 提案内容の構成

人間の感情には様々な状態があり、その表現も多彩である。書評は、書評者が作品から受けた印象や作品の持つ雰囲気を伝える。したがって、書評中の感情表現を分析することで書籍の雰囲気が推測できると考えられる。まず、感情の分類モデルを適用して感情状態を分類する。次に、分類され

た感情状態を示す感情表現を定義する。最後に、書評からそれらの感情表現を抽出する。その書評が、どのような感情状態を示す感情表現をどの程度含んでいるかで書籍の雰囲気を推測する。書評は質的データだが、語句の頻度という量的データとして扱うことで書籍の雰囲気や推測と可視化を実現する。本稿では、感情の分類、表現、抽出の関連性に着目して、書評から書籍の雰囲気を推測し、可視化する。本稿の提案について図2.2に示す。

提案手法の考え方を小川洋子著『博士の愛した数式』の書評を例に説明する。

「内容はとてもきれいに書かれて、小説とは相反的な数学の世界を垣間見せつつも、それが主張しすぎず、算数、数学に抵抗があるひとでも読みやすい小説だと思います。特に数学を出してくるだけでも、意外性があるのに、さらに博士は記憶のとどまる時間のみの中で生活をしていくという博士に課されたハンデ

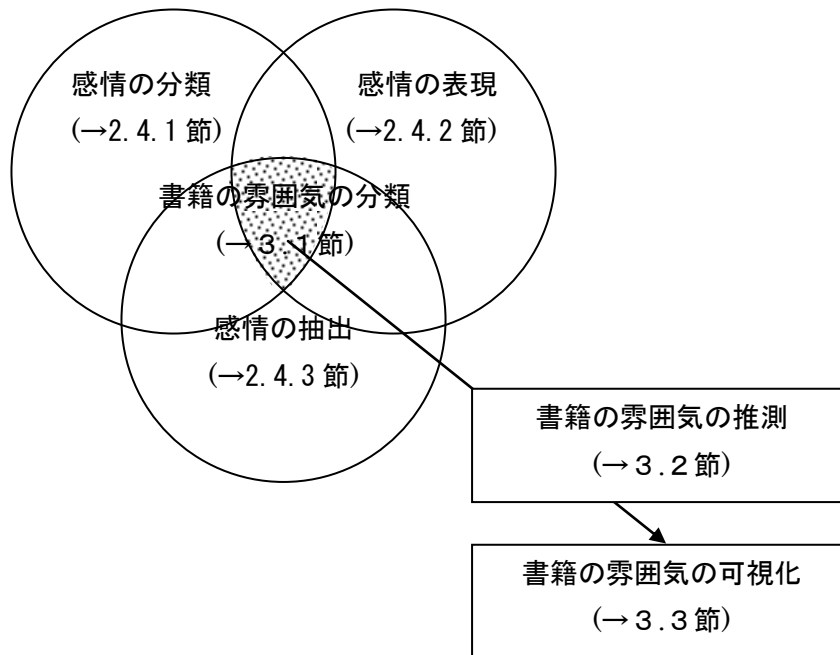


図2.2 提案内容の構成

というものが意外性を引き立て、その2つのテーマが象徴的です。

親子2人が注ぐ愛情は物語を進める中でどんどん湧き上がってくるのを感じることができし、博士は博士なりの愛情というものを紡ぎ出してくれます。

作品としては派手さはないですが、やさしさのある良作です。分量も小説の中でも多い方ではなく、読んでいくにはスムーズに流れていく感じでした。やはり、日常生活に関わりのない数学の世界が私達が知らないだけで実はすごく身近に存在していることの実感と、その数式の美しさ、3人の愛情の満たされた生活、それぞれが心をほっとさせるような作品でした。」

書評から下線部で示される感情表現を抽出し、それを感情の区分ごとに集計する。その集計数から書籍の雰囲気や推測する。表2.2にこの書評から抽出された感情表現の集計数を示す。この例からは、書籍の雰囲気が「おもしろい」と推測される。

## 2.4 関連研究について

### 2.4.1 感情の分類

感情は主に心理学の分野で研究されているが、様々な学説がある。本稿では、書評中の感情表現を分類することで対象書籍の雰囲気や推測することを試みるが、そのためには、感情表現を適切に分類する必要がある。心理学では感情の分類について、大

きく分けると、基本感情説と次元説の2つの立場がある。基本感情説は、感情には各民族や各文化間に共通な基本感情がある、とする立場である。次元説は、感情が次元上の1つのベクトルとして表される、とする立場である。

本稿では、可視化の観点から、次元説による感情の分類方法を検討する。感情を示すベクトルを図として表すことで、可視化が行えると考えた。いくつかあるモデルより、ラッセルの「円環モデル」[03]を感情の分類モデルとして採用する。

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2.3にラッセルの円環モデルを示す。

### 2.4.2 感情の表現

感情については、主に心理学の分野で研究されているが、具体的な感情表現となると言語学や文学の範疇となる。心理学は感情についてのフレーム・ワークを与えるが、そこに含まれる個々の要素には言及しない。

日本語による感情表現を収集したものとして、中村明の『感情表現辞典』がある[04]。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集した。日本語表現の観点から、感情表現を「喜」、「怒」、「哀」、「怖」、「恥」、「好」、「厭」、「昂」、「安」、「驚」の10の区分に分類している。

表2.2 書評から抽出された感情表現の集計数

感情の区分	感情表現	集計数
おもしろい	意外(2回)、上がって、満たされて	4回
いらいら	未検出	0回
リラックス	ほっと	1回
つまらない	未検出	0回

中村明は、心理学における感情の分類としては、次元説を支持している訳ではないが、『感情表現辞典』には複数の区分に登録されている感情表現もあり、基本感情説に立つものでもない。あくまでも、日本語表現の観点から、具体的な感情表現を収集して分類した。日本語による感情表現を網羅的に収集した事例は少ない。本稿では『感情表現辞典』を書籍の雰囲気を探るための辞書として使用する。

### 2.4.3 感情の抽出

インターネットの普及により、様々なテキスト情報がウェブ空間上に蓄積されている。書評もそういったテキスト情報の一部である。それらのテキスト情報には、評価の記述、感情や印象、賛否の表明といった、個人の意見が含まれており、その重要性が認識されている[05]~[06]。特に、サービスや製品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ている。これには自然言語研究の立場からのアプローチがあり、日本語によるテキストの解析を行うツールとして茶筌[07]が広く用いられている。

書評を対象とした自然言語研究の応用例として、原田らは、主として児童書を対象とした書評に含まれる感情や雰囲気など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[08]~[13]。

## 3. 書籍の雰囲気を可視化する手法の提案

### 3.1 書籍の雰囲気の種類

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[14]。住空間の快適性に関連して、ラッセルの円環モデルの各象限を「おもしろい」と「いらいら」と「リラックス」と「つまらない」と命名した(図3.1)。本稿では、書籍の雰囲気の定義に中村芳樹による分類を援用する。

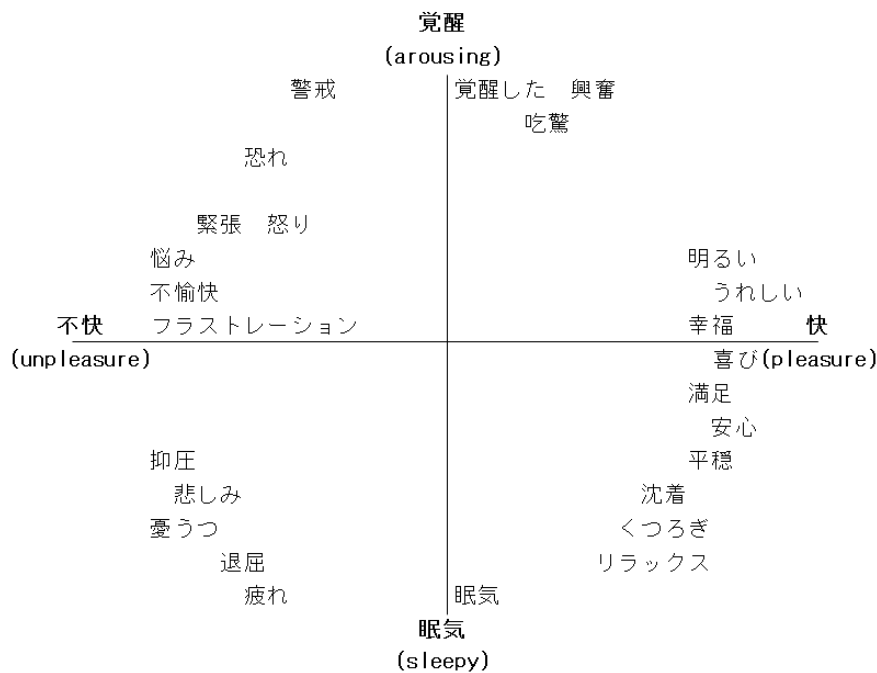


図2.3 ラッセルの円環モデル

ラッセルは、心理学の立場から感情の分類を行った。他方、中村明は、日本語表現の観点から具体的な感情表現を収集した。表3.1にラッセルが引用した語と『感情表現辞典』の分類の対応を示す。両者を組み合わせると、日本語のテキストに円環モデルを適用することが可能になる。本稿では、ラッセルの円環モデルと中村明の『感情表現辞典』を対応付けることを提案する。

表3.1に示す通り、「恥」及び「好」に分類される語は円環モデルには引用されていない。また、「喜」と「昂」と「厭」に分類される語が円環モデル上の複数象限にまたがっている。ラッセルの円環モデルと中村明の『感情表現辞典』が対応していない区分は以下のように処理した。

- ・「喜」は、「快-不快」:「快」 / 「覚醒-眠気」:「覚醒」に対応させる。
- ・「昂」は、「快-不快」:「快」 / 「覚醒-眠気」:「覚醒」に対応させる。

- ・「厭」は、「快-不快」:「不快」 / 「覚醒-眠気」:「覚醒」に対応させる。

・対応のない「恥」と「好」は除外する。  
以上より、感情表現と書籍の雰囲気については表3.2に示す対応付けを行い、4つの書籍の雰囲気を定義した。

### 3.2 書籍の雰囲気への推測

本稿では、書評中の感情表現の出現頻度から求めた点を  $-1 \leq (x, y) \leq +1$  を範囲とする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を推測する。

ラッセルの円環モデルでは、各感情は2次元のベクトルとして示される。4つの区分ごとに感情表現の数を集計することで感情ごとの4つのベクトルを求め、それらを合成することで書籍の雰囲気を平面上の点として示す。以下に、書評から書籍の雰囲気を推測する手順を示す。

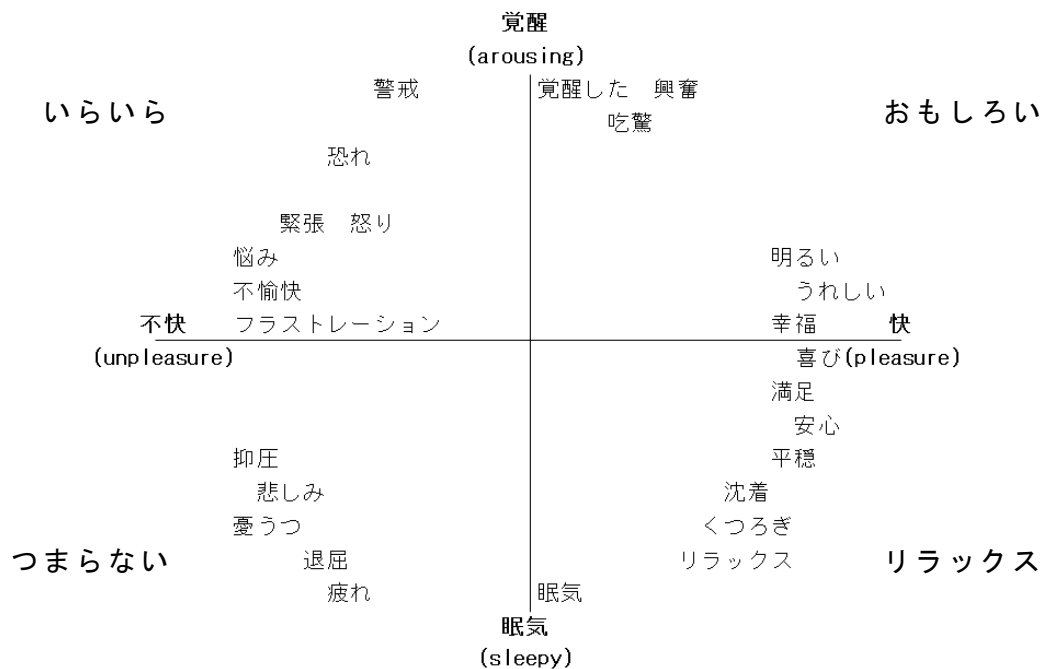


図3.1 感情の分類と書籍の雰囲気への対応

- 手順① 書評を品詞単位に分解する。  
 手順② 区分ごとに感情表現を集計する。  
 手順③ 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。  
 手順④ 正規化した集計値を各象限に割り当てた2次元ベクトルを考える。  
 $X_{i1}=(x_{i1}, x_{i1}), X_{i2}=(-x_{i2}, x_{i2}),$   
 $X_{i3}=(x_{i3}, -x_{i3}), X_{i4}=(-x_{i4}, -x_{i4})$   
 $1 \leq i \leq n$ (書評数).....(式1)
- 手順⑤ 式1の2次元ベクトルを合成する。  
 $Y_{ij}=X_{i1}+X_{i2}+X_{i3}+X_{i4}=(x_{i1}-x_{i2}+x_{i3}-x_{i4}, x_{i1}+x_{i2}-x_{i3}-x_{i4})$   
 $1 \leq j \leq 2$ (次元数).....(式2)
- 手順⑥ 式2より算出した2次元ベクトルを座標の組とする。  
 手順⑦ 求めた座標を、判定基準(表3.3)と比較して、書籍の雰囲気を推測する。

表3.1 ラッセルが引用した語と『感情表現辞典』の分類の対応

ラッセルが引用した語	「円環モデル」の区分		『感情表現辞典』の区分
	快-不快	覚醒-眠気	
幸福、うれしい、明るい	快	覚醒	喜
満足、喜び	快	眠気	
興奮	快	覚醒	昂
緊張	不快	覚醒	
吃驚	快	覚醒	驚
恐れ	不快	覚醒	怖
怒り	不快	覚醒	怒
悩み、不愉快	不快	覚醒	厭
憂うつ	不快	眠気	
哀しみ	不快	眠気	哀
くつろぎ、沈着、平穩、安心	快	眠気	安
該当する語はない。			恥、好
覚醒した	快	覚醒	該当する区分はない。
警戒、フラストレーション	不快	覚醒	
眠気、リラックス	快	眠気	
抑圧、退屈、疲れ	不快	眠気	

表3.2 感情の表現と書籍の雰囲気の対応

書籍の雰囲気の区分	「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



表 3.4 [集計値]は、ある書評について手順①～②を実行し、品詞単位に分解した語句から感情表現を抽出し、書籍の雰囲気の違いごとに感情表現を集計した値である。これを例に、書評から書籍の雰囲気を推測する手順を説明する。表 3.4 [正規化]は手順③、表 3.4 [X座標、Y座標]は手順④を実行し、求めた値である。

- i 集計値を、書評の長さを考慮して正規化する(表 3.4 [正規化])。→手順③
- ii 感情表現を、XY座標に示す(表 3.4 [X座標、Y座標])。→手順④
- iii 書籍の雰囲気を数値化する(表 3.4 [書籍の雰囲気の数値化])。→手順⑤～⑥
- iv 書籍の雰囲気を「いらいら」と推測する(図 3.2)。→手順⑦

表 3.3 書籍の雰囲気の判定基準

「快-不快」		不快	快
		$-1 \leq x < 0$	$0 \leq x \leq +1$
「覚醒-眠気」	$0 \leq y \leq +1$	いらいら	おもしろい
	$-1 \leq y < 0$	つまらない	リラックス

表 3.4 書籍の雰囲気の数値化の例

書籍の雰囲気の数値化の区分	集計値	正規化	X座標	Y座標
おもしろい	3	0.2	+0.2	+0.2
いらいら	6	0.4	-0.4	+0.4
リラックス	1	0.067	+0.067	-0.067
つまらない	5	0.333	-0.333	-0.333
書籍の雰囲気の数値化の計算(XY座標の総和)			-0.466	+0.2

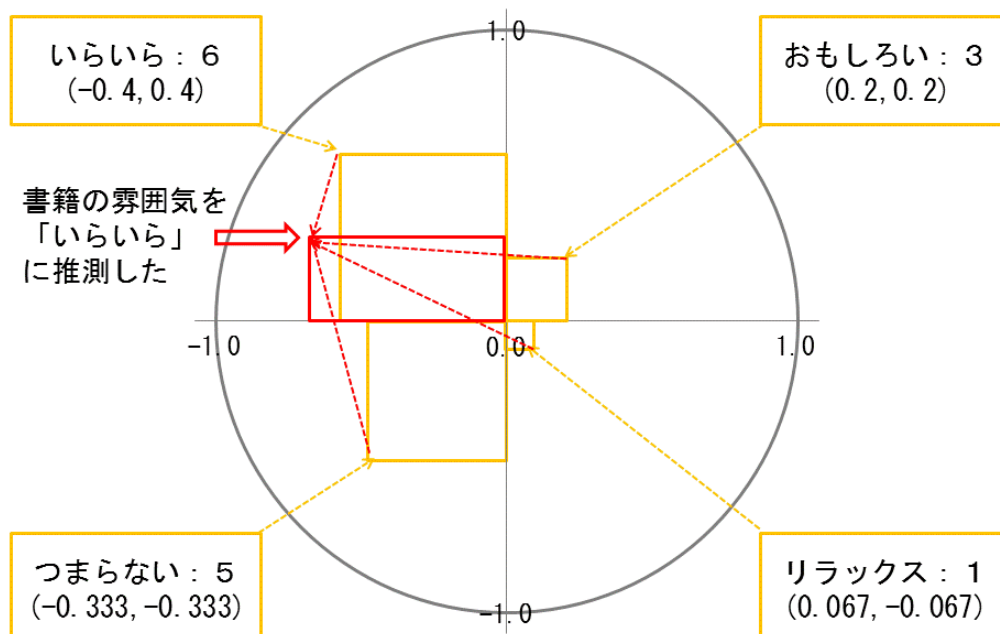


図 3.2 書籍の雰囲気の数値化の例

### 3.3 書籍の雰囲気可視化

#### 3.3.1 散布図による可視化

散布図による可視化では、書籍の雰囲気の全体的な傾向を示す。表3.5と図3.3に小川洋子著『博士の愛した数式』についての書評を分析した結果を散布図として表した例を示す。収集した書評から書籍の雰囲気を推測し、散布図上の1つの点として表す。個々の点は、1つの書評の意見を推測したものであるが、複数の書評から求めた

点を同時に示すことで、書籍が持つ雰囲気の全体的な傾向が推測できる。散布図による可視化は、ラッセルの「円環モデル」にならう。X軸(横軸)は「快-不快」次元を示す。Y軸(縦軸)は「覚醒-眠気」次元を示す。原点を(0,0)とする。XY軸の範囲は、 $-1 \leq (x, y) \leq +1$ とする。数値化された書籍の雰囲気の組をXY座標として、平面上にプロットする。

散布図は要素の関係を示した図である。

表3.5 小川洋子著『博士の愛した数式』の書評の分析結果

書評	「快-不快」次元	「覚醒-眠気」次元	書籍の雰囲気
書評①	-0.467	0.2	いろいろ
書評②	0	1	おもしろい
書評③	-1	0.333	いろいろ
書評④	-1	-1	つまらない
書評⑤	0	1	おもしろい
書評⑥	-0.111	0.556	いろいろ
書評⑦	1	0.6	おもしろい
書評⑧	0.5	0.5	おもしろい
書評⑨	1	1	おもしろい
書評全体	-0.143	0.429	いろいろ

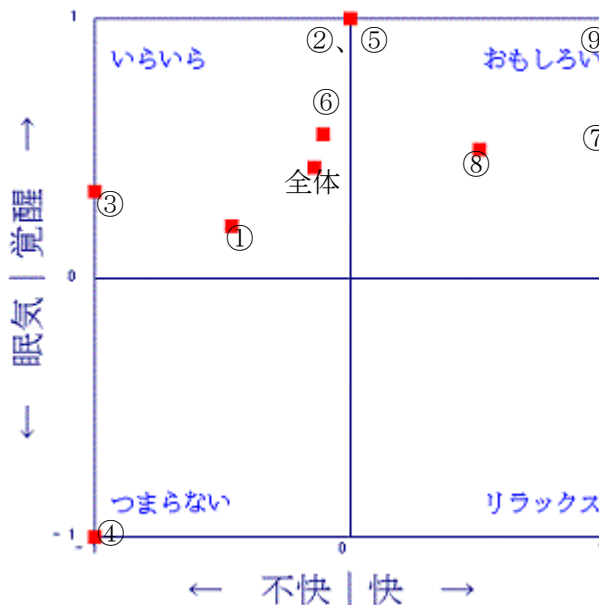


図3.3 小川洋子著『博士の愛した数式』の書評による散布図

書評はテキスト・データであり、比較するためには読み比べなければならないが、数値化することで、散布図を用いた比較が可能となる。図3.3をみると、“書評④”と“書評⑨”が、対照的な評価をしていることが分かる。さらに、全体との比較もできる。“書評⑥”は“書評全体”と近い評価をしていることが分かる。全体的には、「おもしろい」のクラスと「いらいら」のクラスに分かれている。

散布図により、書籍についての書評から、

以下の情報を得ることができる。

- 個別の書評が示す雰囲気の違い感
- 個々の書評と書評全体が示す雰囲気の違い感
- 書評が示す書籍の雰囲気的全体的な傾向

### 3.3.2 レーダー・チャートによる可視化

レーダー・チャートによる可視化では、書籍の雰囲気の詳細な状況を示す。表3.6と図3.4に小川洋子著『博士の愛した数式』についての書評を分析した結果をレー

表3.6 小川洋子著『博士の愛した数式』の書評の分析結果

書評	おもしろい	いらいら	リラックス	つまらない
書評①	0.2	0.4	0.067	0.333
書評②	0.5	0.5	0	0
書評③	0	0.667	0	0.333
書評④	0	0	0	1
書評⑤	0.5	0.5	0	0
書評⑥	0.444	0.333	0	0.222
書評⑦	0.8	0	0.2	0
書評⑧	0.75	0	0	0.25
書評⑨	1	0	0	0
書評全体	0.388	0.327	0.041	0.245

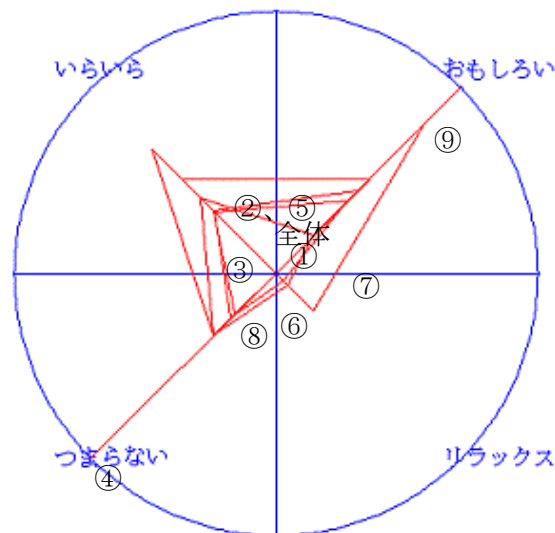


図3.4 小川洋子著『博士の愛した数式』の書評によるレーダー・チャート

レーダー・チャートとして表した例を示す。収集した書評に含まれる感情表現を集計して各象限にプロットした後、各点を線で結ぶ。個々の図形は、1つの書評に含まれる感情表現の分布を示すものであるが、複数の書評から求めた図形を同時に示すことで、書評に含まれる感情表現の詳細な分布を示すことができる。レーダー・チャートによる可視化は、ラッセルの「円環モデル」にならう。X軸(横軸)は「快-不快」次元を示す。Y軸(縦軸)は「覚醒-眠気」次元を示す。X Y軸の範囲は、 $-1 \leq (x, y) \leq +1$ とする。原点を(0,0)とする。感情表現の集計値を、 $y = x$ の示す線上と $y = -x$ の示す線上にプロットし、各点を線で結ぶ。

レーダー・チャートは要素の分布を示した図である。したがって、分析した書評の数だけ、レーダー・チャートが作画される。本稿では、個々の書評から作画された図を1つの座標軸上に表示している。個別の書評については線が重なってしまっていて分析しにくい、線の密集具合から感情表現の分布を確認することができる。

レーダー・チャートにプロットされた点は、円環モデルでの感情ベクトルそのものである。書評にはいくつかの感情表現が含

まれており、それらを感情ベクトルとして示すことができる。その感情ベクトルを書評ごとに線で結んだのが、本稿で示すレーダー・チャートである。散布図上の各点は感情ベクトルを合成したものであるから、このレーダー・チャートは散布図上の各点の根拠を示すものである。例えば、“書評④”や“書評⑨”の分析結果(図3.4)をみると、散布図上の点の位置の根拠は明らかである。“書評④”は、「つまらない」要素しか含まないために、散布図(図3.3)では左下角の(-1,-1)にプロットされる。“書評⑨”は、「おもしろい」要素しか含まないために、散布図(図3.3)では右上角の(+1,+1)にプロットされる。このように、レーダー・チャートによる可視化は、散布図による可視化を補完するものである。

## 4. 提案手法の実装

### 4.1 機能の実装

提案手法の有効性を検証するために、Webクライアント/サーバ・システムを構築する。検証システムはPerlにて実装し、1つのPC内にWebサーバとWebクライアントが共存する構成とした。検証システムの動作環境を表4.1と図4.1に示す。

表4.1 検証システムの動作環境

HW	Windows-PC(NEC LaVie Light BL350/DW6B)
OS	Windows 7 Starter 32bit
SW	ActivePerl 5.12.2 Build 1203(言語処理系) AnHttpd 1.42p(Webサーバ) Internet Explorer 9.0(Webクライアント) 茶筌 2.4.2(形態素解析)
主要要件	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推測 書籍の雰囲気と感情表現の分布を示す図の作画

## 4.2 感情表現辞書の作成

『感情表現辞典』から、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。1つの感情表現は、1つの語または複数の語から成り立つことから、感情表現辞書では、1つの語が1レコードに対応し、1つの感情表現が1つまたは複数のレコードに対応する。感情表現辞書のフォーマットと具体例を表4.2と図4.2に示す。辞書はcsv形式で作成する。

## 4.3 感情表現の抽出の実装

### (1) 茶釜による品詞の分解

日本語には活用のある語があり、単純な比較では感情表現は抽出できない。また、複数の語句が1つの感情表現を示す場合もある。本稿では品詞の分解に茶釜を利用する。茶釜を利用した書評の分解例を表4.3に示す。見出し(基本形)に品詞を分解する。

茶釜により「面白く読みました」を「面白い」、「読む」、「ます」、「た」に分解する。この実装により書評を分解する。

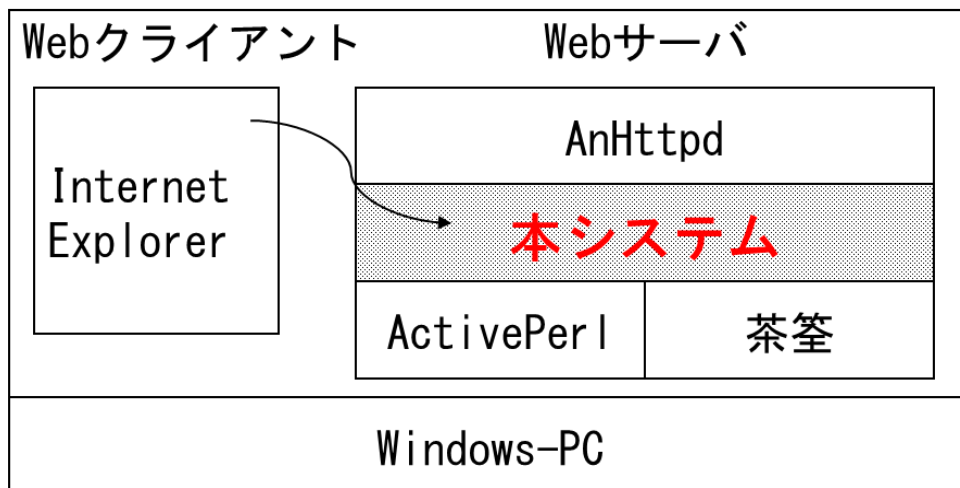


図4.1 検証システムの動作環境図

表4.2 感情表現辞書のフォーマット

第1項目	ラッセルの感情区分	
第2項目	『感情表現辞典』の感情区分	
第3項目	感情表現語	
第4項目	『感情表現辞典』でのページ数	
第5項目	茶釜での分解数	
第6項目	茶釜の出力1	見出し(出現形)
第7項目	茶釜の出力2	読み(カタカナ)
第8項目	茶釜の出力3	見出し(基本形)
第9項目	茶釜の出力4	品詞(1)
第10項目	茶釜の出力5	品詞(2)
第11項目	茶釜の出力6	活用形

(2) 感情表現辞書との比較による感情表現の抽出

感情表現辞書と茶釜を利用した書評の分析結果を比較し、一致した語句を書籍の雰囲気の区分ごとに集計する。感情表現辞書に登録された見出し(基本形)に、「面白い」を含む感情表現語の3語(表4.4)と入力文「面白く読みました」を比較する例を示す。  
 手順①茶釜により「面白く読みました」を「面白い」、「読む」、「ます」、「た」に分解する(見出し(基本形)単位)。  
 手順②見出し(基本形)の語句を比較レコード数だけ比較する。

感情表現辞書「面白い」+「ない」と入力文「面白い」+「読む」→不一致

感情表現辞書「面白い」+「がる」と入力文「面白い」+「読む」→不一致

感情表現辞書「面白い」と入力文「面白い」→一致

手順③感情表現区分として「おもしろい」をカウントする。

感情表現辞書と「面白く読みました」という文を比較することで「面白い」という感情表現を抽出し、感情表現区分として「おもしろい」をカウントすることができる。この実装により感情表現区分ごとの感情表現数をカウントする。

4.4 書籍の雰囲気の数値化の実装

4.3項の実装により求めた感情表現の

おもしろい,喜,面白い,P.24,1,面白い,オモシロイ,面白い,形容詞-自立,形容詞・アウオ段,基本形  
 いらいら,怒,面白くない,P.28,2,面白く,オモシロク,面白い,形容詞-自立,形容詞・アウオ段,連用テ接続  
 ,,,,ない,ナイ,ない,助動詞,特殊・ナイ,基本形

図4.2 感情表現辞書の具体例

表4.3 茶釜を利用した書評の分解例(入力文の例:「面白く読みました」)

見出し(出現形)	読み(カタカナ)	見出し(基本形)	品詞(1)	品詞(2)	活用形
面白く	オモシロク	面白い	形容詞-自立	形容詞・アウオ段	連用テ接続
読み	ヨミ	読む	動詞-自立	五段・マ行	連用形
まし	マシ	ます	助動詞	特殊・マス	連用形
た	タ	た	助動詞	特殊・タ	基本形

表4.4 見出し(基本形)に「面白い」を含む感情表現語

ラッセルの感情分類区分	感情表現語	比較レコード数	見出し(出現形)	見出し(基本形)
いらいら	面白くない	2	面白く	面白い
			ない	ない
おもしろい	面白がる	2	面白	面白い
			がる	がる
おもしろい	面白い	1	面白い	面白い

集計値から書評が示す書籍の雰囲気を数値化して平面上の点を求める。実装は3.2節に示したアルゴリズムにしたがう。この実装により書籍の雰囲気を数値化する。

#### 4.5 書籍の雰囲気の可視化の実装

4.4項の実装により求めた書籍の雰囲気の数値化の結果から散布図とレーダー・チャートを作成する。作画用のライブラリにはGDを利用する(表4.5)。この実装により書籍の雰囲気を可視化する。

#### 5. 提案手法の適用例

提案手法を検証するためのシステムを実装して、芥川賞や直木賞などを受賞した作品や書店員の投票により決定する本屋大賞を受賞した作品について、その書評をインターネット上から収集し分析を行った。25作品についての書評を10本ずつ、合計250本、収集し、192本の書評から書籍の雰囲気を推測することができた。

各書評から書籍の雰囲気を推測した結果

を表5.1に示す。全体の判定率は77%となり、提案手法の有効性が確認できた。純文学を対象とした書評の判定率が72%だったのに対して、娯楽小説を対象とした書評の判定率が88%となった。提案手法の娯楽小説に対する適合度が確認できた。

また、書籍の雰囲気の傾向を比較すると、純文学を対象とした書籍の雰囲気は「おもしろい」と「いらいら」に分かれるが、娯楽小説を対象とした書籍の雰囲気の傾向は「おもしろい」か「いらいら」に偏った分類がされることが確認できた。

純文学を対象とした書籍の雰囲気の例として、図5.1に小川洋子著『ミーナの行進』についての書評を分析した結果を示す。また、娯楽小説を対象とした書籍の雰囲気の例として、図5.2に角田光代著『対岸の彼女』についての書評を分析した結果(「いらいら」に分類される例)を、図5.3に沖方丁著『天地明察』についての書評を分析した結果(「おもしろい」に分類される例)を示す。

本稿では、書籍の雰囲気を「おもしろい」

表4.5 作画に使用したGDメソッド

作画する図	メソッド名	用途
散布図	points	データを点で示すグラフを作成する。
	set	グラフのオプションを設定する。
	plot	グラフを描画する。
レーダー・チャート	line	点と点を直線で結ぶ。
	arc	円を描く。
共通	gif	gif画像を作成する。

表5.1 提案手法により書籍の雰囲気を推測した結果

	おもしろい	いらいら	リラックス	つまらない	判定不能
純文学(180)	71(40%)	49(27%)	4(2%)	6(3%)	50(28%)
娯楽小説(70)	37(52%)	21(30%)	0(0%)	4(6%)	8(12%)
合計(250)	108(43%)	70(28%)	4(2%)	10(4%)	58(23%)

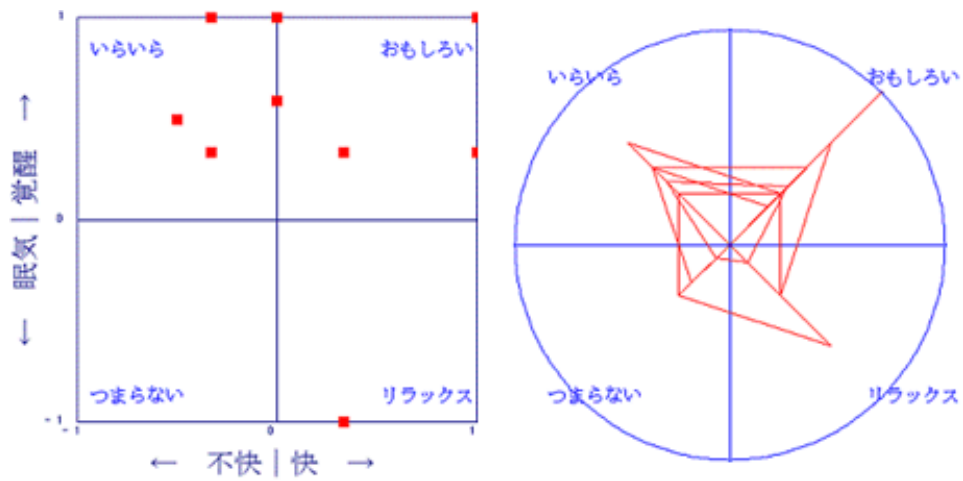


図 5.1 純文学を対象とした書籍の雰囲気可視化

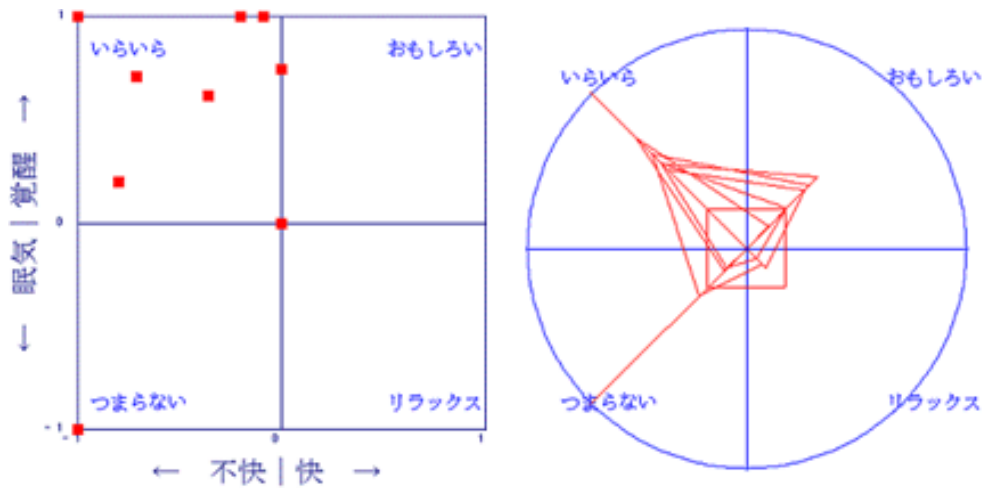


図 5.2 娯楽小説を対象とした書籍の雰囲気可視化  
 (「いらいら」に分類される例)

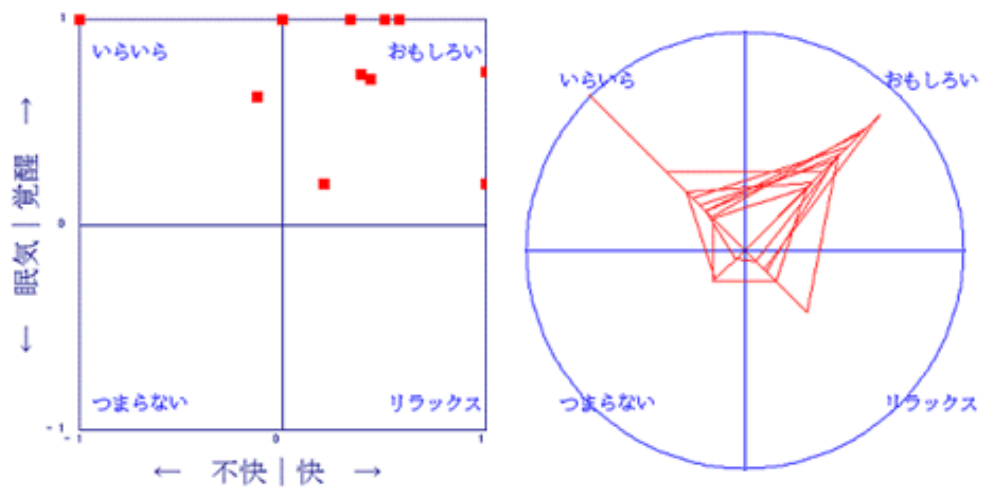


図 5.3 娯楽小説を対象とした書籍の雰囲気可視化  
 (「おもしろい」に分類される例)



と「いらいら」と「リラックス」と「つまらない」の4種類に分類したが、全体的な傾向としては「おもしろい」と「いらいら」の2種類に判定された。分類の基礎となる感情表現辞書の登録語数に偏りがあり、それが反映されたと考えられる。しかし、登録語数の偏りは、表現や語彙の豊かさに起因する日本語表現の問題とし、本稿では、そのような語彙の差を調整しなかった。

## 6. おわりに

本稿では、文芸書を対象としたインターネット上の書評について、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した。また、インターネット上の書評から書籍の雰囲気の全体像を示すことを試みた。

### (1) 書籍の雰囲気分類

書評に感情の分類モデルを適用し、書評から推測される書籍の雰囲気を、「おもしろい」と「いらいら」と「リラックス」と「つまらない」の4種類に分類した。また、感情表現を抽出して、感情の分類モデルを適用するための感情表現辞書を提案した。

### (2) 書籍の雰囲気推測

書評からそれぞれの区分に対応する感情表現を抽出して、その出現頻度から書籍の雰囲気を推測するアルゴリズムを提案した。感情表現辞書と茶筌と組み合わせることで、書評から感情表現を抽出して、感情の分類モデルを適用することを実現した。

### (3) 書籍の雰囲気可視化

書評から推測した書籍の雰囲気を散布図とレーダー・チャートによって可視化した。全体の雰囲気と個々の感情表現の分布が一望できる可視化を行った。

提案手法により書籍の雰囲気可視化が行えたことから、購読する文芸書を選ぶ際の判断材料を与えることが実現できた。

本稿をインターネット上の書評の検索と連動させることで、既存の検索順位とは違ったランキングが行え、書評検索の利便の向上が期待される。また、特に読書コミュニティ系のサイトでは、利用者間の嗜好の近しさを投稿された書評から分析する応用が可能と考える。

書籍の雰囲気推測結果は77%の判定率となったが、20%の判定率しか得られない書籍もあった。感情表現辞書として定義した感情表現は、やや古い文芸作品がベースになっている。また、感情の抽出を単語レベルで行っている。更なる精緻化には、辞書の拡張や文脈からの雰囲気推測、感情以外の要素の加味が必要と思われる。

## 謝辞

研究を進めるに際して、大曾根匡先生、植竹朋文先生から貴重なコメントをいただきました。深く感謝申し上げます。また、情報科学研究所の諸先生方にも様々なご助言をいただきました。お礼申し上げます。

## 参考文献

- [01] 書評コミュニティー本が好き!, <http://www.honzuki.jp/>, 2011年12月28日
- [02] 原田隆史、吉村紗和子, 「オンライン書評と従来の書評との違いの分析」, 情報知能学会誌, 第20巻, 第2号, P.65-P.72, 2010年
- [03] 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年
- [04] 中村明, 『感情表現辞典』, 東京堂出版, 1993年

- [05]大塚裕子、乾孝司、奥村学、『意見分析エンジン—計算言語学と社会学の接点—』, コロナ社,2007年,10月
- [06]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,第13巻,第3号,P.201–P.241,2006年
- [07]『chasen legacy -- an old morphological analyzer』, <http://chasen-legacy.sourceforge.jp/>,2011年10月2日
- [08]原田隆史,「書評中の感性キーワードを用いた小説の分類」,情報知識学会誌,第15巻,第2号,P.57–P.62,2005年
- [09]原田隆史,「最適解の付与に基づく重み付けの自動変更」,情報知識学会誌,第16巻,第2号,P.19–P.22,2006年
- [10]原田隆史、江藤正巳、大西美奈子,「レファレンスデータに対する NDC の自動付与」,情報知識学会誌,第17巻,第2号,P.61–P.64,2007年
- [11]原田隆史、江藤正巳、高柳知世,「書評を用いた図書に対する感性パラメータの自動設定」,情報知識学会誌,第18巻,第2号,P.153–P.160,2008年
- [12]原田隆史、池内正明,「書評中の語を用いた図書に対する感性パラメータの自動設定の最適化」,情報知識学会誌,第19巻,第2号,P.152–P.157,2009年
- [13]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,第21巻,第2号,P.291–P.296,2011年
- [14]『総合科目：文明の生態と居住環境』, <http://www.enveng.titech.ac.jp/sogokamo/ku/sogoVol12.pdf>, 2011年7月20日

## 付録

### 1) 検証システムの構成

検証システムは、システム本体と外部モジュールから構成される。

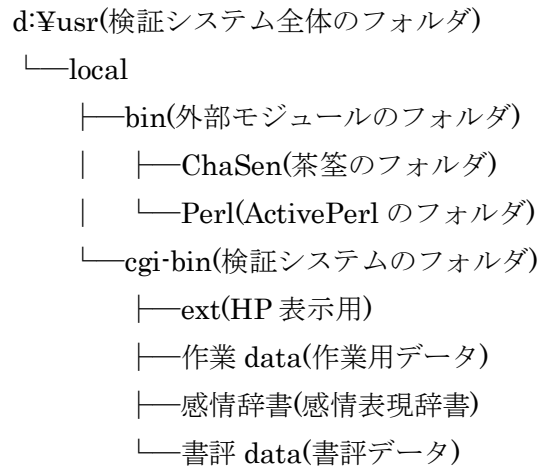


図1 検証システムの構成

### 2) 本体プログラムの構成

検証システムの本体プログラムは、クライアント/サーバ間インターフェースの呼び出し単位に作成されている。

表1 本体プログラムの呼び出し単位

呼び出し名	用途
post	出力画面の表示
graph	散布図の作画
rader	レーダー・チャートの作画
review	書評の表示
demo	デモ・モードでの実行
ohter	入力画面の表示