

実態調査に基づいた感情軸による感情辞書の構築と小説の感情分析

Construction of Emotional Dictionary With Emotional Axes Based on a Survey of Actual Conditions and Sentiment Analysis of Novels

安藤 冴華[†] 伊藤 淳子[‡] 吉野 孝[‡]
Sayaka Ando Junko Itou Takashi Yoshino

1. はじめに

自然言語処理を用いた感情分析研究は盛んに行われており、小説もその対象の一つである。感情と小説に関する研究の目的は、読み手の感情推定や選書支援などである。畠山 [1] や藤田ら [2] は、新聞記事を対象とする熊本ら [3] の手法を参考に感情辞書を構築し、小説の感情分析を行った。また、吉田ら [4] は、感情語辞書を構築し、小説内における感情の変遷を全体俯瞰する研究を行った。しかし、これらの研究で構築された感情辞書は、小説の登場人物が抱く感情や、文章の特徴を表現するためのものであり、読者側の感情を考慮していない。そのため、これらの感情辞書を選書支援に用いる際、どのような読後感を得られるか知りたいユーザーにとって不要な感情も含まれる。また、既存研究では小説の感情分析を、対象の小説を行単位や段落単位の細かい区分で分割し、実施している。細かい区分における感情分析は、局所的な感情表現に結果が左右される可能性がある。そのため、1つの小説の全体的な感情推移がわかりにくく、選書支援に用いるには不十分であると考えられる。

そこで本研究では、実際にユーザーが選書の際に考慮したい感情を対象とした感情辞書を構築する。アンケート調査によってユーザーが考慮したい感情を明らかにし、その感情に対応した感情辞書を構築する。また、小説を前半・中盤・後半に分割し、構築した感情辞書をもとに各区間の感情値を求める。行や段落単位の細かい区分ではなく、3つの話のまとまりの区分にすることによって、長い会話文や著者による意図的な段落分けに対応し、ユーザーが小説を読んだ際に実際に得る印象と近い感情分析が可能である。感情値の推移はグラフで出力し、視覚的に小説の感情遷移を明示する。小説を読んで得た感情と、グラフとを比較する実験を行い、構築した感情辞書を評価する。

2. 関連研究

2.1 新聞記事の印象に関する研究

熊本ら [3] は、新聞記事のデータベースを解析し、印象辞書を構築している。印象辞書には、印象軸の両端に定義される印象語群と、記事内に含まれる任意の単語との共起の仕方を数値化した結果が、その単語の印象値として登録されている。この辞書を用いて新聞記事の印象値を算出し、最終的に、実験によって人が実際に記事を読んだ際の印象値を得、それをもとに、印象辞書の印象値の補正を行った。

熊本らが構築した印象辞書では「楽しい⇔悲しい」、「うれしい⇔怒り」、「のどか⇔緊迫」の3軸6種類の印象を抽出することが出来る。この印象軸は、以下の手順によって新聞記事の印象を表現するのに適した3軸に決定された。はじめに、被験者が新聞記事を実際に読み、記事の印象を「楽しい」、「悲しい」等の印象語について、それぞれ5段階で評価した。この実

験で得られた結果を、因子分析手法やクラスタリング分析を用いて分析し、3つの印象軸を決定した。

本研究では、熊本らの印象値算出の手順を用いて、小説の感情分析のための感情辞書を構築する。

2.2 単語の共起関係に基づいた感情辞書の構築と小説の感情分析に関する研究

畠山 [1] や藤田ら [2] は、2.1 節で挙げた熊本ら [3] の手法を用いて感情辞書を構築し、小説の感情分析をしている。畠山 [1] は読み手の属性を仮定し、その読み手に適した感情辞書を再構築している。しかし、感情軸や感情語は、新聞記事を対象とした熊本らの印象辞書と同様であり、畠山の構築した感情辞書は、感情軸や感情語の選定の観点で小説の感情分析に適していない可能性がある。藤田ら [2] は、Plutchik 理論 [5] をもとに印象軸を定義している。この感情辞書では「喜び⇔悲しみ」、「好意⇔嫌悪」、「恐怖⇔怒り」の3軸6種類の印象について分析が可能であり、印象軸の両端の印象を表す印象語は、類語辞典を参考に各20個を定義している。しかし、この印象軸は選書に使用したいユーザーの感情を考慮しておらず、選書支援において余分な感情が含まれている可能性や、必要な感情が含まれていない可能性がある。また、小説の感情分析を19の区間に分けて実施しているが、藤田らの分析した小説は3000字前後～10000字前後の小説であり、特に3000字前後の小説では、1つの区間における文章量が1文～3文となる。細かい区間分けでは、長い会話文等に対応出来ず、不相当であると考えられる。

2.3 感情表現に基づいた小説の俯瞰分析に関する研究

吉田ら [4] は、感情表現辞典 [6] と感情極性対応表を融合した感情語辞書を構築し、小説を対象に感情抽出を行っている。吉田らの構築した感情語辞書は、語彙に対して、感情表現辞典に記載されている10種類の感情について正負の極性値を付与しており、形態素解析後の文章から感情表現を抽出することが出来る。1つの小説内における感情遷移の分析では、感情表現を行ごとに抽出し、感情値を算出して全体の感情遷移をグラフで表し、一覧できるようにした。また、複数の小説に対して、類似判定を実施した。類似判定では、1つの小説の文章を10分割し、それぞれの区間で正負の感情極性値と10の感情からなる特徴量を求めて1つの小説のデータとした。そして、自己組織化マップを用いて複数の小説の類似度を判定した。

この研究では、感情語辞書の構築の際に、選書に使用したいユーザーの感情を考慮していない。また、行単位で感情分析しているため、長い会話文や、著者による意図的な段落分け等の表現に対応できないと考えられる。

3. ユーザーの感情を考慮した感情辞書の構築

3.1 感情の選定

本研究では、熊本ら [3] や張ら [7] の提案手法を参考に、感情辞書を構築する。この手法は、新聞記事やニュースサイトを

[†] 和歌山大学大学院システム工学研究科, Graduate School of Systems Engineering, Wakayama University

[‡] 和歌山大学システム工学部, Faculty of Systems Engineering, Wakayama University

表 1: 事前アンケート項目

項目番号	質問内容
項目 1	読みたい小説を選ぶ際、考慮したい感情を 6 つ選んでください
項目 2	次の小説(星の銀貨)を読んで、感じた感情を 1 つ以上 6 つ以下の範囲で選んでください
項目 3	次の小説(鮎だま)を読んで、感じた感情を 1 つ以上 6 つ以下の範囲で選んでください
項目 4	次の小説(ねずみの嫁入り)を読んで、感じた感情を 1 つ以上 6 つ以下の範囲で選んでください
項目 5	次の小説(きのこ会議)を読んで、感じた感情を 1 つ以上 6 つ以下の範囲で選んでください

表 2: 感情に関する事前アンケートの項目 1 の結果

項目	感情									
	喜	怒	哀	怖	恥	好	厭	昂	安	驚
項目 1	9	3	9	8	0	9	1	6	5	10

表 3: 感情に関する事前アンケートの項目 2~5 の結果

項目	感情									
	喜	怒	哀	怖	恥	好	厭	昂	安	驚
項目 2	6	0	8	1	1	3	0	0	5	2
項目 3	8	0	0	6	0	4	0	1	6	6
項目 4	7	0	1	0	1	5	0	1	5	4
項目 5	0	2	8	5	3	0	3	0	1	6
項目 2~5 の合計	21	2	17	12	5	12	3	2	17	18



図 1: 決定した感情軸の図

対象としたものであるため、新たに小説の選書や推薦に適した感情を選定する。

感情の選定のため、感情辞書の構築に際しアンケート調査を行った。アンケート調査では、大学 3 年生~大学院 2 年生 10 名に対し、読みたい小説を選書する際に考慮する感情と、小説を読んだ際に実際に感じた感情を問うた。アンケート項目を表 1 に示す。全ての質問項目で、回答の選択肢は「喜」、「怒」、「哀」、「怖」、「恥」、「好」、「厭」、「昂」、「安」、「驚」の 10 種である。この 10 種は、感情表現辞典 [6] に項目のある 10 種の感情である。感情表現辞典は、近現代作家の作品で使われていた感情表現の用例や語句を、感情の種類別に収録しており、小説の感情分析に適していると考え、選択肢に採用した。また、項目 2~項目 5 では、それぞれ、青空文庫の XHTML ファイルのリンクを示した。被験者はリンク先にアクセスして対象の小説を読み、感じた感情すべてにチェックを入れる。項目 2~項目 5 は、被験者によって回答順をランダムに示し、小説を読む順番による影響を軽減した。

アンケート調査に使用した小説は、グリム兄弟作、楠山正雄訳の「星の銀貨」¹、新美南吉作の「鮎だま」²、楠山正雄作

の「ねずみの嫁入り」³、夢野久作の「きのこ会議」⁴の 4 作品である。これらの小説は、ブンゴウサーチ for Kids⁵で検索し、著作権の保護期間が満了した読了時間が 2 分~5 分の作品のうち、人気度の評価が 3 段階評価における最高の 3 であり、表記が新字新仮名の作品である。読了時間の下限を 2 分とした理由は、あまりに短い小説の場合に、被験者が得る感情の数が著しく少なくなることを懸念したためである。また、読了時間の上限を 5 分とした理由は、アンケート調査において、複数の小説を感情の観点から評価するという点で、被験者の負担を考慮したためである。ブンゴウサーチ for Kids は、青空文庫の児童文学を読了時間で検索できるサービスである。このサービスでは、各作品の文字数を 400 字毎分で割って読了時間を算出しており、人気度は過去 10 年分のアクセス数を基に算出している。

項目 1、項目 2~項目 5 のそれぞれの感情 1 つに対する被験者 10 名の得票数を表 2、表 3 で示す。項目 1 より選書の際にユーザが考慮する感情は得票数が多い順に「驚」、「喜」、「哀」、「好」、「怖」、「昂」となり、項目 2~項目 5 の合計値より実際に小説を読んで得られた感情は「喜」、「驚」、「哀」、「安」、「怖」、「好」となった。重複して得られた「喜」、「哀」、「驚」、「怖」、「好」の 5 感情と、感情軸の両端に対極的な感情を選定することを考慮し、これらの感情について、Plutchik 理論 [5] を加味した。その結果、感情軸を「喜び⇔哀しみ」、「怒り⇔怖い」、「期待⇔驚き」の 3 軸に決定した。3 軸を図にしたものを図 1 に示す。感情軸の決定において「喜」、「哀」、「驚」、「怖」の 4 感情は、アンケート調査と Plutchik 理論の基本 8 感情のどちらにも含まれているため、この 4 感情を基準とし、アンケート結果に含まれていない「怒」と「期待」を、Plutchik 理論の基本 8 感情における「怖」と「驚」の感情の対の感情として採用した。アンケート調査に含まれていた「好」の感情については、Plutchik 理論において「joy」と「trust」を混ぜて生成される「love」の感情に類似することから「喜」の部分的感情であ

¹https://www.aozora.gr.jp/cards/001091/files/42316_15930.html(2022 年 9 月 21 日確認)

²https://www.aozora.gr.jp/cards/000121/files/4723_13209.html(2022 年 9 月 21 日確認)

³https://www.aozora.gr.jp/cards/000329/files/18335_11944.html(2022 年 9 月 21 日確認)

⁴https://www.aozora.gr.jp/cards/000096/files/46694_27682.html(2022 年 9 月 21 日確認)

⁵<https://bungo-search.com/juvenile>(2022 年 9 月 21 日確認)

ると考え、本研究では採用しなかった。それぞれの感情軸のうち「喜び」、「怒り」、「期待」を軸の左側の感情とし「哀しみ」、「怖い」、「驚き」を軸の右側の感情とする。また、この6感情を各軸に対する両端の感情と呼ぶこととする。

3.2 感情語の選定

感情辞書を構築する際、感情軸の両端の感情を表す、感情語を決定する必要がある。本研究では、対応する感情について、感情表現辞典 [6] の語句編に収録されている感情表現を感情語として定義した。感情表現辞典には、1つの感情につき少なくとも50以上の感情表現が収録されている。感情辞書の構築には、感情語を含む小説を用いる。この際、既存研究より感情語の数や種類を増やすことによって、感情語を含む小説を多く抽出することが出来るため、感情辞書の構築に必要な小説となる対象小説数を増やすことができる。これによって、選書時にユーザが知りたい感情について、より精度の高い感情辞書を構築できると考えられる。感情表現辞典に記載のなかった「期待」については、類語検索サービス Weblio¹で「期待」、「期待する」を検索した際に検索結果として表示される語句を用いた。

3.3 感情辞書の構築

感情辞書の構築は、熊本ら [3] や張ら [7] の算出式を用い、以下の手順で行う。

手順 1. 小説を左側の感情の感情語を多く含む小説集合 S_L と、右側の感情の感情語を多く含む小説集合 S_R に分類する。

手順 2. 小説内に出現したある単語 w が、 S_L 、 S_R においてそれぞれ何件の小説で出現したかを表す条件付確率を算出する。

手順 3. 手順 2 で求めた条件付確率を用いて、ある単語 w の感情値を求める。

感情辞書の構築に用いた小説は、青空文庫で公開されている、著作権の保護期間が満了した児童文学 957 件である。児童文学を対象とした理由は、比較的短い文章中に感情表現が多数含まれていると考えたためである。児童文学は、一般に、児童に向けて書かれた文学とされる。そのため、児童にとって、話の内容を理解しやすくするため、多種多様な感情表現が数多く含まれていると考える。可能な限り多くの作品から多様な感情表現や単語を抽出し、感情辞書に登録することを目指すため、児童文学を本研究で対象とした。

手順 2 の条件付確率は式 (1) で求めることができ、手順 3 で求める単語の感情値は式 (2) で表すことが出来る。式 (1) では、小説集合 S_L 、小説集合 S_R のそれぞれの要素数を N_L 、 N_R と表記している。また、ある単語 w が小説集合 S_L において出現した小説数を $N_L(w)$ 、小説集合 S_R において出現した小説数を $N_R(w)$ とし、条件付確率を $P_L(w)$ 、 $P_R(w)$ で表している。

$$P_L(w) = \frac{N_L(w)}{N_L}, \quad P_R(w) = \frac{N_R(w)}{N_R} \quad (1)$$

ある単語 w の感情値 $s(w)$ を式 (2) で表す。なお、 $weight_L$ と $weight_R$ は重みであり、小説の数が多いほど、大きくなる。

$$s(w) = \frac{P_L(w) * weight_L}{P_L(w) * weight_L + P_R(w) * weight_R} \quad (2)$$

$$weight_L = \log_{10} N_L, \quad weight_R = \log_{10} N_R$$

¹<https://thesaurus.weblio.jp/>(2022年12月2日確認)

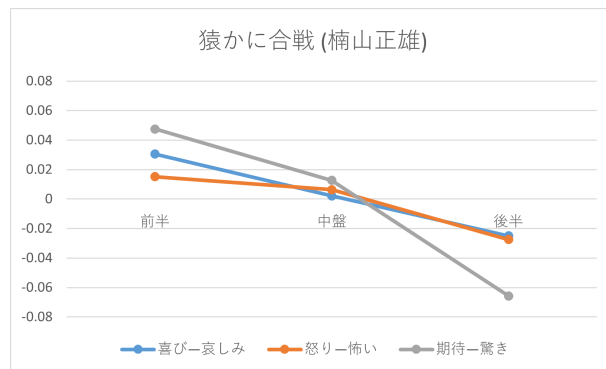


図 2: 「猿かに合戦」の分析結果

4. 感情辞書の評価実験

4.1 小説の感情分析

構築した感情辞書を用いて、小説の感情分析を実施した。実験対象とした小説は、楠山正雄作「猿かに合戦」²、楠山正雄作「花咲かじじい」³、小川未明作「月夜とめがね」⁴、宮沢賢治作「月夜のでんしんぼしら」⁵である。以降、実験対象としたこの4作品を「猿かに合戦」、「花咲かじじい」、「月夜とめがね」、「月夜のでんしんぼしら」と記す。これらの小説は、ブンゴウサーチ for Kids で検索し、著作権の保護期間が満了した読了時間が10分前後の作品のうち、人気度の評価が3段階評価における最高の3であり、表記が新字新仮名の日本の児童文学を選出し、その後被験者へのアンケート調査によって選定した。10分前後で読了できる小説を対象とした理由は、複数の小説に対して、前半・中盤・後半・全体のそれぞれの区間について感情の観点から評価を行う被験者の負担を考慮したためである。また、上記のように、小説を内容に応じて3つの区間に分割するため、全体の文章量が少なく、1つの区間に含まれる文章量が極端に少なくなり、被験者が感情を得られない可能性がある。そのため、1つの区間につき、少なくとも5文以上の文章量を確保できる、読了時間が10分程度の小説を対象とした。ブンゴウサーチ for Kids では、400字毎分で小説の読了時間を算出しているため、10分前後で読了可能な小説は、およそ4000字前後である。そのため、3500字以上5000字未満の小説を選出した。「猿かに合戦」、「花咲かじじい」の2作品は、被験者へ実施した小説の既知・未知を問うアンケート調査により、既知の小説として扱い「月夜とめがね」、「月夜のでんしんぼしら」の2作品は未知の小説として扱う。既知の小説と未知の小説の分類によって、既知の小説はユーザが予め小説の内容を知っているため、その小説を読んで得られる感情についても既知である可能性が高いという点を考慮する。

感情分析は、小説を内容によって前半・中盤・後半に分割し、各区間と全体に対して行った。「猿かに合戦」、「花咲かじじい」は、本文中に「一」、「二」、「三」の漢数字で示された区間が存在するため、それぞれの区間を前半・中盤・後半部分の区間とした。「月夜とめがね」と「月夜のでんしんぼしら」については、実験者がそれぞれの小説を読み、場面の区切りであると感じた部分で前半・中盤・後半の3つの区間に分割した。小説の感情分析結果の一部を、楠山正雄作「猿かに合戦」を例として図2に示す。感情値は、藤田ら [2] の算出方法と同様に、

²https://www.aozora.gr.jp/cards/000329/files/18334_11947.html(2023年1月23日確認)

³<https://www.aozora.gr.jp/cards/000329/files/3391.html>(2023年1月26日確認)

⁴https://www.aozora.gr.jp/cards/001475/files/54404_46246.html(2023年1月26日確認)

⁵https://www.aozora.gr.jp/cards/000081/files/43756_17661.html(2023年1月26日確認)

表 4: 「猿かに合戦」の感情一致度についてのアンケート結果

対象区間	評価値					中央値	最頻値
	1	2	3	4	5		
前半	1	2	1	3	3	4.0	4,5
中盤	0	1	5	4	0	3.0	3
後半	0	1	0	6	3	4.0	4
全体	0	1	0	6	3	4.0	4

表 5: 「花咲かじじい」の感情一致度についてのアンケート結果

対象区間	評価値					中央値	最頻値
	1	2	3	4	5		
前半	0	0	1	6	3	4.0	4
中盤	0	4	1	2	3	3.5	2
後半	1	1	4	1	3	3.0	3
全体	0	2	2	4	2	4.0	4

各区間の感情値と小説全体における感情値を求め、各区間の感情値から全体の感情値を引くことによって求めた。正の値は感情軸の左側の感情を、負の値は感情軸の右側の感情を表す。

感情辞書の評価実験では、前半・中盤・後半のそれぞれの区間について、小説を読んで実際に被験者が得た感情を問うた。また、前半・中盤・後半・全体のそれぞれの区間について分析結果の一致度について問うた。前者は感情表現辞典に収録されていた 10 種の感情に「その他」を追加した計 11 種の選択肢から複数選択、後者は 5 段階で評価する形式である。1~5 の 5 段階評価では、1 が「あてはまらない」、5 が「あてはまる」を表す。さらに、自由記述形式の回答項目を設け、実験全体に対する意見や、選択肢形式では表現出来ない感情について問うた。これらの実験結果と小説の感情遷移のグラフから感情辞書の評価した。

4.2 実験結果と考察

「猿かに合戦」の実験結果を、表 4 に示す。実験の結果「猿かに合戦」では、前半・中盤・後半・全体の全ての区間で、中央値が 3 以上であった。また、最頻値についても、前半・中盤・後半・全体の全ての区間で 3 以上であった。このことから、感情辞書による感情分析結果と、被験者が得た感情は概ね一致していると推測される。

「花咲かじじい」に対するアンケート結果を表 5 および図 3 に、感情分析結果を図 4 に示す。実験の結果「花咲かじじい」では、中盤部分では 5 段階評価のうち、一致しないを表す 2 を選んだ被験者が計 4 名、3 を選んだ被験者が計 1 名、一致することを表す 4, 5 を選んだ被験者が計 5 名となった。この作品の中盤部分の「驚」の感情について、自由記述形式の回答より、強さが一致していると感じる被験者と、一致していないと感じる被験者の双方が確認できた。また、図 3 より被験者の半数が「驚」の感情を得たと回答している。これらの回答と結果から「驚」の感情の感じ方は特に個人差が大きいと考えられる。また、中盤の「哀」の感情について、図 3 より、10 名のうち 9 名が得たと回答している。これに対し、感情辞書による分析結果では、図 4 より「喜び⇔哀しみ」軸において「哀しみ」を示しているが、その絶対値が他の軸に比べて小さくなっている。そのため「哀」の感情についての感じ方の度合いの点で、感情辞書による分析結果と被験者が得た感情は不一致であると評価した被験者が存在したと示唆される。

「月夜とめがね」に対するアンケート結果を表 6 および図 5 に、感情分析結果を図 6 に示す。なお、図 6 では、感情の変化がわかりやすいよう、拡大表示をしている。実験の結果「月夜とめがね」では、前半部分では 5 段階評価のうち、一致しな

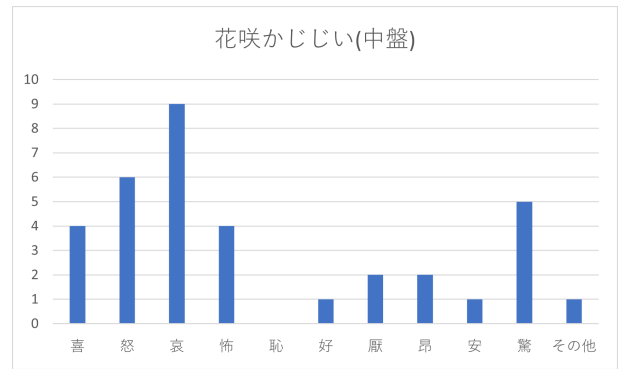


図 3: 「花咲かじじい」中盤の感情得票数

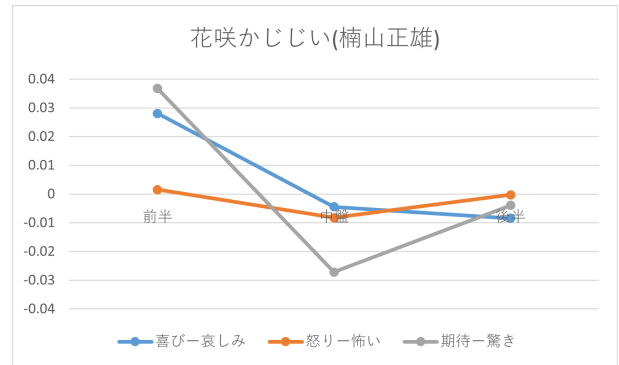


図 4: 「花咲かじじい」の分析結果

表 6: 「月夜とめがね」の感情一致度についてのアンケート結果

対象区間	評価値					中央値	最頻値
	1	2	3	4	5		
前半	1	6	2	1	0	2.0	2
中盤	1	3	2	4	0	3.0	4
後半	1	1	2	6	1	4.0	4
全体	0	3	3	4	0	3.0	4

いを表す 1, 2 を選んだ被験者が計 7 名、3 を選んだ被験者が計 2 名、一致することを表す 4 を選んだ被験者が計 1 名となった。この作品の前半部分について、自由記述形式の回答より、「怒」の感情を感じなかった被験者が複数確認できた。また、図 5 より「月夜とめがね」前半において「怒」の感情を得た被験者は 0 名であったことがわかる。一方、感情辞書による分析結果の図 6 では「怒り⇔怖い」軸において「怒り」を示しており、絶対値も他の 2 軸と比べて大きく、強く「怒り」の感情を示している。以上のように、被験者が得なかった感情について、感情分析の結果で大きく出現している部分があったため、前半部分の一致度が低かったと推測される。

前半部分の感情辞書による分析結果において「怒り」の感情値が大きく算出された理由について考察する。感情辞書の「怒り⇔怖い」軸における「月夜とめがね」の前半部分に含まれる単語と感情値の組み合わせを参照すると「しごと」、「わかい」、「孫娘」等の単語に強い怒りを示す感情値が設定されていた。感情値の算出過程をみると、これらの単語は、 $N_R(w)$ 、つまり小説集合 S_R において出現した小説数の値が 0 であった。 $N_R(w)$ の値が 0 であるとき、3.3 節の式 (1)、式 (2) より、単語の感情値は最大値の 1 となる。以上より、小説集合における、ある単語の出現小説数が少ない場合は、単語に設定される感情値が極端になることが明らかになった。小説本文の感情値は、単語に設定された感情値を用いて算出するため、単語の感情値

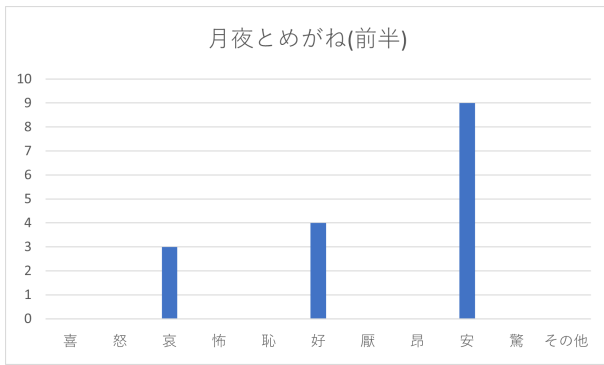


図5: 「月夜とめがね」前半の感情得票数

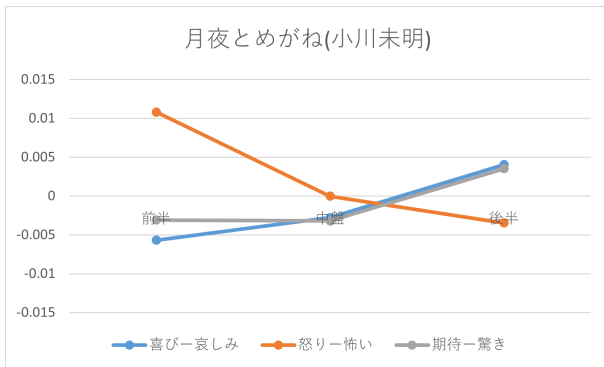


図6: 「月夜とめがね」の分析結果

によって、感情分析の結果は大きく左右される。よって、本研究で用いた感情値の算出方法は、小説集合におけるある単語の出現小説数が少ない場合には適さないと考えられる。

「月夜のでんしんばしら」について、実験結果を表7に、感情遷移のグラフを図7に示す。なお、図7では、感情の変化がわかりやすいよう、拡大表示をしている。「月夜のでんしんばしら」では、中盤以外の全ての区間で4、5を選んだ被験者の数よりも1、2を選んだ被験者の数の方が多く、概ね不一致の結果となった。図7より「月夜のでんしんばしら」の感情辞書による分析の感情遷移グラフでは、大幅な感情推移や変化が見られ、大きな感情変動が見られる。しかし、実際に読んだ被験者は感情の変化を強く感じず、不一致と評価した。感情辞書による分析で感情値が大きく変動した理由は、感情辞書の構築の際、出現回数が少ない単語に対して極端な感情値が設定される計算式になっていたためである。「月夜のでんしんばしら」には、作者独自の独特な表現が多く含まれていたこともあり、被験者と感情辞書による分析の結果が乖離する結果となったと考えられる。

評価実験全体に対しては、より多くの感情についても提示してほしいという指摘や、提案手法の3軸6感情だけでは表現出来ない感情がある、感情の対極性に違和感を感じるという指摘が見られた。これらの指摘により、感情の選定方法にも課題があると考えられる。

5. おわりに

本研究では、ユーザの感情を考慮した感情辞書を構築し、感情辞書の分析の妥当性について評価した。その結果、本研究で用いた感情辞書の構築手順では、小説集合全体において出現回数が少ない単語には、極端な感情値が単語に設定される算出式となっており、小説の感情分析に使用する際には算出式の工夫が必要であることが明らかになった。また、構築した感情辞書によってそのような単語を多く含む小説を感情分析した場合、

表7: 「月夜のでんしんばしら」の感情一致度についてのアンケート結果

対象区間	評価値					中央値	最頻値
	1	2	3	4	5		
前半	1	3	3	2	1	3.0	2,3
中盤	2	3	0	5	0	3.0	4
後半	1	3	6	0	0	3.0	3
全体	1	5	3	1	0	2.5	2

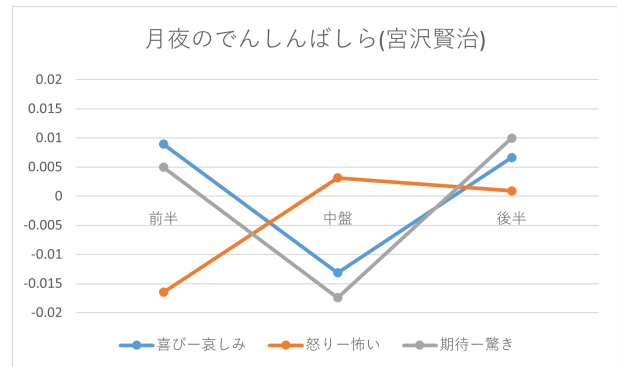


図7: 「月夜のでんしんばしら」の分析結果

感情辞書による分析の結果とユーザが実際に小説を読んで得る感情との間に大きく差が生じることが明らかとなった。さらに、本研究で選定した6つの感情だけでは表現できない感情や、足りない感情が存在し、ユーザはそれらの感情についても考慮したい可能性があることが明らかになった。

今後の課題として、感情辞書に採用する感情の選定方法や、単語の感情値の計算方法について再検討する必要がある。そのため、感情値を求めるための計算式について、改善と検証を行う。計算式は、単語の出現回数によって重みをつける等の改善を検討中である。

謝辞

本研究の一部は、JSPS 科研費基盤研究 (C) (JP22K12110) の助成を受けたものである。

参考文献

- 1) 島山侑也: 感情辞書による小説の分析に基づいた読み手の感情推定, 法政大学大学院紀要. 情報科学研究科編, Vol.14, pp.1-6 (2019).
- 2) 藤田佳恵, 鈴木伸崇: 小説データの本文分析による選書支援手法の提案, 情報処理学会第76回全国大会講演論文集, Vol.2014, No.1, pp.519-520 (2014).
- 3) 熊本忠彦, 河合由起子, 田中克己: 新聞記事を対象とするテキスト印象マイニング手法の設計と評価, 電子情報通信学会論文誌D, Vol.J94-D, No.3, pp.540-548 (2011).
- 4) 吉田知世, 小林一郎: 感情極性に基づく文書の俯瞰分析への取り組み, 情報処理学会第73回全国大会講演論文集, Vol.2011, No.1, pp.387-388 (2011).
- 5) R. Plutchik: The Nature of Emotions, American Scientist, Vol.89, No.4, pp.344-350 (2001).
- 6) 中村明: 感情表現辞典, 東京堂出版 (1993).
- 7) 張建偉, 河合由起子, 熊本忠彦, 白石優旗, 田中克己: 多様な印象に基づくニュースサイト報道傾向分析システム, 知能と情報 (日本知能情報ファジィ学会誌), Vol.25, No.1, pp.568-582 (2013).