

第 28 回年次大会予稿

感情語辞書に基づく物語の登場人物の感情抽出 Emotion Extraction of Characters in Stories Based on Emotion-Word Dictionary

逢坂 駿也^{1*}, 村井 源¹
Shunya OSAKA, Hajime MURAI

¹ 公立はこだて未来大学

Future University Hakodate

〒041-0804 北海道函館市亀田中野町 116-2

* 連絡先著者 Corresponding Author

本研究では、物語から登場人物の感情状態を抽出するため 12 種類の感情に分類した感情語の辞書を構築した。分析対象として「キノの旅 -the Beautiful World-」という小説を選択し、辞書を用いて感情語を抽出した。感情語の抽出回数を多次元データとして扱い因子分析をかけた結果、皮肉な物語の因子・意外性の強い物語の因子・悲喜劇的な物語の因子の 3 つに分類された。また各因子での登場人物の役割ごとの感情の差異を抽出した。

In this study, we constructed a dictionary of emotion words classified into 12 types of emotions in order to extract the emotional state of the characters from the story. We selected a novel called "Kino's Journey -The Beautiful World-" as an analysis target, and extracted emotion words using a dictionary. The number of emotional words extracted was treated as multidimensional data and factor analysis was performed. As a result, it was classified into three factors: ironic story factors, surprising story factors, and tragic story factors. In addition, we extracted the differences in emotions for each role of the characters in each factor.

キーワード: 自然言語処理, 登場人物, 感情
Natural Language Processing, Characters, Emotion

1 背景

自然言語処理技術を用いて物語テキストから登場人物の特徴や関係性などの物語に置いて重要な情報を抽出する研究はこれまで数多く行われている。馬場ら [1] は、小説テキストから登場人物の人名、性別、年齢、年代、職業、身体的特徴、性格といった属性を抽出することに成功している。Liu ら [2] は、SNS へのユーザーの投稿をテキストコーパスとして利用した性格予測モデルを独自に作成し、それを小説の登場人物に当てはめて登場人物の性格の推定を行った。しかし、こういった研究は実用的に十分な精度がいまだに得られていないものが多い。理由としては、同様の属性の表記であっても記述の仕方が様々に異なるということや文章中において主語などの省略が多い可能性が考えられる。そこで本研究ではデータソースを物語テキストに限定し、抽出対

象を登場人物の感情状態に限定した上でテキストからの情報抽出を試みた。これらの限定化により、物語分析に必要な辞書構築やアルゴリズムの推定が汎用的なテキストの場合よりも容易になり精度の向上が期待できると考えられる。

本研究では、複数の感情に分類した高精度の感情語の辞書を構築し、その辞書を用いて物語テキストにおける登場人物の感情状態の抽出を行い、その後抽出した情報に基づいて考察を行うことを目的とした。感情語の辞書は物語に限らず、会話文やインターネット上の口コミなど広範囲に応用可能である。そして本研究の成果が文章の自動生成などの関連研究のために有用なものになることを目指した。

2 関連研究

人間の感情の推定にはこれまでに多くの方法が提唱されている。その一つとして、人間の発話や記述した言葉から感情を推定するという方法がある。高村ら [3] や佐藤ら [4] は単語をポジティブなものとネガティブなものに分類する、極性分類の研究を行った。

また、極性分類では感情をポジティブなものとネガティブなものの2種類に分類されるが、人間の感情は2種類では収まらないと主張する学者は多く存在する。心理学者の Ekman [5] は感情を6種類に分類し、Plutchik [6] は感情を8種類に分類した。また、Russell [7] は全ての感情は「快-不快」「覚醒-非覚醒」の2次元平面上で表される円環モデルで表現可能であると提唱した。以上のように感情を多種類に分類する手法を用いて感情語辞書の構築を行うことで物語の登場人物の感情状態をより細かく抽出することができると考えられる。

3 対象データ

本研究では分析対象およびデータソースとして「キノの旅 -the Beautiful World-」[8] というライトノベル小説を選択した。以下では表現の簡単のため「キノの旅」と記載する。

「キノの旅」という作品の特徴としては、旅人である「キノ」が相棒である「エルメス」と旅をしながら様々な国を訪れるという短編および一話完結型のファンタジーであることが挙げられる。上記の二名以外にも、「師匠と相棒」「シズと陸」「フォトとソウ」が主人公となる話もあるが、主人公が二名であるという話の構成は変わらない。基本的には主人公らが、独特の制度や価値観、技術を持つ国家および国民と関わっていくという内容のストーリーになっている。

「キノの旅」を分析対象およびデータソースとして選択した理由は大きく三つある。一つ目は、本研究を行うためには、さまざまな属性を持った登場人物が登場し、登場人物間の多種多様かつ複雑な人間関係が描写されている物語作品を選択する必要があるためである。二つ目は、話数が豊富なためである。「キノの旅」は20年に渡って連載が継続されており、話数も258と分析を行うには充分な量であると判断した。三つ目は、社会的に評価されている作品であ

るためである。「キノの旅」は二度のアニメーション化およびラジオドラマ化がされている。また、長期連載が継続されていることや新聞で週刊連載がされていたこと、シリーズが累計820万部を超えていることなどからも社会的に高い評価がされている作品であるということが推測できる。以上の3点の理由から「キノの旅」を分析対象およびデータソースとして選択した。

4 辞書構築

辞書を作成するにあたり、中村 [9] の感情分類手法を参考にし、感情を「諦」「無」「喜」「怒」「哀」「怖」「恥」「好」「厭」「昂」「安」「驚」12種類に分類することに決定した。以下に各感情についての言葉の特徴を述べる。

- 諦：諦めの印象を与える言葉
- 無：虚無的で呆けた状態を表す言葉
- 昂：興奮状態を表す言葉
- 恥：恥ずかしさや緊張状態を表す言葉
- 好：対象の人や物を好きな状態や満足感を表す言葉
- 怖：恐怖状態を表す言葉
- 驚：驚きを表す言葉
- 厭：嫌悪感を表す言葉
- 怒：怒りを表す言葉
- 喜：喜びを表す言葉
- 哀：悲哀や憂鬱状態を表す言葉
- 安：安心感を表す言葉

辞書を作成する際、評価値表現辞書 [10] および分類語彙表 [11] を一部引用した。これらの辞書およびテキストコーパスは本研究とは別の用途、目的のために作られたものであり、本研究に直接用いることはできない。しかし、辞書を作成するにあたり日本語の語彙の収集から初めてしまうと膨大な手間、時間がかかると推測した。そこで既存の辞書から語彙のみを引用することで辞書作成の短縮化を図った。引用した語彙に対して手動で感情の分類を行った。

辞書を作成した結果、2018 語を感情語として採用した。また、作成した辞書中には一つの語彙に複数の感情を持つと定義されているものも存在する。例えば「うなだれる」という言葉には「哀」「恥」の2種類を定義した。それぞれの感情における語彙の数量は以下の表 1 に記載する。

表 1: 分類の一覧と数量

分類	主体的	客体的
諦	69	0
無	22	0
昂	249	5
恥	53	10
好	227	7
怖	170	12
驚	130	10
厭	539	25
怒	204	8
喜	223	22
哀	197	4
安	132	13
合計	2215	116

以下の表 2 に感情語の例を記載する。

表 2: 感情語の例

種類	例
諦	手に余る, 諦観する, 敷居が高い
無	つまらない, 愛想の無い, 興ざめ
昂	熱狂する, 奮い立つ, 笑い出す
恥	緊張する, 照れる, 赤面する
好	好き, 素敵, 慈しむ
怖	恐い, ぞっとする, 不安
驚	目が点になる, 動転する, 呆然
厭	疎む, 煙たがる, 不快
怒	憤る, 腹が立つ, 目に余る
喜	鼻が高い, 清々しい, 喜ぶ
哀	傷心する, 寂しい, 泣きわめく
安	おだやか, 心温まる, 安らぐ

5 分析

物語テキスト内の感情語の出現回数を 1 話ごと、各感情ごとに計測した。具体的には形態素解析エン

ジン MeCab [12] を用いて物語テキストを分かち書きし、同じく分かち書きされた感情語との一致回数を各感情ごとに計測した。抽出した結果、各感情ごとの抽出回数に大きな差が生じた。因子分析をする際には、データに 0 の割合が多いと一般的には計算が上手く行われないことが知られている。そのため、データを一部結合し、12 次元から 8 次元のデータに変更した。具体的には「驚」「昂」, 「厭」「恥」「諦」, 「哀」「無」を結合した。以下それぞれを「驚(昂)」, 「厭(恥, 諦)」, 「哀(無)」と定義する。「驚(昂)」はどちらも興奮状態であり、感情状態としては近いものであると判断し、結合した。「厭(恥, 諦)」はいずれもネガティブな印象を持っているが、「怒」ほど激しくはなく「哀」ほどネガティブな度合いが強くない。よって感情状態としては近いものであると判断し、結合した。「哀(無)」においても感情状態としては近いものであると判断し、結合した。

5.1 因子分析

物語を分類するにあたって、本研究では分析手法として因子分析を用いた。因子分析とは多変量解析手法の一つであり、複数の変数を持つデータの元にある少数の要因を特定する際に用いられる。分析環境としてはプログラミング言語「R」を用い、プロマックス回転を採用し分析を行った。分析結果を表 3 に示す。

表 3: 因子分析結果

	因子 1	因子 2	因子 3
好	-0.070	-0.118	-0.188
怖	0.023	-0.079	0.480
驚(昂)	-0.044	0.712	0.003
厭(恥, 諦)	-0.425	-0.592	0.656
怒	-0.171	0.138	0.146
喜	1.039	-0.182	0.305
哀(無)	-0.089	-0.037	-0.435
安	-0.121	0.510	-0.017

5.2 因子得点による物語分類

各データの因子得点を用いて物語を 3 つに分類した。本研究では因子得点が 0.4 以上のデータを分類

の対象とした。また、複数の因子において因子得点が0.4以上のデータに関しては分類の対象外とした。因子1の因子得点が規定の値に達している物語データを第1グループと表現し、因子2、因子3についても同様にして第2グループ、第3グループと表現する。

5.3 登場人物ごとの感情抽出

5.2節で記載した物語の分類結果を用いて、物語の登場人物ごとの感情抽出を行った。「キノの旅」の物語形式の特徴は3章で述べたが、それぞれの話に主人公が二名登場し、毎話ごとに新しい国に訪れ、その国で登場する人物と関わっていくという特徴が挙げられる。そこで、主人公二名とその他の登場人物一名の計三名ずつの感情状態を計測した。主人公の二名は主人公A、主人公Bと表記し、その他の登場人物についてはサブキャラクターと表記する。サブキャラクターは物語に登場した回数が一番多い人物を抽出回数とした。また、サブキャラクターが登場しない物語は今回の解析の対象外とした。抽出には構文解析エンジンKNP [13]を使用した。抽出結果を表4表5に示す。

6 考察

6.1 各因子についての考察

本節では因子分析の結果および因子得点を用いた物語分類結果より各因子ごとの考察を行う。また、因子負荷量の絶対値0.3以上の感情語を考察の対象とする。

因子1では、「喜」について1.039、「厭(恥, 諦)」について-0.425となっている。また、第1グループの物語の特徴として、シニカルな雰囲気演出している作品が多いということが挙げられる。具体例として、第1グループのある物語のあらすじを記載する。主人公が訪れた国の国民は、自国にある氷河を崩して地球を暖かくし、住みやすい世界に変えるという目的を持っており、それに対して国民が喜びや幸せを感じているという話である。このように、主人公らと訪れた国の国民の間で価値観の違いが生じている話が多い。グループ全体として「喜」の感情語が多数抽出されていることから、作者は「喜」の感情語のような、明るい・幸せそうといったイメー

ジを持つ言葉を多用することでシニカルな喜びを表現しているのではないかと推測した。以上のことから因子1を皮肉的物語因子と名付けた。

因子2では、「驚(昂)」について0.712、「安」について0.510、「厭(恥, 諦)」について-0.592となっている。また、第2グループの物語の特徴として、登場人物への理不尽さや、読者への意外性を演出している作品が多いということが挙げられる。具体例として、第2グループのある物語のあらすじを記載する。主人公が訪れた国では、月に数回地面に直径200メートルほどの穴が空いてしまいそれによって死人も出てしまうが、国民たちは誰もそれを問題視していないことに主人公が驚愕するとともに理不尽さを感じるという話である。グループ全体として「驚」の感情語が多数抽出されていることから、抽出結果と作品の特徴が一致しているのではないかと推測した。以上のことから因子2を意外性物語因子と名付けた。

因子3では、「厭(恥, 諦)」について0.656、「怖」について0.480、「喜」について0.305、「哀(無)」について-0.435となっている。また、第3グループの物語の特徴として、幸福な明るい情景から一転して不幸な暗い情景になるといった作品や、反対に不幸な暗い情景から幸福な明るい情景に変化する作品が多いことが挙げられる。具体例として、第3グループのある物語のあらすじを記載する。主人公が訪れた国では、投票されない人を処分する「投票祭り」というイベントがあるが、いままで投票されなかった人はいないらしいということを、主人公たちに優しく対応してくれていたホテルのオーナーの男が主人公に教えてくれた。次の朝、そのオーナーが死んでしまい、実は彼は誰からも投票されていなかったの

で国の政府が秘密裏に処分していたことがわかったという話である。男の目線に立つと、ホテルのオーナーという充実・安定した職について幸せに暮らしていたが、実は国の誰からも必要とされていなかったという事実が判明する。幸せな描写から一転して不幸な描写になっている。グループ全体として「厭」「怖」「喜」が多数抽出されていることから妥当な結果ではないかと推測した。以上のことから因子3を悲喜劇的物語因子と名付けた。

表 4: 各グループの合計抽出結果

	登場人物	好	怖	驚(昂)	厭(恥, 諦)	怒	喜	哀(無)	安	合計
第1グループ	主人公 A	3	2	15	7	2	16	2	1	48
	主人公 B	3	0	16	3	2	9	1	1	35
	サブ	4	2	16	7	1	22	5	1	58
第2グループ	主人公 A	5	2	35	12	3	11	0	2	70
	主人公 B	0	1	24	6	3	8	0	0	42
	サブ	1	3	38	9	0	26	2	0	79
第3グループ	主人公 A	0	0	0	4	0	2	0	0	6
	主人公 B	0	0	6	3	0	1	0	0	10
	サブ	0	0	11	4	2	7	0	2	26

表 5: 各グループの抽出割合

	登場人物	好	怖	驚(昂)	厭(恥, 諦)	怒	喜	哀(無)	安
第1グループ	主人公 A	0.063	0.042	0.313	0.146	0.042	0.333	0.042	0.021
	主人公 B	0.086	0	0.457	0.086	0.057	0.257	0.029	0.029
	サブ	0.069	0.034	0.276	0.121	0.017	0.379	0.086	0.017
第2グループ	主人公 A	0.071	0.029	0.5	0.171	0.043	0.157	0	0.029
	主人公 B	0	0.024	0.571	0.143	0.071	0.19	0	0
	サブ	0.013	0.038	0.481	0.114	0	0.329	0.025	0
第3グループ	主人公 A	0	0	0	0.667	0	0.333	0	0
	主人公 B	0	0	0.6	0.3	0	0.1	0	0
	サブ	0	0	0.423	0.154	0.077	0.269	0	0.077

6.2 各因子の登場人物ごとの感情についての考察

第1グループにおいて、サブキャラクターの「喜」の抽出回数が多く、登場人物3名の「驚」の抽出割合が高いという特徴が挙げられる。6.1節では、第1グループの物語にはシニカルな雰囲気演出している作品が多く、「喜」の感情語のような明るい、幸せそうといったイメージを持つ言葉を本文中で多用することによりシニカルな雰囲気を表現しているという推測を述べた。サブキャラクターの「喜」の抽出回数が多いという抽出結果は、その推測を補強するような形になったのではないかと考えられる。また、第1グループの物語には主人公らと訪れた国の国民の間で価値観の違いが生まれている話が多いと述べたが、登場人物3名の「驚」の抽出割合が高いという抽出結果においても登場人物らがお互いの価値観が違うことに驚いているため、登場人物3名の「驚」の抽出割合が高くなったと考えられる。

第2グループにおいて、登場人物3名の「驚」の

抽出回数および抽出割合が第1グループ以上に高いという特徴が挙げられる。6.1節では、第2グループの物語には登場人物への理不尽さや、読者への意外性を演出している作品が多いと述べた。本結果はその考察を裏付ける形になったのではないかと考えられる。

第3グループにおいて、主人公2名の感情抽出回数が他の2つのグループと比較して少なくなっていることから、作者は主人公らの感情を極力表現しないことで、物語における傍観者のような役割を主人公らに与えているのではないかと考えられる。また、サブキャラクターの「驚」の抽出割合が高いという特徴が挙げられる。6.1節で、第3グループの物語には幸福な明るい情景から不幸な暗い情景に一転する作品やその反対の作品が多いと述べた。幸せから不幸もしくは不幸から幸せの描写へ一転する際にサブキャラクターの「驚」の感情が多く抽出されたのではないかと考えられる。

7 今後の課題

本研究では、物語から登場人物の感情状態を抽出するため、12種類の感情に分類した感情語の辞書を構築した。そして、構築した辞書を用いて「キノの旅」という物語から感情語を抽出し、因子分析を用いて感情の観点から複数の因子への分類に成功した。その結果、皮肉な物語の因子・意外性が強い物語の因子・悲喜劇的な物語の因子の3つに分類された。そして、各因子の物語から登場人物の感情を抽出した結果とも整合性があった。

今後の課題として、まず作成した辞書の客観性を担保する必要があることが挙げられる。感情の分類手法に関しては中村 [9] の手法を参考に行っているが、本研究における実際の割り当ては一人の分析者で行った。そのため、現状では作成した辞書の感情語の分類は主観的であると言わざるを得ない。したがって評価実験により複数人による一致度の検証を行う予定である。

次に、本研究では、因子分析を行う際に「恥」「諦」「昂」「無」の感情状態を他の感情と結合して分析を行った。これらの感情の抽出回数が他の感情と比較して少なかった原因として、辞書の語彙数が不足していたという可能性が挙げられる。今後、辞書の語彙を増加させることで、より精密な感情抽出をすることが可能であると考えられる。

また、本研究では分析対象およびデータソースとして「キノの旅」というライトノベル小説の1タイトルを選択した。因子分析を行うことで3つの因子に分類可能であったが、今後異なる作者やジャンルの物語を対象に分析を行いデータ量を増加させることで、汎用的な物語の感情分類が可能であると考えられる。

最後に、本研究では登場人物ごとの感情を抽出する際に構文解析器 KNP [13] を用いたが、現状では日本語の構文解析は精度が高いとは言えない。精度の問題が解決された場合は、登場人物の感情抽出の精度の向上が見込められる。

参考文献

- [1] 馬場こづえ, 藤井敦: 「小説テキストを対象とした人物情報の抽出と体系化」, 言語処理学会第13回年次大会発表論文集, pp. 574- 577, 2007.
- [2] Mingming, L. et al.: "Literary intelligence analysis of novel protagonists' personality traits and development", Digital Scholarship in the Humanities, Vol. 34, No. 1, pp. 221-229, 2018.
- [3] 高村大也; 乾考司; 奥村学: 「スピンモデルによる単語の感情極性抽出」, 情報処理学会論文誌, Vol. 47, No. 2, pp. 627- 637, 2006.
- [4] 佐藤貴俊; 高村大也; 奥村学: 「分散表現を用いた単語の感情極性抽出」, 情報処理学会研究報告, Vol. 2016-NL-228, No. 12, pp. 1- 6, 2016.
- [5] Ekman, P.; Friesen, W.V.: Constants across cultures in the face and emotion, Journal of Personality and Social Psychology, Vol. 17, No. 2, pp. 124-129, 1971.
- [6] Plutchik, R.: The Multifactor-Analytic Theory of Emotion, Psychology, Vol. 50, pp. 153-171, 1960.
- [7] Russell, J.: A circumplex model of affect, Journal of Personality and Social Psychology, Vol. 39, No. 6, pp. 1161-1178, 1980.
- [8] 時雨沢恵一: 「キノの旅 -the Beautiful World-」第一巻, 電撃文庫, 2000.
- [9] 中村明: 「感情表現辞典」, 東京堂出版, 1993.
- [10] 小林のぞみ; 乾健太郎; 松本裕治: 「意見情報の抽出/構造化のタスク仕様に関する考察」, 情報処理学会研究報告, Vol. 2006, No. 1, pp. 111- 118, 2006.
- [11] 国立国語研究所: 「分類語彙表増補改訂版データベース (ver.1.0)」, 2004.
- [12] Kudo, Taku: "MeCab: Yet Another Part-of-Speech and Morphological Analyzer". <https://taku910.github.io/mecab/> (2019年10月23日参照)
- [13] "KNP". <http://nlp.ist.i.kyoto-u.ac.jp/index.php?KNP> (2019年10月23日参照)