

絵本における登場キャラクターの行動に着目した感情極性評価 Evaluating Emotion polarity focusing on the character 's behavior in a picture book

伊良原 諒^{1*} 松下 光範¹
Ryo Irahara¹ Mitsunori Matsushita¹

¹ 関西大学総合情報学部
Faculty of Informatics, Kansai University

Abstract: The aim of the research is to develop a search system of picture books that have close transition patterns in terms of the emotional polarity received from the story. Our method calculates and visualizes the transition of the emotional polarity for each character appeared in the story to make the searcher grasp them. Currently, picture book search systems have been developed that use not only the bibliographic information of the book but the content information such as texts and images. However, the transition pattern of the emotional polarities which calculated automatically may differ from the impression the reader received from the story. This is because various characters' actions are mixed in the text of a picture book. To solve the problem, we focus on the action sequence of each character and calculates the transition of the emotional polarity based on the sequences character by character. To verify the validity of the method, we conducted an experiment to evaluate the accuracy of extracted verbs and calculated emotional polarity. Our experiment so far revealed the necessities of estimation of a speaker of each utterance and adjustment of the values of the used emotional polarity dictionary.

1 はじめに

現在、絵本は膨大な数が出版されており、全国出版協会・出版科学研究所の2016年版出版指標年報によると、1年あたりの新刊絵本の冊数は1500冊を超えている[8]。そのため、非常に多くの絵本の中から購入する絵本を選択することが難しい。

現在、絵本選択の手段として、絵本情報サイト「絵本ナビ¹」のように、タイトルや著者名、出版社などの書誌情報を基にした検索や全ページの試し読みなど、絵本のレビューの投稿や閲覧などが利用可能である。

絵本の購入目的には、情操教育のための絵本の読み聞かせ²や、絵本から癒しや心理的变化を受ける[7]ための購入などが挙げられる。そのため、絵本の選択の意図は多岐に渡る。これらの目的に応じた検索を実現するため、絵本の内容情報を基にした検索手法や子供の興味・関心を基にした検索手法など様々な検索の手法が必要であるが、現在利用される絵本検索サイトで

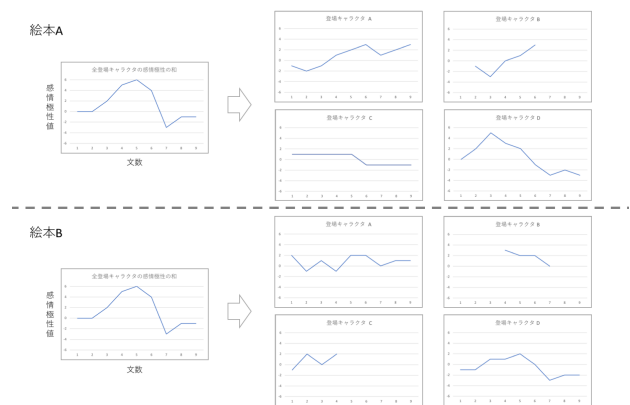


図 1: ストーリーとキャラクターごとの感情極性の例

はこのような検索手法は実現されていない。この問題に対して、意図に沿った絵本を発見するために様々な絵本検索手法が提案されている。

絵本に出現する単語の感情極性の推移を算出することで絵本同士の類似度をはかり、類似絵本を検索を行う手法を検討が行われている。しかしストーリーには複数のキャラクターの感情極性が混在するため、出現するすべての単語から感情極性の推移を算出すると、読み

*連絡先：関西大学総合情報学部
〒569-1095 大阪府高槻市豊山寺町 2-1-1
E-mail: mat@res.kutc.kansai-u.ac.jp

¹絵本ナビ <https://www.ehonnabi.net> (2019/2/14 確認)

²文部科学省 http://www.mext.go.jp/a_menu/shotou/youji/asobou/try/yomikikase.htm (2019/2/14 確認)

手がストーリーから受ける感情極性の推移と異なる結果になることがある。例えば、図1では絵本A,絵本Bのそれぞれにおいて、左側には絵本における全てのキャラクタの感情極性が混合したグラフを示し、右側には絵本における登場キャラクタごとに感情極性を分類したグラフを示している。全てのキャラクタの感情極性を混合したグラフを比較すると絵本Aと絵本Bのグラフは形状が類似していることが読み取れるが、登場キャラクタごとに感情極性を分類したグラフを比較すると、キャラクタごとの感情極性のグラフはそれぞれ異なる形状であることが読み取れる。このようにキャラクタの感情極性が混合している場合、各キャラクタの感情極性が類似していなくても、ストーリー展開が類似していると判定される問題が考えられる。

本研究では各キャラクタに着目したストーリー展開を基に絵本の検索を目指す。これを実現することで、より詳細な絵本の類似度を判定することが可能になる。さらに映画や小説などにおいて、人はストーリーに登場するキャラクタに感情移入することが明らかになっていること [1] から、登場キャラクタごとの感情の流れを把握することで、ユーザがストーリーから受け取る感情極性の推移に近い絵本検索も行えると考えられる。

そこで、本稿ではストーリーを各登場キャラクタの行動に関する記述の連続として表現されるものと仮定し、行動を表す動詞をキャラクタごとに分類して感情極性を付与することでストーリー展開の把握を試みる。

2 関連研究

藤田らは、子どもの興味に合う内容および子どもの年齢に応じた読みやすさの絵本を提案するため、絵本の内容情報を基にした絵本検索システム「ぴたりえ」を提案している [9]。子供の興味に合う内容の絵本を提案するため、グラフ構造を索引とした高速汎用検索アルゴリズムを実装し、特徴量による絵本の検索を可能にしている。これによって検索に厳密に一致する絵本を提案するだけでなく、内容語や書誌情報など様々な要因を用いた検索を実現している。今後の展望として、ハッピーエンドやバッドエンドのようなストーリー展開をもとにした検索や、読み聞かせにかかる時間をもとにした検索の必要性が述べられている。

安尾らは、絵本の中に出現する単語の感情極性に着目し、その推移を検索に用いることで類似した絵本を検索可能にする手法を提案している [11]。テキスト中に出現する感情極性が類似している絵本同士はストーリー展開が類似しているという仮定のもと、ページ単位で感情極性を計数している。しかし、全てのキャラクタの感情極性が混合していることから、立場が対立するキャラクタが存在する場合に両方の極性を取得する

ことが起こり、ストーリーの感情極性評価に影響を与えると考えられる。

絵本検索に関する研究では、テキストデータや書誌情報などを用いて、単語の難易度やストーリー展開に着目した絵本の検索方法が提案されている。しかし、登場キャラクタごとの感情極性を基にした絵本の検索手法は実現されていない。

ストーリー分析に関する研究として、Reaganらは、小説における感情極性の変化は、大きく分けて6種類に分類できることを明らかにした [3]。この研究では感情極性の変化を把握するため、感情極性の時系列変化を示す曲線であるエモショナルアークに着目し、ストーリーの分析を行った。

吉田らは、小説全体の雰囲気を感じの観点から捉えた選別を行うため、感情を用いて小説の俯瞰分析を行っている [6]。文章中に現れる感情表現に対して、感情の極性や感情の種類を判断する辞書を使用して文書の感情の変遷を取得している。しかし、抽出された感情表現が誰の感情表現であるか判断できないといった問題や、感情表現辞典においてどの感情の種類に分類するか不明瞭な語句が存在するという問題がある。

以上のようにストーリー分析は文章に含まれる単語の感情極性や感情の種類を評価し、それらの時系列変化を把握することで行われている。これらの手法では、ストーリー全体を俯瞰的に把握することができる。しかし、登場キャラクタごとにストーリーを分析する研究は行われていない。

レビュー文やSNSの投稿、発話などからユーザの感情抽出を行う研究が多く存在する。その際、感情を評価する基準として Positive や Negative 等の極性や R.Plutchik の基本感情モデル [2] のように複数の感情に分類を行い、感情を抽出するものがある。

熊本らは、新聞記事を読んだ人々が感じる印象の記事から抽出するテキスト印象マイニング手法を提案した [10]。印象語辞書を構築し、記事の印象値を算出した。被験者実験を行い、それぞれの印象尺度における印象値を人手で付与した印象値と比較し、それらの平均誤差から提案手法の精度を評価した。

これらのように感情抽出に関する研究では、感情極性辞書や感情表現辞典など、様々な感情モデルを用い、テキストからの感情抽出が試みられている。本稿では、これらの手法から本研究に適した感情抽出手法の選定を行う。

3 提案手法

絵本テキストデータからの、登場キャラクタおよび動詞の抽出・感情極性付与を行う。その手順を図2に示す。

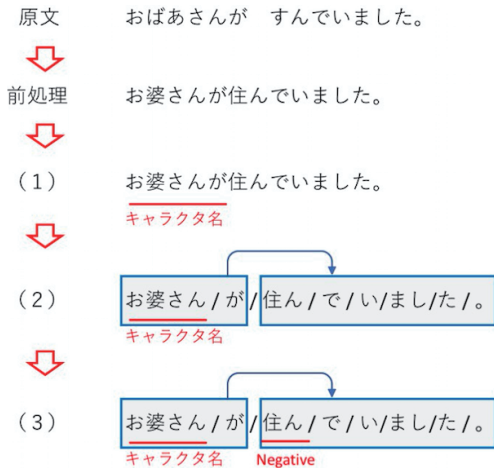


図 2: 提案手法における処理の例

テキストデータの前処理として、鉤括弧で囲まれた文(以下、台詞文と示す)を1つの文とし、その他の文は“。”を基準に1文とする区切りを定めた。例外として、鉤括弧の中に3文以上が含まれている場合、係り受け解析に失敗することがあるため、“|”の中身を“。”で区切り、それぞれを1文とした。また、“～と、赤頭巾は言いました”のように台詞文と結合している後ろに現れる表現は、1文とした。

登場キャラクタとする名詞の選定条件として、(1) 絵本テキストの文において主格となる、(2) 鉤括弧で括られた台詞の発話者となる、を共に満たす名詞とした。条件(2)を選定条件に設定した理由として、例えば、「雨が降る。」という文があった場合、条件(1)のみでは文中の主格となる語が“雨”であるため、登場キャラクタでないものを登場キャラクタとして選定してしまう場合が考えられるためである。

ここで、“一/匹/目/の/子豚”のように複数の形態素で表される登場キャラクタが存在する。これを防ぐため、ユーザ辞書に一意的な登場キャラクタ(以下、一意キャラクタ)を全て登録した。さらに、“三匹の子豚”のような、複数の登場キャラクタをまとめて指す単語(以下、複合キャラクタ)は、ユーザ辞書に登録し、それらが指す各々のキャラクタと動詞を人手で対応づけた。

また同じ表現でも、異なるキャラクタを指している場合がある。例えば、「子豚が食べられました。」、「次の子豚が現れる。」、「子豚が食べられる。」という表現の場合、“子豚”という表現が何匹目の子豚を指示しているのか機械的に判断することが困難であるため、人手で判断し対応付けを行い、係り受け解析を行った。

係り受けから判定した動詞の感情極性を評価するための辞書を、日本語極性辞書(用言編)[5]と、PNtable[4]の中から選定する。選定の条件として、登場キャラクタに係る動詞の感情極性を取得することが必要である

(例) そこへ、一匹目の子豚が逃げてきました。



図 3: 正解データの作成例

ため、登録されている動詞の数が多い感情極性辞書が適切だと考え、日本語極性辞書(用言編)とPNtableの動詞登録数の比較を行った。PNtableに含まれる動詞の数は4140語であり、感情極性辞書(用言編)は638語であった。動詞の登録数の多いPNtableを使用する。

PNtableでは、Negativeとして登録されている語数がPositiveとして登録されている語数に対して多いため、PositiveとNegativeの語数が均等に近づくよう、全ての単語の感情極性値の平均が0、分散が1となるz-scoreによる正規化を行い感情極性値として再付与した辞書を作成した。作成した辞書と係り受けで抽出した動詞を用いて感情極性を判定する。

4 実験

提案手法での動詞抽出精度と感情極性付与精度を評価するため、提案手法から得られたデータと、人手で作成した正解データを比較する実験を行った。

登場キャラクタごとに感情極性を付与するためには、登場キャラクタとそれに係る動詞が紐づくことが必要である。そのため、登場キャラクタを主語とする動詞の抽出精度を評価する実験(i)を行う。また、PNtableは単語の感情極性を評価するものであるが、それが絵本の登場キャラクタの感情極性と一致するかを検証する必要がある。そのため、抽出された動詞の感情極性の精度を評価する実験(ii)を行う。実験に使用した絵本と、それぞれの作品に付与したタグ(識別名)を表??に示す。以下ではそれぞれの作品を示す場合、タグによって表記する。

4.1 正解データの作成

実験に用いる正解データの作成について記述する。提案手法での登場キャラクタを主語とする動詞の抽出と抽出した動詞の感情極性が正しく取れているかを判定するために人手でこれを評価し正解データの作成を行った。評価者は、20～23歳の大学生男女9名であっ

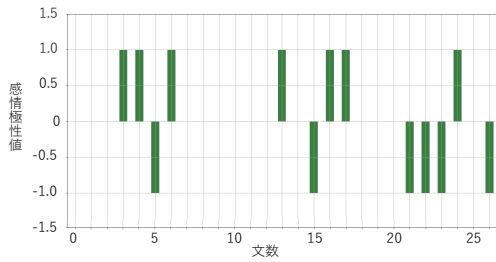


図 4: 感情極性のみの正解データのグラフ

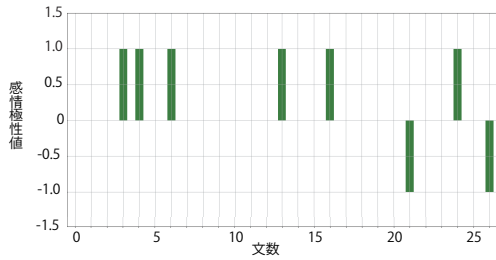


図 5: 本研究で用いる正解データのグラフ

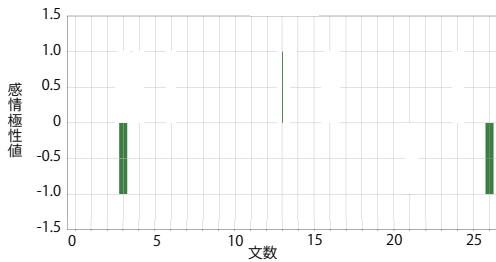


図 6: 提案手法データのグラフ

た．1冊の絵本に対して3人の評価者が評価し，そのうち，2人以上の評価が一致したものを正解データとして採用した．前処理後の絵本テキストデータの各文に対して，動詞部分の抽出と登場キャラクターごとの文の Positive か Negative の感情極性の判定したものを評価とする．動詞部分の抽出では，絵本テキストデータの各文の動詞部分に評価者が線を引き，その行動の主語となる登場キャラクターを記すタスクを課した．感情極性の判定では，評価者がキャラクターにとって動詞が Positive であるか Negative であるかを判定し記述するタスクを課した．正解データの作成例を図3に示す．感情極性を評価するタスクから得られた感情極性値をグラフ化し，その一部を図4に示す．

得られたデータを精査したところ，他のキャラクターの動作が感情極性値に影響を与えている場合があった．そこで本実験では，提案手法がキャラクターを主体とする動作による感情極性の変化を取得できるかを検証する．そのため，人手によって取得した感情極性値のうち，キャラクター自身の動作が起因する変化のみを正解

表 1: 実験 (i) における動詞の抽出率

	Precision	Recall
実験条件 a	0.84	0.31
実験条件 b	0.85	0.38

データとした．その正解データをグラフ化した一部を図5に示す．この処理を1冊分の絵本テキストデータに対して行う．

4.2 動詞抽出精度の比較実験

実験 (i) では，提案手法を用いて抽出した動詞と正解データに含まれる動詞の一致率を求めた．ここで，複合キャラクターのストーリー展開に対する影響を確認するため，一意キャラクターのみを係り受け解析に用いる手法を実験条件 a とし，一意キャラクターと複合キャラクターを用いる手法を実験条件 b として各々の条件下での抽出精度を比較した．実験 (i) における動詞の抽出率を表1に示す．

動詞抽出数を比較した結果，実験条件 b の方が多く抽出を行えた．これは，複合キャラクターの表現が数多くテキスト内に含まれるためであると考えられる．複合キャラクターが登場する作品では，実験条件 b 用いることでより正確に動詞を抽出できることがわかった．

また感情極性の推移を正確に把握するためには，感情極性の付与対象となる動詞を多く抽出することが望ましいが，動詞の再現率は実験条件 a では31%，実験条件 b では38%であり，感情極性の推移の把握に十分とは言い難い．これは，絵本に多く含まれている鉤括弧で括られた台詞文には主語が記述されない場合が多く，台詞文中の動詞と登場キャラクターとを紐づけることができなかつたためだと推察される．

4.3 感情極性付与精度の比較実験

実験 (ii) では，実験 (i) の実験条件 b で正解データと一致した動詞のみを対象に，動詞の感情極性と正解データを比較し，その一致率を評価した．提案手法から得られたデータをグラフ化した一部は図6のようになった．対象となる動詞の表記を PTable に登録されている表記と一致させる処理を行った．その結果，対象とした動詞の41%は PTable に登録されていない動詞であったため実験対象から除外した．実験 (ii) における感情極性の一致率を表2に示す．実験結果から，PTable に登録されている感情極性の多くは正解データの感情極性と異なることが明らかになった．特に正解データ

表 2: 実験 (ii) における感情極性の一致率

		提案手法	
		Positive	Negative
人 手	Positive	0.07	0.41
	Negative	0.01	0.13
	Even	0.04	0.32

では Positive であるが、提案手法では Negative である場合が多くみられた。

5 発話者の推定

本稿では、テキストから動詞を抽出する際、登場キャラクターがその動詞の主語であることを動詞の抽出条件とした。しかし、実験において鉤括弧で括られた台詞文中では、主語が省略されることが多いため、動詞の抽出制度が低くなることが明らかになった。そのため、台詞文において主語が省略される場合、それを補うことにより動詞の抽出精度を向上させることができると考え、これを実装した。

台詞文における主語を推定するにあたり、平常文の台詞で省略される主語は発話者であることが多いことに注目する。絵本において“(主語)が言いました。”という表現では、この文の「言う」の前後に現れる台詞文の発話者は“(主語)”であると推定できる。また、“(主語)は～と言いました。”という表現では、“～”の主体は“(主語)”であると推定できる。このことから、「言う」という表現に着目し、発話者の推定を試みる。これを実現するため、本節では、台詞文の発話者推定の手法を考察し、動詞の抽出制度の向上を図る。

実装の条件として、1. テキストの地の文において「言う」とその主語を抽出できたとき、その文の前後に台詞文があること、2. 「言う」の直前に「そう」が含まれること、3. 1. で抽出した主語が台詞文中に含まれないことの3点を定めた。

条件 1. は、上に述べた着目点に沿って定めた。条件 2. は“桃太郎はそう言いました。”という文において「そう」は前の台詞文を指すため、台詞中の“(主語)”から前の台詞文の発話者の推定が可能であると考え定めた。条件 3. は、発話者自身の名前が台詞中に含まれることは自己紹介をする際など限定的な場面であり、頻度が少ないと考え定めた。

そのため、条件 1. から 3. の全てに当てはまる場合は「言う」の主語を、その前の台詞文の発話者とし、条件 1. と 3. に当てはまる場合は「言う」の主語を、前後の台詞文のうち、発話者の名前を含まない台詞文の発話者とした。

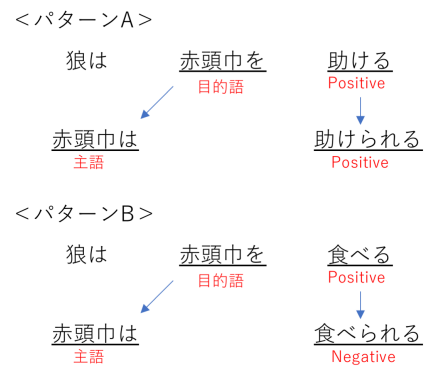


図 7: 目的語に対する感情極性付与のパターン

これらの条件に沿って、台詞文の発話者の推定を行い、人手による発話者の推定結果と比較した。いくつかの文では、人手による判断と一致する発話者を推定することができ、発話者推定が動詞の抽出精度の向上に繋がること示唆された。しかし、複数の台詞が連続した場合や一つの台詞文に複数の推定結果が重複した場合に正しく推定できないことがわかった。今後は発話が行われている場面に、どのキャラクターが存在するのかに着目することで本稿で推定できなかった発話者の推定が可能になると考えられる。これらの処理を行うことで、感情極性の精度が向上すると考えられる。

6 今後の展望

6.1 キャラクターの関係性

図 4 に示した感情極性のグラフは、ぶた 1 に登場する狼の感情極性を表したものである。このグラフは、読み手が感じるストーリー中のキャラクターの感情極性を最も反映していると考えられる。これは他の登場キャラクターの動作が、そのキャラクターの感情極性に影響を与えることを加味した評価を行ったためである。しかし、正解データと提案手法から得られたデータはそのキャラクターのみの動作から感情極性を評価したものであった。今後、キャラクターの関係性を考慮し組み合わせることで、図 4 のようなグラフに近づくと考える。そのため、提案手法においても、キャラクターごとの関係性を考慮する必要がある。

6.2 感情極性の評価

実験から、登録されている単語や感情極性値は人手の判断と異なることが示された。辞書に登録されていない単語があることや登録されている感情極性値が異

なることが原因であると考えられる。そのため、辞書に登録されていない単語の登録や、人手による判断と異なる感情極性の単語の特徴を調査し辞書の感情極性の変更を行う必要がある。それに加えて本節では(1) 目的語のキャラクタ、(2) 動詞以外の品詞に着目して、感情極性の評価の向上が期待される処理を述べる。

(1) に関して、提案手法では、登場キャラクタが主語である場合に紐づく動詞に対して感情極性の付与を行った。しかし、正解データを分析した結果、目的語にある登場キャラクタも感情極性付与の対象となっていた。そのため、登場キャラクタが目的語として出現する場合に対しても感情極性の評価を行うことで精度が向上すると期待される。図7は、動詞に紐づく目的語を主語とし、動詞の部分を受け身の表現にする処理を示している。パターンAは、主語の登場キャラクタと目的語の登場キャラクタに係る動詞の感情極性が同じである。パターンBは、主語の登場キャラクタと目的語の登場キャラクタに係る動詞の感情極性が異なる場合である。このように目的語のキャラクタを主語に置き換えた場合、動詞は受け身の表現になると考えることができる。パターンA、パターンBそれぞれに対応する受け身の表現の単語を辞書に登録する。それにより登場キャラクタが目的語として出現する場合に紐づく動詞に対して、感情極性の付与が行えるようになり、感情極性評価の精度の向上が見込める。

(2) に関して、提案手法では動詞から感情極性を評価していたが、動詞の前の記述によって動詞単独の感情極性評価とは異なる評価がされるものがある。例えば、“思う”という単語は、動詞単独の感情極性評価ではNegativeと判断されているが、“思う”の前に“嬉しく”が記述されていた場合、Positiveと評価されることが考えられる。このような語に考慮した感情極性評価を行うためには、動詞以外の品詞と組み合わせることで人手の判断に近い感情極性が取得できると考えられる。

7 おわりに

本稿では、登場キャラクタごとの行動に着目した感情極性の評価手法を検討した。この手法の有用性を検証するために、提案手法による動詞抽出の精度と、付与される感情極性の精度の評価を行った。今後、感情極性辞書にない単語の登録を行うこと、及び人手による判断と異なる感情極性の単語の特徴を調査し辞書の感情極性の変更を行うことにより、感情極性評価の精度向上に取り組む。また今回の手法では、登場キャラクタが主格である場合のみを感情極性の評価対象としたが、登場キャラクタが目的格である場合も考慮した手法を検討し、読み手の受ける感情極性により近いストーリー把握を目指す。

参考文献

- [1] Davis, M. H.: A Multidimensional Approach to Individual Difference in Empathy, *JSAS Catalog of Selected Documents in Psychology*, Vol. 10, p. 85 (1980).
- [2] Plutchik, R.: *The Emotions*, University Press of America (1991).
- [3] Reagan, A. J., Mitchell, L., Kiley, D., Danforth, C. M. and Dodds, P. S.: The emotional arcs of stories are dominated by six basic shapes, *EPJ Data Science*, Vol. 5, No. 1, p. 31 (2016).
- [4] Takamura, H., Inui, T. and Okumura, M.: Extracting Semantic Orientations of Words Using Spin Model, *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, ACL '05, Stroudsburg, PA, USA, Association for Computational Linguistics, pp. 133–140 (2005).
- [5] 小林的ぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一: 意見抽出のための評価表現の収集, *自然言語処理*, Vol. 12, No. 3, pp. 203–222 (2005).
- [6] 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, 第73回全国大会講演論文集, Vol. 2011, No. 1, pp. 387–388 (2011).
- [7] 笹倉剛: 絵本を中心としたビブリオセラピーに関する研究, *神戸親和女子大学言語文化研究*, Vol. 9, pp. 21–39 (2015).
- [8] 全国出版協会・出版科学研究所: 2016年版出版指標年報 (2016).
- [9] 藤田早苗, 服部正嗣, 小林哲生, 奥村優子, 青山一生: 絵本検索システム「ぴたりえ」~子どもにぴったりの絵本を見つけます~, *自然言語処理*, Vol. 24, No. 1, pp. 49–73 (2017).
- [10] 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, 情報・システム*, Vol. 94, No. 3, pp. 540–548 (2011).
- [11] 安尾萌, 服部正嗣, 藤田早苗, 松下光範: 物語の類型に着目した絵本の類似探索手法に関する一検討 (ヒューマンコミュニケーション基礎), *電子情報通信学会技術研究報告*, Vol. 116, No. 436, pp. 103–108 (2017).