

絵本のレビュー解析と画像特徴量による直感に沿った画像評価

藤田 早苗 木村 昭悟 服部 正嗣 小林 哲生 奥村 優子

NTT コミュニケーション科学基礎研究所

{fujita.sanae,kimura.akisato,hattori.takashi,kobayashi.tessei,okumura.yuko}@lab.ntt.co.jp

1 はじめに

絵本やイラストを探するとき、人はどのように検索するのだろうか。好きな画家やイラストレーターから検索する場合もあるだろうが、「優しい絵」や「迫力のある絵」といった直感的な評価に合う絵本やイラストを探したい場合も多いだろう。

近年、画像研究では、写真の感情極性（ポジティブ、ネガティブ）を推定したり [1, 5]、人物の笑顔度や風景の natural 度を推定する [3] など、より抽象的な特徴を推定する研究が盛んになってきている。しかし、「優しい」「迫力がある」のような、より人間の直感的な評価に近い特徴推定を目指した場合、学習データの獲得は容易ではない。そこで、本稿では、レビューから絵に関する評価表現を収集し、各評価表現を推定するための学習データとクラスを自動獲得して人間の直感的評価に近い画像評価を行う方法を提案する。

ツイッターやブログ、レビューを用いた評判分析の研究は多くなされている [4, 6, 2, 7]。これらの研究は、評判が記述されているものについて、記述内容に基づいた感情極性の推定 [2, 6]、記述内容からの評価やクレーム情報の抽出・要約 [4, 7] を行っている。本稿では、レビューから評価表現を抽出するだけでなく、その評価表現を推定するために適切な学習データを自動的に収集し、画像特徴量と組み合わせることで、その評価に関する記載がない画像についても評価を推定することを目指す。

つまり、本提案手法は、画像処理研究の視点からみると、ある評価軸に沿った学習データを自動獲得すること、より人間の直感的な評価に近い推定を行うこと、イラストを主な対象とする点でもチャレンジングである。また、評判分析の視点からみると、評判の記述がある画像だけでなく、評判の記述がない画像に対しても画像特徴量から評価を推定する点でチャレンジングである。

提案手法の概要は次の通りである。まずレビューから、絵に関する評価表現を自動抽出する (Step 1, 2 章)。次に評価表現の共起情報に基づき、評価表現ごとに学

習データとクラスを自動的に得る (Step 2, 3 章)。さらに、各評価表現に対する一致度を測るためのランキング学習を画像特徴量を用いて行う (Step 3, 4 章)。最後に、評価が未知の画像に対しても、各評価表現に対する一致度を推定する (Step 4, 5 章)。以降、各章で提案手法の各ステップと評価について述べる。

2 Step 1: 評価表現の抽出

実験対象 実験には、絵本のレビューと表紙の画像データを利用する。絵本の場合、他の一般書とは異なり、絵の比重が大きいため、レビュー中に絵に関する記述が比較的多いと期待できる。また、表紙画像も通販サイトから得やすいためである。絵本に関するレビューと表紙画像を Web から収集した結果、レビューを獲得できた絵本が 17,521 冊¹、そのうち表紙画像が獲得できた絵本は 15,484 冊² だった。

レビューからの評価表現抽出方法 絵に関する評価表現の抽出方法は次の通りである。まず、全レビューを UniDic 短・長単位によって形態素解析する [8, 9]。次に、短単位が絵に関するキーワード³と一致する場合、同一文中のキーワードの前後 4 語内に含まれる内容を長単位の基本形で、キーワードと組にして抽出する。以降、キーワードと内容語の組を評価表現と呼ぶ。ただし、内容を表わすキーワード⁴が出現する場合、それ以降の語は利用しない。また、これらの内容語の後ろに否定表現が出現する場合、NOT+を付与する。

例えば文 (1) では、下線部 をキーワードとして、「絵=決して」「絵=NOT+可愛らしい」「絵=思う」「色=毒々しい」の 4 組が評価表現として抽出される。

- (1) 色 は 毒々しい し、決して かわいらしい 絵 ではない と思いました。

¹2015.11.27 に収集を実行。

²2015.12.14 時点。収集未完了。

³「表紙」「絵」「色」「線」「タッチ」「イラスト」「デザイン」「コントラスト」を利用。ただし、「絵」「色」は、「挿し絵」「色彩」「色合い」など、語の一部にこれらを含む語も対象とした。ただし、「絵本」「色々」「異色」など、そのものの意味ではない語は除いた。

⁴「ストーリー」「御話」「内容」「文」など。

表 1: 評価表現の抽出精度と分析

	No.	%
抽出できた	199	64.0
抽出誤り	1	0.3
抽出できなかった	111	35.7
(4語以上離れている)	68	61.3
内 (複合表現)	20	18.0
訳 (対象語の追加が必要)	18	16.2
(形態素解析誤り)	5	4.5
合計	311	

表 2: 評価表現の抽出結果

	出現回数	異なり	例 (頻度順)
NOT+ を含まない	1,427,464	14,188	絵={可愛い, 素敵, 優しい, 可愛らしい, 奇麗, 楽しい, ...}
NOT+ を含む	16,813	744	絵=NOT+{可愛い, 怖い, 写真, 可愛らしい, 派手, ...}
合計	1,444,277	14,932	

評判分析における評価表現としては形容詞が利用されることが多いが、本稿では全内容語を評価表現として抽出している。これは、形容詞以外でも、組み合わせによっては評価表現として有用な場合があると考えたためである。例えば、動詞である「押さえる」も「色数を押さえた絵」のような表現であれば、評価表現として有用だろう。内容語すべてを利用するため、評価表現としては不適切な表現も多く抽出してしまうが、本稿では調査も兼ね、広く抽出している。レビューが一つ以上獲得できた17,521冊のうち、絵に関する評価表現を抽出できた絵本は13,579冊(77.5%)だった。

評価表現抽出精度 抽出した絵の評価表現候補数が10以上15以下だった絵本から、ランダムに選んだ21冊、877レビューを対象に、絵に関する評価表現を人手抽出した。人手で抽出した評価表現と同じ評価表現を自動的に抽出できたかどうかを表1に示す。

表1から、人手で抽出した評価表現のうち、64.0%は正しく抽出できた。抽出誤りとなった文は「いまどきの可愛い絵、じゃない」であり、「可愛い」にはNOTが付与できたが、「今時」にはNOTが付与できなかったことが原因である。また、35.7%の評価表現は抽出できなかったが、その半分以上は、キーワードから4語以上離れた個所に出現していた。否定のスコープ、4語以上はなれた個所の評価表現の抽出は今後の課題としたい。

評価表現抽出結果と分析 表2に、評価表現の抽出結果を示す。表2から、否定表現と一緒に出てくる評価表現(NOT+)は非常に少ないことがわかる。NOT+を含む評価表現で最も頻度が高いのは「絵=NOT+可愛い」だが、出現回数は156回(116冊)であり、「絵=可愛い」の出現回数9,486回(3,707冊)と比べると2%にも満たない。さらに、否定表現とは全く一緒に出

てこない評価表現も多い。例えば、「絵=優しい」は、3,372回(1,644冊)出現しているが、否定表現と一緒に出現しなかった⁵。

各評価表現に対して学習データを獲得する場合、「絵=優しい」と「絵=NOT+優しい」のようなNOT+を含まない/含む評価表現を対として、それらが付与された絵本をクラス分けすることが考えられる。しかし、前述のように、NOT+を含む評価表現は圧倒的に少なく、全く出現しないことも多い。さらに、「タッチ=優しい」「絵=温かい」のように、「絵=優しい」という評価表現と類似した内容の評価表現も多く存在する。そのため、「絵=優しい」という評価がレビュー中に出てこない絵本がすべて優しい絵ではないとは言えず、「絵=優しい」という評価表現が出てくるかどうかだけでクラス分けすると大量の誤りが含まれる可能性がある。また、人手判断も時間とコストがかかり、容易ではない。そうした問題点を踏まえ、評価表現毎に学習データとクラスを自動的に獲得する方法を次章(Step 2)で提案する。

3 Step 2: 学習データの自動獲得

図1に学習データ獲得方法を示す。ここで、対象とする評価表現と共起する評価表現が付与された絵本は学習データとして利用しない。表現が異なっても内容が類似している評価表現による影響を避けるためである。また、ClassA, ClassBの分類は感情極性とは異なり、反対の特徴で分類することを目指している。例えば図1の例では、ClassAには「絵が細かい」絵本がはいり、ClassBには「タッチが大胆」な絵本が含まれるが、これらはどちらもネガティブではなく、単に評価軸上の遠い位置にあるのだとみなすことができる。

図1の手法により、のべ出現回数が300回以上の評価表現58語を対象として、学習データを自動獲得した⁶。その結果、ほとんどの評価表現では、ClassBもClassAと同程度獲得できた。ただし、非常に出現頻度の高い「絵=可愛い」など4つの表現については共起する評価表現が多く、ClassBのデータ数がClassAの1/10に満たなかった。

4 Step 3: ランキング学習

学習器 本稿ではランキング学習を行う。Step 2では、学習データを2つに分類したが、目指す評価軸は直感的な評価であり、本来ははっきりと2値分類でき

⁵否定表現は想定を覆す場合に多く使われていると考えられる。例えば、「絵=NOT+可愛い」多いのは、絵本が通常「可愛い」と想定されているためだと考えられる。

⁶ただし、絵に関する何等かの評価表現を獲得でき、かつ、表紙画像も獲得できた絵本8,410冊の中から抽出

対象とする評価表現を $expA$ 、その否定形を $expB$ とする。
 (例: $expA$ 「絵=細かい」, $expB$ 「絵=NOT+細かい」)
 同じ絵本のレビューから得られた評価表現同士は共起すると呼ぶ。

Step2-1: 全評価表現が $expA$ 、 $expB$ と共起するか調査。
 $expA$ と共起する評価表現群を A 、
 (例: $A = \{ \text{色遣い=細かい, 絵=描き込む, ...} \}$)
 $expA$ と共起しない評価表現群を B と分類。
 (例: $B = \{ \text{表紙=ポップ-pop, タッチ=touch=大胆, ...} \}$)
 また、 $expB$ と共起し、 A に含まれない評価表現群も B に追加する。

Step2-2: 各絵本を、ClassA、ClassB、その他 (ClassO) に分類。
 ある絵本について、絵に関する評価表現が全部で n 個、そのうち $expA$ が n_{expA} 個、 $expB$ が n_{expB} 個、評価表現群 A と一致する表現が n_A 個、 B と一致する表現が n_B 個あるとすると、分類方法は以下の通りである。

```

if  $n_{expA} > 0$  and  $n_{expB} = 0$  then ClassA
elseif  $n_{expA} > 0$  and  $n_{expB} > 0$  then
    if  $n_{expA} - n_{expB} > th$  then ClassA
    elseif  $n_{expB} - n_{expA} > th$  then ClassB
    else ClassO
elseif  $n_{expA} = 0$  and  $n_{expB} > 0$  then ClassB
elseif  $n_A = 0$  and  $n_B > 0$  then ClassB
else ClassO
    
```

ただし、ClassB の数は、ClassA の数 N を上限とし、ClassB の数が N を越える場合、絵に関する評価表現 n がより多い絵本を優先的に利用し、残りの絵本はすべて ClassO とする。また、 th は本稿では 10 とした。

図 1: 評価表現ごとの学習データ獲得方法

る性質のものではない。将来的には評価表現に対する一致度でクラスを詳細化することを見据え、ランキング問題として捉えるのが良いと考えた。学習器には、 SVM^{RANK7} を利用し、ClassA を評価値 2、ClassB を評価値 1 として学習する。

画像特徴量 利用する画像特徴量と特徴数は表 3 の通りである⁸。表 3 で、カラーヒストグラムからは利用されている色・明度・輝度などの情報、統計量からは利用されている色数など、その他の特徴量は絵の複雑さ等を反映するために利用している。

交差検定 Step 2 で獲得した学習データを用いて 10 分割交差検定を行った。表 4 に、ClassB と ClassA が同数得られた評価表現 (つまり、ベースラインが 50 %) について、精度が高かった上位 10 例と、最も精度が低かった例を示す。

表 4 で最も結果が良かったのは「絵=迫力」の 68.8% である。最も結果が悪かったのは「絵=写真」の 6.9% である。表 2 からわかるように、「絵=NOT+写真」という表現は比較的多く出現しており、これを ClassB と

⁷http://www.cs.cornell.edu/people/tj/svm.light/svm_rank.html

⁸値に非常にばらつきがあることから、特徴量は正規化して与えている。

表 3: 画像特徴量と特徴数

特徴量	数
カラーヒストグラム 3 種 (RGB,HSV,YCrCb)	各 96
各カラーヒストグラムの統計量 (平均, 標準偏差, エントロピー, 非零要素数)	各 4
GIST 特徴量	960
HOG 特徴量	1,764
特徴点の数 (SIFT, ORB, MSER, Harris)	4
合計	3,028

表 4: 10 分割交差検定

評価表現	学習	
	データ数 (*2)	精度 (%)
絵=迫力	317	68.8
色=カラフル-colorful	336	63.5
色彩=鮮やか	144	63.5
絵=インパクト-impact	240	63.1
絵=癒す	212	63.0
タッチ=touch=優しい	382	61.3
絵=懐かしい	135	61.1
絵=御洒落	207	59.7
絵=細かい	474	59.3
絵=はっきり為る	304	59.2
絵=写真	580	6.9

評価表現は、長単位での解析結果の表記をそのまま利用。

して利用している。しかし、「写真ではない」と言及される絵本は、「一見写真のような絵」が多く、画像特徴量での区別が困難だったのだと考えられる。

5 Step 4: 新規画像の評価推定

本稿では、画像特徴量を用いた推定を行っているため、レビューや評価表現が得られなかった画像に対しても、対象評価表現に対する一致度によるランキングができる。このようにして得られたランキングの妥当性を検証するため、前節の交差検定で精度が 50 % 以上となった 34 の評価表現⁹を対象に、ClassO に対して各評価表現に対するランキングを行い、人手によるランキングとの相関係数を調査した。

人手のランキング方法は次の通りである。まず、評価者 7 名に絵本の表紙画像 41 冊分と形容語 30 語¹⁰を提示し、各形容語に対象絵本の表紙がどの程度当てはまるかを 5 段階評価してもらおう。次に、7 名の付与した平均値によって対象絵本をランキングする。このランキング結果と、本稿の実験でのランキング結果を比較する。ただし、もともと本稿の評価用として実施した実験ではないため、本稿での評価表現とは必ずし

⁹タッチ=touch=優しい, 絵={ はっきり, はっきり為る, インパクト-impact, カラフル-colorful, シンプル-simple, リアル-real, 仄々, 優しい, 凄い, 分かる易い, 可愛い, 可愛らしい, 大きい, 奇麗, 小さい, 御洒落, 懐かしい, 楽しい, 温かい, 独特, 癒す, 素敵, 素晴らしい, 細かい, 美しい, 迫力, 雰囲気, 面白い, 鮮やか }, 色={ カラフル-colorful, 奇麗, 鮮やか }, 色彩=鮮やか

¹⁰カラフル, 明るい, 暗い, きれい, 優しい, かわいい, 怖い, 楽しい, 写实的, アニメっぽい, 判別しやすい, 細かい, やわらかい, シンプル, 細い, 美しい, ごちゃごちゃ, 抽象的, 象徴的, 幾何学的, 不思議, 派手・目立つ, 躍動的, 静か, のんびり, レトロ・古い, 現代的, おしゃれ, 非現実的, 日常的, SF 的

表 5: 中程度以上の相関がある組み合わせ ($|\rho| > 0.4$)

形容語	正負	評価表現
明るい	-	雰囲気
暗い	+	雰囲気, 独特
優しい	-	細かい
かわいい	-	細かい
写実的	+	素晴らしい
写実的	-	シンプル-simple, 分かる易い, カラフル-colorful
アニメっぽい	+	分かる易い
アニメっぽい	-	美しい
細かい	+	迫力, 素晴らしい, 温かい
細かい	-	シンプル-simple, 分かる易い, カラフル-colorful はっきり為る, 可愛い, 素敵
シンプル	+	シンプル-simple, 分かる易い, カラフル-colorful
シンプル	-	迫力
ごちゃごちゃ	+	迫力
抽象的	-	迫力, リアル-real
象徴的	+	リアル-real, 迫力
幾何学的	+	色=カラフル-colorful
躍動的	+	鮮やか
のんびり	-	凄い

ただし、評価表現の「絵=」の場合は省略。

も一致していない。

表 5 に、スピアマンの順位相関係数 (ρ) で、中程度以上の相関があるとされる $\rho > 0.4$ となった組み合わせを示す。人手評価で提示した形容語と同じ評価表現は少なかったが、例えば「アニメっぽい」に対して「分かる易い」が正の相関を示すなど、異なる評価表現でも相関する語が多くみられた。また、逆の概念になると考えられる「明るい」と「暗い」、「細かい」と「シンプル」、「抽象的」と「象徴的」等は、同じ評価表現に対し逆の相関を示しており、妥当な結果といえる。

また、「絵=シンプル-simple」は評価語「シンプル」を含んでおり、相関係数も 0.56 と比較的高かった。一方で、「絵=カラフル-colorful」は、人手評価での「カラフル」とほとんど相関がないという結果になった。例えば、「シニガミさん」(作絵 宮西達也, 2010, えほんの杜)の表紙はほとんど黒一色であり、7人中5人が全くカラフルではないと評価した。しかし、はっきりとした色使いで、カラーヒストグラムや明度の差が大きいことなどから本手法ではカラフルだと推定された。実際、中の絵を見るとカラフルでないとは言い難い。レビューでは、絵本の中の絵を踏まえて、カラフルかどうか記載されている事が多いと考えられる。しかし、人手による評価実験では、表紙のみから判断してもらったため、評価が一致しなかったと考えられる。一方で、「シンプル」かどうかは、表紙でも中身でも変わらないため、比較的相関が高くなったと考えられる。

こうした問題を避けるため、表紙に関する評価だけを利用することも考えられるが、キーワードが「表紙」である評価表現は、全体の 8.5% と非常に少ないため、表紙と中身の乖離については今後の課題としたい。

6 まとめと今後の課題

近年、画像処理の研究では、画像に対してより抽象的な特徴を推定する研究が盛んになってきている。本稿では、レビューの解析結果を用いて、「優しい」や「迫力がある」のような、より人間の直感的な評価に近い特徴にそった画像評価を行う方法を提案した。

提案手法では、まず絵本のレビューから、絵に関する評価表現を自動抽出した (Step 1, 2 章)。次に評価表現の共起情報に基づき、評価表現ごとに学習データとクラスを自動的に獲得した (Step 2, 3 章)。さらに、各評価表現に対する一致度を測るためのランキング学習を画像特徴量を用いて行った (Step 3, 4 章)。これにより、評価が未知の画像に対しても、各評価表現に対する一致度を推定した (Step 4, 5 章)。

これは、評価表現や学習データの選択まですべて全自動で行った上で、画像特徴量だけを用いて人間の直感的な評価に沿った推定を行うという、非常にチャレンジングな課題設定である。にも関わらず、「迫力がある」かどうかを 68.8% の精度で推定できるなど、評価表現によっては比較的高い推定精度を得られた。また、人手でランキングした結果とも、多くの語で中程度以上の相関が得られた。

今後は、評価表現の抽出精度の向上と、ラベルの付与方法の改善に取り組みたい。また、提案手法による画像評価の結果を用いて、「優しい絵の絵本を探したい」などの要求にこたえられる検索システムの構築を目指したい。

参考文献

- [1] D. Borth, R. Ji, T. Chen, T. Breuel, and S. Chang. Large-scale visual sentiment ontology and detectors using adjective noun pairs. In *the 21st ACM international conference on Multimedia*, pp. 223–232, 2013.
- [2] T. Nakagawa, K. Inui, and S. Kurohashi. Dependency tree-based sentiment classification using CRFs with hidden variables. In *Human Language Technologies: NAACL-2010*, pp. 786–794, 2010.
- [3] D. Parikh and K. Grauman. Relative Attributes. In *ICCV-2011*, 2011.
- [4] I. Titov and R. McDonald. Joint Model of Text and Aspect Ratings for Sentiment Summarization. *Journal ACL*, 8:308–316, 2008.
- [5] Y. Wang, S. Wang, J. Tang, H. Liu, and B. Li. Unsupervised Sentiment Analysis for Social Media Images. In *IJCAI-2015*.
- [6] 岡野原 大輔, 辻井 潤一. レビューに対する評価指標の自動付与. *自然言語処理*, 14(3):273–295, 2007.
- [7] 乾 孝司, 梅澤 佑介, 山本 幹雄. 評価表現と文脈一貫性を利用した教師データ自動生成によるクレーム検出. *自然言語処理*, 20(5):683–706, 2013.
- [8] 鈴木 潤, Kevin Duh, 永田 昌明. 拡張ラグランジュ緩和を用いた同時自然言語解析法. *NLP-2012*, C5-6.
- [9] 鈴木 潤, 永田 昌明. 動的変化する文章を対象とした自然言語解析に適した解析アルゴリズムの考案. *NLP-2013*, A6-1.