

ニコニコ動画におけるスラングの感情極性辞書構築

小椋 翔太[†] 桂井麻里衣^{††}

〒610-0394 京都府京田辺市多々羅都谷 1-3

†† 同志社大学理工学部 〒610-0394 京都府京田辺市多々羅都谷 1-3

E-mail: $\dagger\{\text{ogura,katsurai}\}@mm.doshisha.ac.jp$

あらまし ニコニコ動画における評判の良い動画やシーンの検索には、動画コメントの感情分析が効果的と考えられる。一般に、テキストの感情分析には、対象ドメインでよく用いられる語を収録した感情極性辞書が必要となる。そこで本稿では、ニコニコ動画における特有の隠語、略語、俗語（スラング）の感情極性辞書の構築手法を提案する。提案手法では、フォーマルなテキストを対象とした既存の感情極性辞書をシードセットとして用いる。具体的には、まず各動画をシーンに分割し、コメント集合の単語と既存の辞書からシーンの感情極性を決定する。このとき、ニコニコ動画では同一シーンを視聴するユーザ間に一体感があると仮定し、多数派の感情極性を優先する。次に、シーン内の単語集合に対し、先に決定した感情極性を伝播させる。このように、シーンを媒介とした感情極性と共起語の関係に着目することで、既存の辞書にない未知語に対して感情極性を割り当てる。本文の最後には、ニコニコ動画のカテゴリ別コメント集合を用いた実験を行い、提案手法の有効性を示す。

キーワード ニコニコ動画, 感情分析, スラング

1. はじめに

Youtube^(注1) やニコニコ動画^(注2), FC2 動画^(注3) などの利用者投稿型動画共有サービスでは, 日々膨大な数の動画が投稿されている. これらのサイトは CGM (Consumer Generated Media) と呼ばれ, 社会に大きな影響を与えるメディアに成長している [1]. 中でもニコニコ動画では動画や楽曲などの創作が活発に行われており, 初音ミクを代表とするニコニコ動画発の文化は全世界で注目を集めている [2, 3].

ユーザは CGM 動画を視聴する際、望む動画内容を表すキーワードをクエリに用いてタイトルやタグから動画を検索する。得られた動画集合から、再生数やマイリスト数（お気に入りに登録したユーザの数）などの指標をもとに視聴動画を決定する [4]。しかし、これらの指標には不正アクセス等の意図的な工作が行われることが多いほか、動画投稿時期に左右される側面もある。一方、ニコニコ動画では図 1 に示すように、ユーザが動画上に自由にコメントを付与できる。一人のユーザからのコメントは一行で表示され、投稿時のシーンに合わせて右から左へ流れる。これらのコメントには、動画に対する感想や賞賛、共感表現などが含まれている。ゆえに、評判の良い動画やシーンを効率的に検索するためには、動画コメントの感情分析が効果的と考えられる。

テキストの感情分析は、機械学習に基づく手法と辞書に基づく手法に大別される。後者の手法では、単語にポジティブまたはネガティブの感情極性を付与した辞書（以降、感情極性辞書）を参照し、テキスト内における各極性の単語の出現割合を

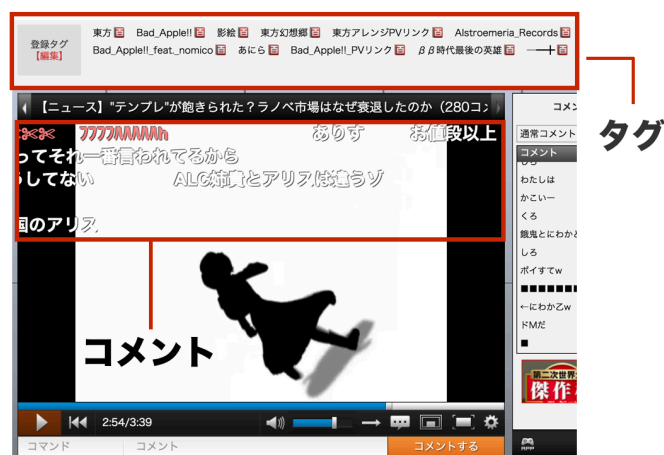


図 1 ニコニコ動画再生画面の例。動画上部にタグが提示され、コメントは投稿時のシーンにあわせて動画上に表示される。一行のテキストが一人のユーザからのコメントに対応する。コメントは右から左に流れる。

算出する。このアプローチでは、分析対象とするドメインの単語が感情極性辞書に収録されている必要がある。これまでに、レビューサイトなどのウェブページにおける頻出語を収録した辞書が構築されてきた [5, 6]。しかしながら、ニコニコ動画には特有の隠語、略語、俗語（スラング）が多く投稿されており、フォーマルなテキストを対象として構築された辞書のみでは網羅的な分析が困難である。

そこで本稿では、ニコニコ動画におけるスラングの感情極性辞書の構築手法を提案する^(注4)。提案手法は、既存の感情極性辞書 [5, 6] をシードセットとして用いる。具体的には、まず各動

(注1) : <https://www.youtube.com/?hl=ja&gl=JP>

(注2) : <http://www.nicovideo.jp/>

(注3) : <http://video.fc2.com/>

(注4)：本稿では単語を一般的な語またはニコニコ動画特有の語へ分類する方法を提案しないため、便宜上、既存の辞書に含まれていない単語をスラングと呼ぶ。

画をシーンに分割し、コメントの単語と既存の辞書からシーンの感情極性を決定する。このとき、ニコニコ動画では同一シーンを視聴するユーザ間に一体感があると仮定し、多数派の感情極性を優先する。次に、シーン内の単語集合に対し、先に決定した極性の情報を伝播させる。このように、シーンを媒介とした感情極性と共起語の関係に着目することで、未知語に対して感情極性を割り当てる。本文の最後には、ニコニコ動画のカテゴリ別コメント集合を用いた実験を行う。実験では、(i) 既存の感情極性辞書のみをシードセットとする場合、(ii) 既存の感情極性辞書を更新し、再びシードセットに用いる場合という二つのアプローチに対し、スラングの感情極性分類の評価結果を報告する。さらに、構築した感情極性辞書に基づく動画視聴ツールを構築し、動画シーンのコメント感情分類への適用例を示す。

本稿の構成は以下の通りである。2章では、ニコニコ動画のテキストに着目した従来研究と、テキストの感情分析に関する従来研究を説明する。3章では、ニコニコ動画におけるスラングの感情極性辞書の構築手法を提案する。4章では提案手法の有効性を確認するための実験を行う。最後に5章で本文をまとめ、今後の方向性について検討する。

2. 関連研究

2.1 ニコニコ動画のタグやコメントに関する研究

ニコニコ動画のタグやコメントの特徴に着目した検索支援技術が盛んに研究されている。例として、特定の感情の遷移を呼び起こすであろう動画をニコニコ動画に投稿し、付与されたコメントを分析した研究 [7] がある。大量の動画とコメントデータの分析例 [8] では、ニコニコ動画特有の「もっと評価されるべき」というフレーズタグの有無を教師とみなし、コメント、タグ、マイリスト数などを特徴として動画の識別器を構築した。文献 [9] では、検索クエリに関連する動画集合から様々な種類のコメント特徴を算出し、人手で評価されたコメント集合を教師データとして有用なコメントの識別器を構築した。このような機械学習アプローチに加え、限られた数の感情語をシードセットとしたコメントの感情分類手法とその応用が提案されている。「面白い」という感情に着目した文献 [10] では、コメント内で笑いを表すスラング“w”の出現回数をカウントし、動画のランク付けに応用した。文献 [11] では、ニコニコ動画で頻出する顔文字に手動で感情ラベルを付与し、動画の評判分析に応用した。文献 [12] では、手動で感情極性辞書を構築し、動画内の感情極性の遷移とコメント数に基づく動画信頼性推定システムを構築した。文献 [13] においても、コメントの感情分類のための前処理として、ニコニコ動画の頻出単語に対し手動で感情極性が付与されている。以上の研究に対し、本研究で構築するニコニコ動画特有の感情極性辞書を導入することで、感情表現のカバレッジや分類精度の向上が期待できる。

2.2 テキストの感情分析に関する研究

テキストの感情分析は、検索・推薦の高度化やイベント検出など様々な応用先があり、メジャーな研究トピックの一つである。多くの手法は単語と感情極性の対応付けを必要とするため、これまでに感情極性辞書の構築方法が検討されてきた。文献 [6]

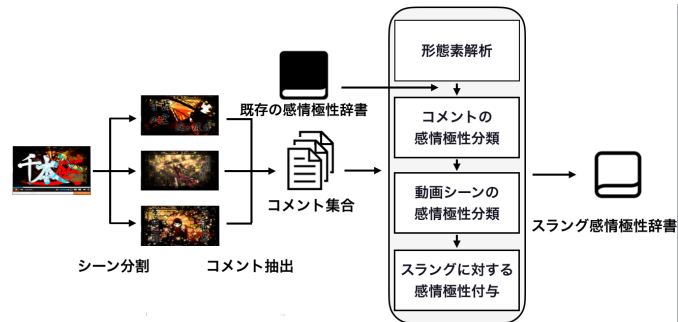


図2 提案手法の概要。

は、対象・属性・評価値の共起パターンから収集された感情表現候補を人手で評価する方法を提案した。文献 [5] では、名詞・格助詞・述語の共起パターンに基づき、予め極性の付与された単語を正解とみなして名詞の極性識別器を構築した。構築後の辞書は人手のチェックを経てオンライン公開されており、本研究のシードセットとして妥当と考えられる。

単語の同義関係や共起関係に基づき既存の感情極性辞書を拡張する技術も種々提案されている。文献 [14] は、Affect database [15] に収録されている 2,438 個の感情語をシードセットとし、外部シソーラスの同義語に感情スコアを付与した。文献 [16] では、文脈によって単語の感情極性が異なる点に着目し、既存の感情辞書やシソーラス、レビューデータ、複数のヒューリスティックを組み合わせることで文脈依存な感情極性辞書を構築した。本研究に最も関連した文献 [17] では、Urban Dictionary (UD)^(注5) に収録されているスラングの感情極性を分類した。具体的には、UD と既存の感情極性辞書の両方に収録されている単語をシードセットとし、UD で定義された同義語に極性を伝搬した。さらに、Twitter^(注6) からツイートを収集し、感情語と頻繁に共起する単語を感情極性辞書に追加した。本研究においても従来研究で有効性が示されてきた語の共起を用いるが、ニコニコ動画のコメント機能が及ぼすユーザの一体感を前提とし、同一シーンを媒介として未知語に感情極性を伝搬させる点が新規といえる。

3. 提案手法

本章では、ニコニコ動画におけるスラングの感情極性辞書構築手法を提案する。提案手法の概要を図2に示す。動画は複数のシーンから構成されており、必ずしも全てのシーンに同一の感情極性が当てはまるとは限らない。そこで提案手法では、動画カテゴリごとにシーンの時間幅と分割最大数を設定し、予め各動画をシーン分割する。分割最大数の設定には、一つの長時間動画によるバイアスを防ぐねらいがある。時間幅による分割が最大数を超えた場合、コメント数の少ないシーンは除外する。以降、図2枠内の各ステップの内容について説明する。

3.1 形態素解析による単語抽出

はじめに、各シーンのコメント集合に対し、以下の前処理を

(注5) : <http://www.urbandictionary.com/>

(注6) : <https://twitter.com/>

適用する.

- 英単語を大文字に統一 (例: web → WEB) .
- 「あ, い, う, え, お, わ」の小文字を大文字に置換 (例: つよい→つよい) .
- 一つのコメントで同じ単語が連続する場合は一つの単語に置換 (例: 好き好き好き → 好き) .
- 一つのコメントで同じ文字が3回以上連続する場合は3文字に置換 (例: 8888888 → 888) .
- 「!, ?, ~」などの記号を省く.
- コメント末尾の「w」を除去する.
- 顔文字の分割を避けるため, 丸括弧でくくられた範囲は一つの単語としてみなす.

次に, 各コメントに形態素解析エンジン MeCab^(注7) を適用し, 名詞, 動詞, 形容詞, 副詞を抽出する. 形態素解析で異なる活用形が生じた場合, 各単語の基本形を用いる. なお本稿では, 5つ以上の動画に付与されたタグをニコニコ動画特有の単語とみなし, 形態素解析用ユーザー辞書として導入する. 加えて, Web 上の新語を収録した mecab-ipadic-NEologd [18] も導入する. しかし, 我々の予備実験では, 「マイリス」というスラングが「マイ」と「リス」に分割されるなど, これらの辞書では不十分な例が存在した. これを解決する簡便な方法として, 単語のバイグラムを用いる. ここで, 提案手法のバイグラムは「名詞+名詞」に限定する.

3.2 コメントの感情極性分類

日本語評価極性辞書・名詞編 [5], 用言編 [6] をシードセットとして用いて各コメントの感情極性を分類する. 前者の辞書では名詞, 後者の辞書では動詞, 形容詞, 副詞に対して感情ラベルが付与されている. 以降, ポジティブ, ネガティブ, ニュートラルのインデックスをそれぞれ p, n, ne とおく. あるコメント内で, 感情ラベル s をもつ単語が出現した回数を $Num(s)$ ($s \in \{p, n, ne\}$) で表したとき, コメントの感情極性 s^* を次式で決定する.

$$s^* = \begin{cases} \arg \max_s Num(s), & \text{if } \max_s Num(s) > 0 \\ unknown, & \text{otherwise} \end{cases} \quad (1)$$

上式のように, コメントがシードセットの単語を一つも含まない場合は, 感情ラベルとしてニュートラルではなく不明 ($unknown$) を割り当てる.

3.3 動画シーンの感情極性分類

シーン i 内のコメント集合のうち, 感情 s ($s \in \{p, n, ne\}$) に割り当てられたコメント数を $Count(i, s)$ とおく. 提案手法では, シーン i がもつ各感情のスコアを, 次式のように感情別コメント数の割合として定義する.

$$P(s|i) = \frac{Count(i, s)}{\sum_{s' \in \{p, n, ne\}} Count(i, s')} \quad (2)$$

ニコニコ動画のコメント機能の性質から, 同一シーンを視聴するユーザー間には一体感があると仮定できる. つまり, 既存コメ

表 1 各カテゴリの動画シーンの時間幅と最大分割数. 本稿では実験的に設定した.

	ゲーム	歌ってみた	政治
時間幅 (秒)	30	20	20
最大分割数	30	13	35

ントが, 新たに動画を視聴したユーザーの意見や単語の選び方に少なからず影響すると考えられる. したがって本稿では, 多数派の感情極性を優先し, $s^* = \arg \max_{s \in \{p, n, ne\}} P(s|i)$ となる感情 s^* をシーン i の感情極性とする.

3.4 スラングに対する感情極性付与

最後に, 感情別シーン集合における語の出現割合に基づき, 未知語の感情極性を決定する. データセット中で各感情 s に割り当てられたシーンのうち, 単語 w をコメントに含むシーン数を $V_s(w)$ ($s \in \{p, n, ne\}$) で表す. 提案手法は, 感情 s に対する単語 w の感情スコアを次式で定義する.

$$P(s|w) = \frac{V_s(w)}{\sum_{s \in \{p, n, ne\}} V_s(w)} \quad (3)$$

$P(s|w)$ の値が高いほど, 単語 w は感情 s の意味で使われる傾向にあることを意味する. 以上のように提案手法は, シードセットの感情極性をシーンに伝搬させたあと, シーンを媒介としてその他の単語の感情極性を決定する.

2.2 節の関連研究で前述したように, 単語は文脈によって感情極性が異なる場合がある. ニコニコ動画においても, ゲームに関する動画と政治に関する動画では, 単語の表す感情が異なる可能性がある. ゆえに, 提案手法を動画カテゴリごとに適用することで, 様々なドメインにおける感情表現を獲得する.

4. 実験

本章では, 提案手法の有効性を評価するためにニコニコ動画のカテゴリ別コメント集合を用いた実験を行う. まず 4.1 節で実験用データセットの詳細を説明し, 4.2 節でポジティブ・ネガティブ単語の抽出結果を報告する. 最後に 4.3 節において, コメントの感情分析結果を表示可能な動画視聴ツールを構築し, 提案手法の有効性を定性的に評価する.

4.1 データセット

国立情報学研究所によって提供されているニコニコ動画コメント等データ^(注8)のうち, カテゴリタグ「ゲーム」, 「歌ってみた」, 「政治」のいずれかが付与された動画を実験に用いた. 各カテゴリタグに対し, 表 1 に示す時間幅と最大分割数を実験的に設定した. シーン分割後, 各カテゴリでポジティブ・ネガティブ・ニュートラルに分類されたシーンをランダムに 50,000 件ずつ選択した. ユニグラムとバイグラムを抽出後, データセット中の出現回数が 1,000 回未満の単語は感情スコア算出対象から全て除外した. 最終的に 3,521 語が本実験の感情スコア算出対象となり, それらのうち 198 語がシードセットに収録されていた.

(注7): <http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html>

(注8): <http://www.nii.ac.jp/cscenter/idr/nico/nico.html>

表 2 「ゲーム」カテゴリにおける両極性の上位 10 個の単語.

更新なし		順位	更新あり	
ポジティブ	ネガティブ		ポジティブ	ネガティブ
GJ	おわた	1	P 乙	オワタ
KTK	あーあ	2	UPOTU	おわた
綺麗だ	オワタ	3	うぼつ	あーあ
楽しみだ	ひっ	4	うぼつー	ひでえ
うぼつ	ひでえ	5	うぼつです	これはひどい
P 乙	死んだ	6	うぼ	ひっ
上がる	これはひどい	7	888	ざまあ
いいね	ざまあ	8	KTK	死
うぼつー	おま	9	綺麗だ	ひどい
うぼつです	ぶ	10	楽しみだ	おいおい

表 3 「歌ってみた」カテゴリにおける両極性の上位 10 個の単語.

更新なし		順位	更新あり	
ポジティブ	ネガティブ		ポジティブ	ネガティブ
綺麗な声	へた	1	綺麗な声	これはひどい
きれー	これはひどい	2	うぼつー	ひでえ
も素敵	ひでえ	3	うぼつです	腹筋
素敵です	ひどい	4	声綺麗	やめて
声綺麗	死ぬ	5	もっと評価されるべき	へた
癒される	やめて	6	癒される	ひどい
声きれい	しね	7	も素敵	おいおい
素敵すぎる	きも	8	素敵です	きも
※:	おま	9	素敵すぎる	やめろ
絵も	ふいた	10	聞き惚れる	こわい

4.2 ポジティブ・ネガティブ単語の抽出結果

各カテゴリのコメント集合に対し、提案手法がとりうる以下の二つのアプローチをそれぞれ適用し、ポジティブ・ネガティブ単語を抽出した.

(1) **シードセットの更新なし**. 3 章で提案した手法を一度のみ適用する.

(2) **シードセットの更新あり**. 3 章で提案した手法を一度適用し、シードセットに収録された単語の感情スコアを更新してから再び提案手法を適用する.

「ゲーム」、「歌ってみた」、「政治」カテゴリにおいて二つのアプローチを適用した結果、ポジティブまたはネガティブスコアが上位 10 個となった単語をそれぞれ表 2, 3, 4 に示す. いずれの表においても「うぼつ」という単語がポジティブ上位として抽出されている. うぼつとは「う p 乙 (アップロードお疲れ様です)」の意であり、投稿者への労いを表す表現のため、ポジティブへの分類は妥当といえる. 「歌ってみた」カテゴリの結果を表す表 3 では、声を賞賛する単語がポジティブ上位となり、動画内容に適した表現が得られたといえる. 表 3 の最左列で 9 位となった「※:」は「弾幕」と呼ばれるニコニコ動画特有の表現である^(注9). また、いずれの表においても、くだけた日本語表現である「かけえ」「ざまあ」「ひでえ」が抽出された.

提案手法の性能を定量的に評価するために、各カテゴリタグ

表 4 「政治」カテゴリにおける両極性の上位 10 個の単語.

更新なし		順位	更新あり	
ポジティブ	ネガティブ		ポジティブ	ネガティブ
すばらしい	おまえは	1	うぼつ	こいつの
ありがとう	おまえ	2	すばらしい	クズ
888	クズ	3	かけえ	こいつも
うぼつ	死ぬ	4	素晴らしい	おまえは
素晴らしい	こいつも	5	ありがとう	おまえ
かけえ	あほ	6	かけえ	だこいつ
いいね	こいつの	7	888	嘘を
かけえ	お前	8	GJ	嘘をつく
かわいい	嘘つく	9	かわいい	こいつは
GJ	ねーの	10	麻生さん	をつく

においてポジティブ・ネガティブスコアの降順に単語を並べ、各極性の上位 100 語からなるリストを作成し、3 名の評価者 (22~23 歳の男性 2 名, 女性 1 名) が感情極性を手動で付与した. このとき、単語のみでは評価が難しい場合は全てニュートラルへの分類を依頼した. 2 名以上の評価者で一致した極性を正解ラベルとみなし、各極性の上位 N_{word} 語に対して次式の Precision を算出した.

$$Precision(N_{word}) = \frac{N_{agree}}{N_{word}} \quad (4)$$

ここで、 N_{agree} は手法により正しく極性を付与できた単語の総数を表す. 「ゲーム」、「歌ってみた」、「政治」カテゴリに対するポジティブ・ネガティブ単語抽出の Precision をそれぞれ図 3, 図 4, 図 5 に示す. いずれの図も、上段、下段がそれぞれポジティブ、ネガティブ単語の抽出結果を示し、横軸は N_{word} 、縦軸は $Precision(N_{word})$ の値を表す. 全てのカテゴリで、ネガティブ単語に比べポジティブ単語が精度良く抽出できていた. 特に、「歌ってみた」カテゴリは最も精度良くポジティブ単語を抽出できたといえる. 一方、本実験評価で最も性能が悪かったのは「政治」カテゴリのネガティブ単語の検出である. 実際、表 4 に示した上位 10 語のネガティブ単語は、半数以上が評価者からニュートラルとしてラベル付けされていた. 例えば、「お前」「こいつ」などの単語については、それ自体が強い極性をもたないと判断された. しかし、「政治」カテゴリでは動画の出演者に否定的な意見を述べる際に「お前」「こいつ」などの乱暴な表現が見られることが多く、提案手法の結果がネガティブコメント・シーン検出に有効な可能性がある. したがって、カテゴリの文脈を考慮できるよう実験の評価方法を再検討する必要がある. 加えて、実験結果から、シードセットの更新の有無が手法に影響を与えることを確認した. 各カテゴリのコメントの傾向をもとにシードセットの感情スコアを更新することで、次の適用時にシーンの感情分類性能が向上する場合がある. 今後は Precision が高くなる条件を調査し、スコアの確信度に基づきシードセットを更新する方法を検討する必要がある. その一つの方法として、式 (1) や式 (2) に閾値を導入することで、感情スコアの高いコメントやシーンのみをシードセットの更新に用いることも考えられる.

(注9): 弾幕は「歌ってみた」カテゴリで多用される傾向にあり、動画の盛り上がりシーンでは画面全体に多数の「※:」が出現する.

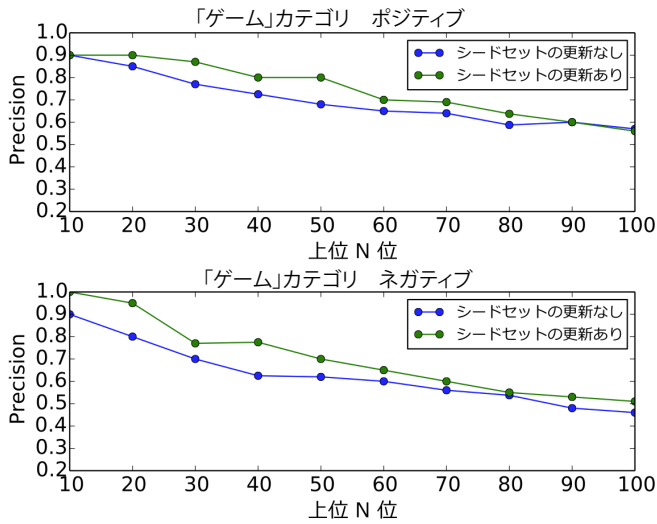


図3 「ゲーム」カテゴリにおける二つのアプローチの性能比較。

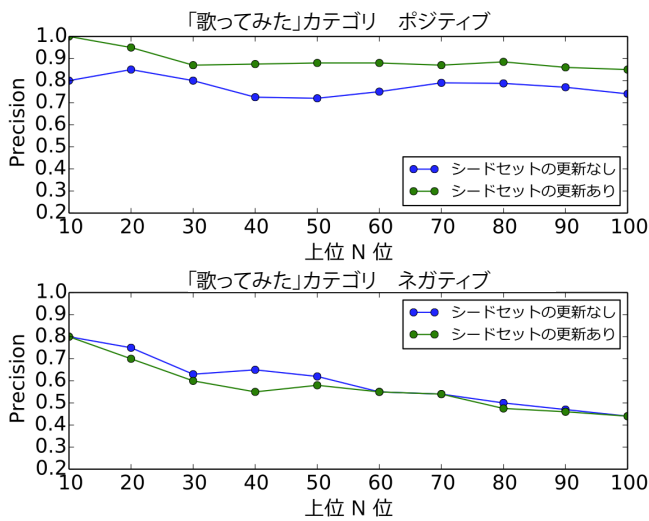


図4 「歌ってみた」カテゴリにおける二つのアプローチの性能比較。

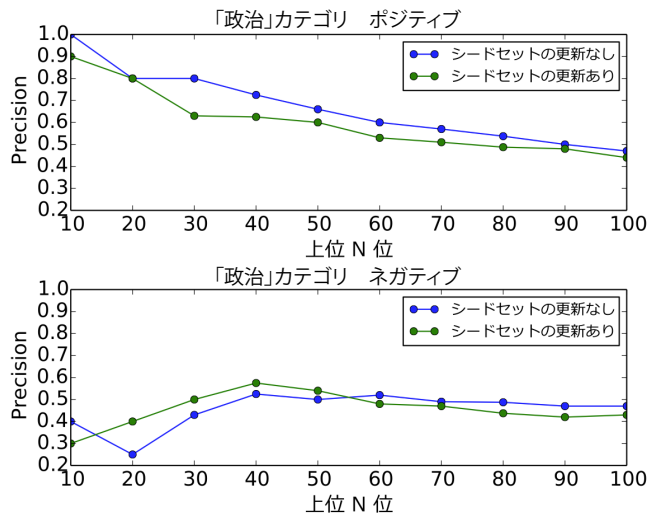
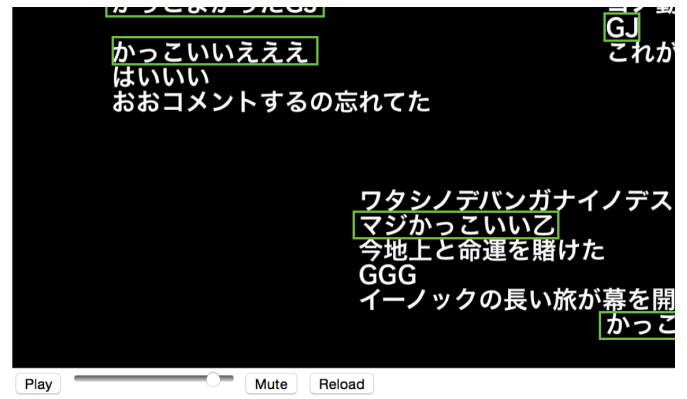


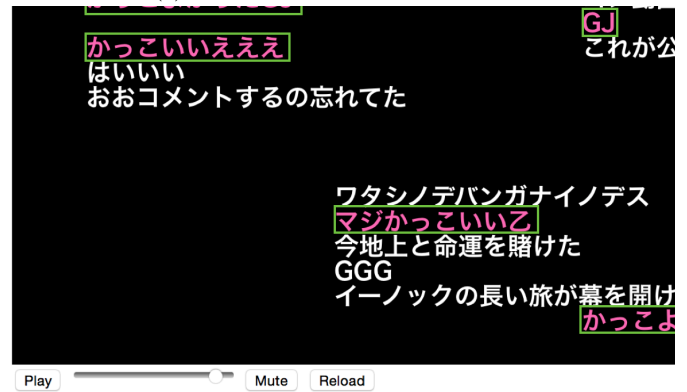
図5 「政治」カテゴリにおける二つのアプローチの性能比較。

4.3 ニコニコ動画の感情極性辞書に基づく動画視聴ツールの構築

最後に、提案手法で構築したニコニコ動画の感情極性辞書の



(a) 既存の感情極性辞書を用いた場合。



(b) 提案手法により構築した感情極性辞書を用いた場合。

図6 感情極性辞書に基づくコメント感情分析結果の一例。(a) 既存の感情極性辞書を用いた場合、(b) 提案手法により構築した感情極性辞書を用いた場合。

応用として、コメントの感情分析結果を表示可能な動画視聴ツールを構築した。提案するツールは、与えられた辞書に基づき式(1)でコメントの感情を分類し、ポジティブなら赤色、ネガティブなら青色、ニュートラルなら緑色、不明(unknown)ならば白色で文字を表示する。既存の辞書の適用例を図6(a)、各極性スコア上位30語を辞書に追加した場合の適用例を図6(b)に示す。既存の感情極性辞書は「かつこいい」「GJ」などの単語を収録しておらず、図6(a)では白色で表示されている。一方、提案手法により構築した感情極性辞書を用いた場合、これらのコメントは図6(b)に示すように正しく赤色で表示できた。ゆえに、本研究で構築した感情極性辞書は、ニコニコ動画で評判の高いシーンの検出に役立つといえる。

反対に、提案手法で構築した辞書を適用した結果、感情分類が失敗した例を図7緑枠内に示す。図7の左のシーンでは、「面白かった」というポジティブな単語に対し、「やばい」「中毒」というネガティブな単語の影響でコメント全体がネガティブにみなされた。右のシーンでは、「乙の意味知らんガキ消えろ」というコメントが「乙」という単語の影響でポジティブに分類されているが、実際には他のユーザへの批判であり動画シーンへの賞賛には該当しない。これらの問題を解決するには、文中の係り受け構造や共起単語を考慮してコメントの感情を分類する必要がある。

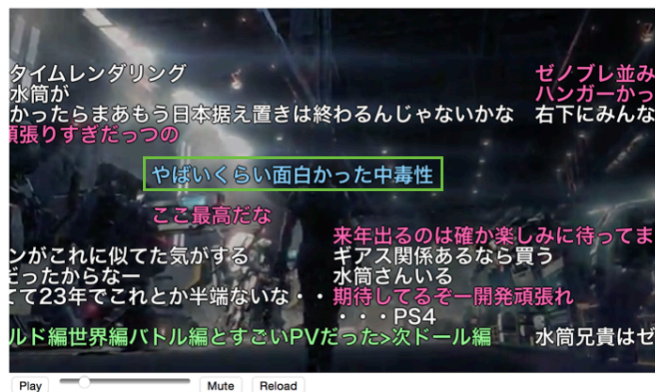


図7 「ゲーム」カテゴリの動画シーンに対し提案手法で構築した辞書を適用した場合の失敗例。

5. まとめと今後の課題

本稿では、ニコニコ動画におけるスラングの感情極性辞書の構築手法を提案した。提案手法では、フォーマルなテキストを対象とした既存の感情極性辞書をシードセットとして用い、同一シーンにおける単語の共起情報から未知語の感情スコアを算出した。三つの動画カテゴリを対象とした評価実験では、ポジティブ、ネガティブスコアがそれぞれ上位30語までは高いPrecisionを示す傾向にあった。また、シードセットを更新して再び適用するアプローチによる精度向上の可能性を示唆した。提案手法はニコニコ動画のコメント機能がもたらすユーザーの一体感を利用して未知語の感情スコアを算出した。ニコニコ動画のみならず、中国の動画共有サイト Bilibili^(注10) のような類似サービスのスラング抽出も可能と考えられる。

実験の最後には、感情極性辞書に基づくコメント分析結果を文字色で示すことのできる動画視聴ツールを構築した。ツールを通じ、既存の感情極性辞書では分析困難なコメントに対し、提案手法が感情検出のカバレッジを向上できる点を確認した。一方、正確なコメントの感情分類には係り受け構造や単語の共起の解析が必要となる例も示した。

提案手法の性能を向上するために検討すべき点がいくつかある。例として、コメント・シーン感情分類への閾値の導入や、信頼度の高い感情スコアのみを用いたシードセットの更新が挙げられる。また本稿では動画カテゴリごとにシーン分割の時間窓を実験的に設定したが、今後はコメント数に基づく適切なシーン分割方法を検討する必要がある。さらに、評価者による正解ラベル付与において、実際に用いられているシーンやコメントを提示する必要性が考えられる。今後は実験用データセットを大規模化して有効な改善方法を調査する予定である。

文 献

- [1] MyVoice：動画共有サイトに関するアンケート調査,. <http://research.nttcoms.com/database/data/000785/>. Last accessed: 01/16/2017.
- [2] M. Hamasaki, H. Takeda, and T. Nishimura. Network Anal-



- ysis of Massively Collaborative Creation of Multimedia Contents: Case Study of Hatsune Miku videos on Nico Nico Douga. In *Proc. Int. Conf. Designing Interactive User Experiences for TV and Video*, pp. 165–168. ACM, 2008.
- [3] 後藤真孝. 初音ミク, ニコニコ動画, ビアプロが切り拓いた CGM 現象. 情報処理, Vol. 53, pp. 466–471, 2012.
- [4] M. Richardson, E. Dominowska, and R. Ragno. Predicting Clicks: Estimating the Click-Through Rate for New Ads. In *Proc. Int. Conf. World Wide Web*, pp. 521–530. ACM, 2007.
- [5] 東山昌彦, 乾健太郎, 松本裕治. 述語の選択選好性に着目した名詞評価極性の獲得. 言語処理学会第14回年次大会論文集, pp. 584–587, 2008.
- [6] 小林のぞみ, 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.
- [7] 亀井且有, 豊田晃史, 串田淳一. 擬似同期を用いた動画共有によるビデオ視聴者の感情高揚. 知能と情報, Vol. 24, No. 5, pp. 944–953, 2012.
- [8] 平澤真大, 小川祐樹, 諏訪博彦, 太田敏澄ほか. ニコニコ動画のログデータを用いたソーシャルノベルティのある動画の発見に関する研究. 情報処理学会研究報告, Vol. 2011, pp. 1–8, 2011.
- [9] K. Tsukuda, M. Hamasaki, and M. Goto. SmartVideo-Ranking: Video Search by Mining Emotions from Time-Synchronized Comments. *Proc. IEEE Int. Conf. Data Mining Workshops*, pp. 960–969, 2016.
- [10] N. Murakami and E. Ito. Emotional video ranking based on user comments. In *Proc. Int. Conf. Information Integration and Web-Based Applications and Services*, pp. 499–502. ACM, 2011.
- [11] 高木潤, 中村健二, 小柳滋. 顔文字の感性情報を用いた動画コメントの評価. 情報処理学会第77回全国大会, Vol. 5, p. 02, 2015.
- [12] S. Nakamura, M. Shimizu, and K. Tanaka. Can Social Annotation Support Users in Evaluating the Trustworthiness of Video Clips? In *Proc. ACM Workshop on Information Credibility on the Web*, pp. 59–62. ACM, 2008.
- [13] H. Sakaji, J. Ishibuchi, and H. Sakai. Extracting Polarity Comments from Nico Nico Douga. In *Proc. Int. Conf. Network-Based Information Systems*, pp. 669–672, 2015.
- [14] A. Neviarouskaya, H. Prendinger, and M. Ishizuka. Sentifun: Generating a Reliable Lexicon for Sentiment Analysis. In *Int. Conf. Affective Computing and Intelligent Interaction and Workshops*, pp. 1–6. IEEE, 2009.
- [15] A. Neviarouskaya, H. Prendinger, and M. Ishizuka. Textual Affect Sensing for Sociable and Expressive Online Communication. In *Affective Computing and Intelligent Interaction*, pp. 218–229. Springer, 2007.
- [16] Y. Lu, M. Castellanos, U. Dayal, and C. Zhai. Automatic Construction of a Context-Aware Sentiment Lexicon: An Optimization Approach. In *Proc. Int. Conf. World Wide*

(注10) : <https://www.bilibili.com/>

Web, pp. 347–356. ACM, 2011.

- [17] L. Wu, F. Morstatter, and H. Liu. SlangSD: Building and Using a Sentiment Dictionary of Slang Words for Short-Text Sentiment Classification. *arXiv preprint arXiv:1608.05129*, 2016.
- [18] T. Sato. Neologism dictionary based on the language resources on the Web for Mecab. <https://github.com/neologd/mecab-ipadic-neologd>, 2015. Last accessed: 01/12/2017.