

ニコニコ動画における感情語辞書の半自動構築に関する検討

小椋 翔太[†] 桂井 麻里衣[†][†] 同志社大学理工学部情報システムデザイン学科

1. はじめに

ニコニコ動画では、日々膨大な数のコメントが動画に投稿されている。コメントは動画に対する感想だけでなく、主張の対立や同意、それに伴うユーザの感情や興味嗜好なども反映されている。したがって、コメントに着目することで動画の評判を分析することができる。

一般に、テキストの感情分析には対象ドメインに適した感情語辞書が必要となる。しかしながら、単語ごとに人手で感情タグを付けるには多大なコストが必要となる。そこで本稿では、既存の感情語辞書とコメントを用いてニコニコ動画特有の感情語辞書を半自動構築する。

2. 提案手法

本研究では、動画コメントと日本語評価極性辞書[1]を用いて、ニコニコ動画に特化した感情語辞書を構築する。提案手法の概要を図1に示す。まず、動画データとそのコメントを取得し、形態素解析により名詞のみを抽出する。次に、日本語評価極性辞書を用いて、各動画のユーザ評価をポジティブ(p)・ネガティブ(n)・ニュートラル(ne)の感情極性に分類する。具体的には、その動画に対する全てのコメントに出現する単語の感情極性をカウントし、全体の単語数に対する割合を算出する。

上記のように求めた各動画の感情極性に基づき、データセット中の単語の感情極性を再評価する。具体的には、単語 w の感情スコアを次式で定義する。

$$P(s|w) = \frac{V_s(w)}{\sum_{s \in \{p,n,ne\}} V_s(w)} \quad (1)$$

上式において、 $V_s(w)$ ($s \in \{p,n,ne\}$) は単語 w が出現する動画のうち感情 s に分類された動画数を表す。最終的に、最もスコアの大きな感情をその単語の感情とみなす。以上により、従来の日本語評価極性辞書にない単語に対しても感情スコアを算出することが可能となる。

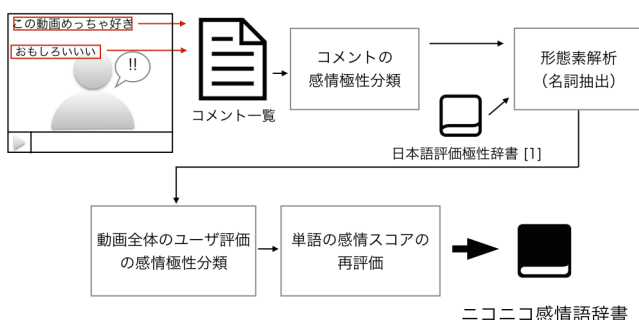


図1 提案手法の概要。

3. 実験

株式会社ダウンゴが提供するニコニコ動画コメント等データ[2]を国立情報学研究所のダウンロードサービスにより収集し、提案手法を適用した。本実験では、「ゲーム」、「アニメ」、「歌ってみた」、「東方」、「音楽」タグが付与されている動画のうち、各感情に分類された 20000 本、合計 60000 本の動画を用いて単語の感情スコアを算出した。ポジティブまたはネガティブスコアが上位 10 個の単語を表 1, 2 に示す。この結果より、「8888」、「GJGJ」、「歌詞職人」、「これはひどい」といった日本語評価極性辞書に含まれていない単語が抽出できていることがわかる。

表1 各動画カテゴリのポジティブ単語上位 10 個。

	ゲーム	アニメ	歌ってみた	東方	音楽
1	画質	高音質	きれい	音質	音質
2	高画質	高画質	綺麗	画質	すてき
3	デッキ	音質	素敵	高音質	高画質
4	綺麗	画質	すてき	綺麗	素敵
5	きれい	すてき	キレイ	アレンジ	キレイ
6	楽しみ	素敵	GJGJ	原曲	美人
7	好き	歌詞職人	なぜ伸びない	好き	きれい
8	攻め	名曲	画質	ボーカル	綺麗
9	槍	想い	8888	綺麗	高音質
10	ゲット	エコノミー	(*ω*)	雰囲気	大好き

表2 各動画カテゴリのネガティブ単語上位 10 個。

	ゲーム	アニメ	歌ってみた	東方	音楽
1	下手	無限ループ	音痴	ループ	カオス
2	地雷	ループ	不協和音	カオス	これはひどい
3	自爆	病院	下手	病気	音痴
4	ゾンビ	カオス	鼻声	テスト	最悪
5	へたくそ	病気	これはひどい	これはひどい	ループ
6	へた	腹筋崩壊	カオス	バグ	おまw
7	爆弾	いじめ	不安定	狂気	かわいそう
8	へた	鬼畜	へた	中毒	違和感
9	カス	最低	がん	最低	下手
10	バグ	これはひどい	最悪	鬼畜	変態

4. まとめ

本稿では、ニコニコ動画における感情語辞書の半自動構築手法を提案した。既存の日本語感情語辞書の情報を伝播させることで、ニコニコ動画特有の単語に対しても感情スコアが算出できていることを確認した。今後は構築した感情語辞書を用いた評判分析や、動画カテゴリごとの辞書構築について調査する予定である。

参考文献

- [1] 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 pages 584-587(2008).
- [2] 大学共同利用機関法人情報・システム研究機構国立情報学研究所 (NII): 情報学研究データリポジトリ <http://www.nii.ac.jp/cscenter/idr/nico/nico.html>