

日本語解析による Twitter の感情分析と シーンインデキシングへの応用

山内 崇資[†] 林 佑樹[‡] 中野 有紀子[‡]

成蹊大学大学院 理工学研究科 理工学専攻[†]

成蹊大学 理工学部 情報科学科[‡]

1 はじめに

テレビの高画質化・多チャンネル化や番組録画用ストレージの大容量化等の技術の大幅な進歩により、テレビ番組を大量に録画する人が増えた。しかし実際には、大量の録画番組を全て視聴する時間はない場合が多く、効率的な番組視聴を支援するために、録画した番組に対するインデキシング技術が注目されている。

従来のインデキシング技術では、番組内の映像や字幕等、番組配信側が作成する情報を用いてインデキシングを行うのが一般的だった[1]。

しかし、この方法では番組配信側が提供するコンテンツを反映させることはできるが、ユーザと同様に番組を視聴している人々の意見を反映させることは不可能である。

本研究では Twitter を用いて、視聴者の感情に基づくテレビ番組のインデキシングを提案する。テレビを視聴しながら番組内容に関する Tweet をする人が多いため、このような Tweet を解析することで、Tweet メッセージに表現される視聴者の感情を抽出できると考えた。本研究では、これを用いて番組配信側の意図ではなく、視聴者の意見に基づくインデキシングを行うことを目標とする。

2 Twitter の感情分析に基づくシーンの特徴づけ

我々はこれまでに Twitter からテレビ番組に関する Tweet を収集し、それらに対し自然言語処理の手法を適用することにより感情表現分析を行い、これに基づき、シーンインデキシングを行う手法を開発してきた[2]。具体的には、クエリ検索で番組放映時間中の番組関連 Tweet を収集し、それらの Tweet を形態素解析器 MeCab[3]を用いて分割し、各形態素と感情極性辞書を照らし合わせ、文章中で感情を表現する形態素（感情語）に合わせた得点を計算することで、

投稿したユーザの感情極性を抽出した。そして、同じ時間に投稿された Tweet の感情極性値を足し合わせることで、その時間の視聴者全体のシーンに対する感情を推定した。

感情極性辞書は感情語を 2278 語収録した辞書である。一部の感情語に対して、喜び、予期、怒り、嫌悪、悲しみ、嘆き、不安、信頼の 8 軸として表現された Plutchik の感情モデル[4]に基づいてアノテーションし、それらの形態素を種にシソーラス WordNet[5]を用いて拡張している。

これまでの研究では文の構造を全く考慮していなかったため、番組の登場人物の容姿に関する評価が、あたかも番組自体の評価であるかのように計算されてしまうといった問題があった。そこで、本研究では文構造を考慮するために、係り受け関係に基づいた感情の重み付けを行う。

3 文構造を考慮した Twitter 感情分析

感情語の周辺の係り受け関係を調べ、係り受けルールと照らし合わせて重み付けを行う。例えば、「とても楽しい」という文があった場合、「とても」が「楽しい」に係っており、「とても」が強調を表す副詞であるため、「楽しい」の感情極性値が大きくなるように重み付けされる。一方、「全く楽しくない」という文は「全く」と「ない」が「楽しい」に係っているため、「楽しい」の感情極性値は負の値になるように重み付けされる。図 1 に提案手法の処理構成を示す。

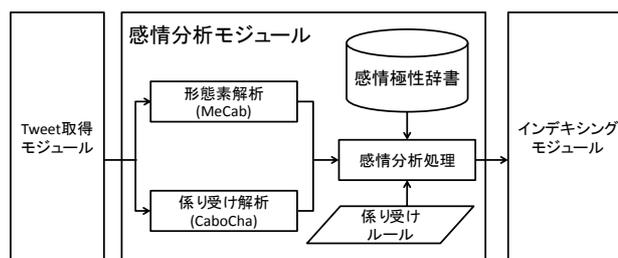


図 1: 感情分析の処理構成

具体的な処理の流れとしては、形態素解析結果を用いた従来手法による感情分析の後に、係り受け解析器から出力される Tweet メッセージの木構造に係り受けルールを適用し、ルールに合致した感情表現に重みを掛け合わせる。係り受け解析器には MeCab をベースにした CaboCha[6]を用いた。

4 係り受けルール

係り受けルールは XML で定義され、ユーザが設定することができる。図 2 に XML の例を示す。図 2 では喜びと強調のクラスタを用意し、それらを用いたルールを定義している。この例では、強調を表す副詞の「とても」が感情語に係っているため、その感情語の重みが 1.5 として設定されている。ルールには、前から修飾するルール、後ろから修飾するルール、以降の一文全体に係るルールの 3 種類を type 属性として設定できる。

```
<?xml version="1.0" encoding="UTF-8"?>
<root>
  <wordset>
    <cluster name="cluster1">
      <word>楽しい</word>
      <word>嬉しい</word>
      <word>面白い</word>
    </cluster>
    <cluster name="cluster2">
      <word>すごく</word>
      <word>とても</word>
      <word>非常に</word>
    </cluster>
  </wordset>
  <grammar>
    <rule word="cluster1" relation="cluster2" type="forward" value="1.5" />
    <rule word="かわいい" relation="ない" type="backward" value="-0.5" />
    <rule word="だけど" relation="null" type="sentence" value="1.2" />
  </grammar>
</root>
```

図 2 : 係り受け解析ルール

5 シーンインデキシング

Tweet の感情分析の応用先として Twitter 内で多く投稿されていることや、分析の容易さから、本研究では TV 番組に関連した Tweet に適用し、感情を用いたインデキシングを行った。

同じ時間に投稿された TV 番組に関連した Tweet を集め、感情極性値を足し合わせ、その合計を投稿された時間の視聴者の感情とし、時間軸を横軸にしたグラフを録画番組と一緒にユーザに提示するシステムを構築した。

図 3 にシステムのインタフェースを示す。この例では Plutchik の 8 種類の感情全てをグラフ化した場合、重なり合って見づらくなってしま



図 3 : インタフェース

ため、喜びと悲しみの感情極性値のみを提示している。

6 終わりに

本研究では、Tweet 中の感情語とそれに関わる文構造を考慮した感情解析手法を提案した。係り受け解析を行い、感情語の周辺の係り受け関係を調べ、ユーザが自由に定義した係り受けルールで感情語に対して重み付けを行うことにより、より正確に感情極性を算出できる。

今後の課題として、実際に感情が出現した時間と、Tweet が投稿された時間の時間差を考慮する必要がある。また、本研究では感情分析の応用として TV 番組のインデキシングシステムを構築したが、他の応用として GPS 情報や場所名を利用することで、楽しい感情が多く投稿されている場所などが分析可能となり、地図アプリケーションと組み合わせることによって旅行計画支援を行うシステムなども考えられる。

7 参考文献

- [1] Intille, S.S., and Bobick, A.F.: Closed-World Tracking, Proc. of ICCV'95, pp. 672-678 (1995)
- [2] Yamauchi, T., and Nakano, Y.: A Scene Explorer for TV Programs Based on Twitter Emotion Analysis, ICIC Express Letters, Vol.6, No.12, pp.3069-3075 (2012)
- [3] Kudo, T., Yamamoto, K., and Matsumoto, Y.: Applying Conditional Random Field to Japanese Morphological Analysis, Proc. of EMNLP'04, pp.230-237 (2004)
- [4] Plutchik, R.: The Emotions, University Press of America (1991)
- [5] Bond, F., Baldwin, T., Fothergill, R., and Uchimoto, K.: Japanese SemCor: A Sense-tagged Corpus of Japanese, Proc. of GWC'12, pp.56-63 (2012)
- [6] Kudo, T., and Matsumoto, Y.: Fast Methods for Kernel-Based Text Analysis, Proc. of ACL'03, pp.24-31 (2003)