

ソーシャルメディアにおける嫉妬感情の検出の一検討

上野 貴弘 森 慎太郎 大橋 正良

福岡大学工学部

1. はじめに

近年, スマートフォン, タブレット端末, PC が広く普及することにより, ソーシャルメディアが身近な存在になっている. ソーシャルメディアでは, 日々, 大量のメッセージが送受信されており, それらのなかには特定の感情に基づいて記述されている場合がある. そのため, 誹謗中傷をはじめとしたメディアリテラシーの問題が生じている[1]. この問題の一例としては Web 掲示板に投稿された誹謗中傷文が挙げられる. そこで, 磯野ら[2]の研究において, ソーシャルメディアの一つである Web 掲示板に投稿されたテキストを対象として, 皮肉と誹謗中傷の含んだ文を検出するシステムを開発した. しかし, 本来意図しない文章まで誤って検出する課題が未解決である.

本研究では, 検出精度を向上させるための足掛かりとして, ソーシャルメディアに投稿された文章から嫉妬感情に基づいた皮肉と誹謗中傷表現を検出することを目的とする. 検出した嫉妬感情に基づいた表現を学習させることで, 皮肉と誹謗中傷を含んだ文の検出要素が増え, 精度を向上させることができる. そこで, Web 掲示板に投稿された文章を対象として, 皮肉文と誹謗中傷を含んだ文から嫉妬感情を検出できるシステムを開発する. 提案手法では, 検出対象文に対して, 日本語評価極性辞書に当てはまった文字列を含んだ文を抽出する. 抽出した文章から, 皮肉及び誹謗中傷を含んだ文及びそれが嫉妬感情に基づいているかを手動で判定し, 検出された文は新たなテキストデータに加える. このシステムを開発することで, 皮肉と誹謗中傷に基づいた文章の検出精度の妥当性を明らかにすると共に, 精度を向上させるための辞書を作

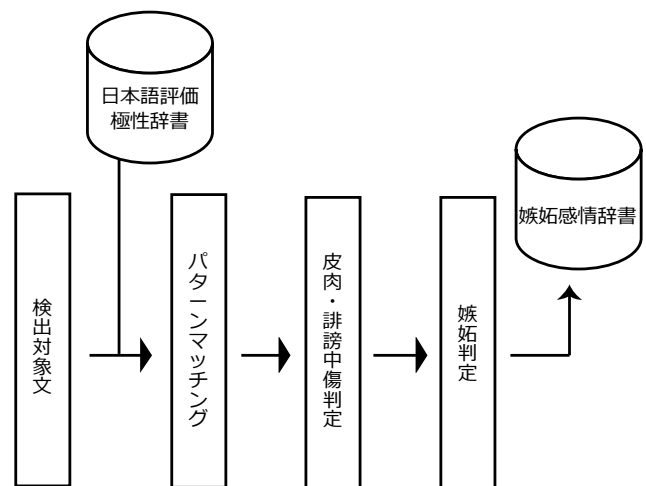


図1 開発システムのブロック図

成する.

2. 開発したシステム

図1に開発したシステムのブロック図を示す. 検出対象文については, Web 掲示板の一つである Yahoo! JAPAN ニュースのコメント欄に投稿された文章 [3] からコメントのテキスト部分を抽出し, 1037 文を検出対象とした. 嫉妬感情を検出するためのリストは日本語評価極性辞書を使用した [4]. 嫉妬感情の検出のためのシステムは, Python 言語を用いて実現した. 具体的には, 検出対象文にリストをパターンマッチングさせる. 抽出した文章から皮肉と誹謗中傷を含んだ文を手動で判定した後, それが嫉妬感情に基づいているかについても手動で判定する. 文章が皮肉と誹謗中傷を含み, 且つ嫉妬感情に基づいていると判定した場合, 嫉妬感情辞書としてテキストデータに加えるといったものである.

図2に Python が皮肉と誹謗中傷の含む文を抽出するプログラムの処理手順をリスト化した日本語評価極性辞書に記載されている文字列が検出対象文に一致するか一行ずつパターンマッチングさせ, 部分一致した場合その文字列含む文

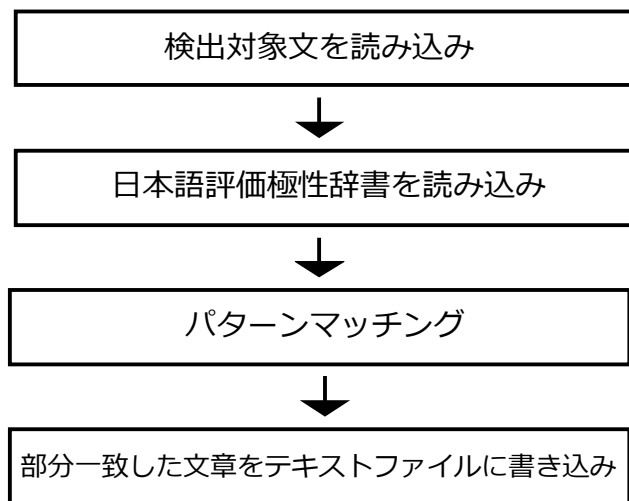


図2 プログラムのフローチャート

をほかのテキストファイルに書き込んでいく。

3. 実験結果

図3に皮肉と誹謗中傷を含んだ文の検出結果及び嫉妬感情に基づいた文の検出結果を示す。全体では、検出対象文 1037 文のうちパターンマッチングした文が 129 文、皮肉と誹謗中傷を判定した文が 6 文、そのなかで嫉妬感情に基づいている文が 1 文という結果になった。

4. 考察

実験では抽出した文章から嫉妬感情を検出できた。しかし、対象全文から検出した皮肉文からさらにフィルターをかけるため、皮肉文を検出するよりも精度が悪くなる。また、検証する対象文によって精度にムラが生じる可能性がある。今後は、嫉妬感情をもとにした文章を検出するという観点では皮肉文に限らずほかの文章からも実験する必要がある。また、嫉妬判定システムは手動のため、判定嫉妬感情に基づいた文章をまとめた辞書を作成するには膨大な時間を要する。

5. おわりに

本稿では、ソーシャルメディアに投稿された、嫉妬感情に基づいた文章の検出を目的として、嫉妬感情検出の一検討を行った。

Yahoo!JAPAN ニュースのコメント欄に投稿された文章を対象に Python を使用してパターンマッチン

```

>>> f = open(r'C:\Users\takahiro ueno\Desktop\辞書\検出対
象.txt', 'r', encoding='utf-8')
>>> taisho = f.readlines() #readlinesでリストとして読
み込む # ファイル終端まで全て読んだデータを返す
>>> f.close()
>>> b = open(r'C:\Users\takahiro ueno\Desktop\辞書\resulte
1.txt', 'r', encoding='utf-8')
>>> one = b.readlines() # ファイル終端まで全て読んだデー
タを返す
>>> b.close()
>>>
>>> file = open(r'C:\Users\takahiro ueno\Desktop\辞書\resu
lt3.txt', 'w', encoding='utf-8') #書き込みモードでオープン
>>>
>>> for i in range(len(one)):
...     a=[line for line in taisho if (one[i]) and ("であ
る" in line) or ("になる" in line) or ("する" in line) or ("があ
る" in line) or ("高まる" in line) or ("に行く" in line)]
>>>
>>> for line in a:
...     print("\n")
...     print(str(line))
...
  
```

個人なので懸賞にも当たらないと言っている弁護士がいるが、前

図3 パターンマッチングシステムのデモンストレーション

グを行い、主観で嫉妬感情を判定した文章を嫉妬感情辞書に加えるシステムを構築した。実験結果により、ソーシャルメディアから投稿された文章から嫉妬感情が検出できることが明らかになった。今後の課題としては、検出部分の辞書作成に膨大な時間を要するため、機械学習を用いた自動解析を導入する必要がある。

参考文献

- [1] 総務省：ICT メディアリテラシーの育成，総務省（オンライン），入手先 http://www.soumu.go.jp/main_sosiki/joho_tsusin/kyouiku_joho-ka/pdf/0705_ictmedia.pdf（参照 2019-1-9）。
- [2] 磯野史弥，松吉俊，福本文代，Web 掲示板における皮肉の分類および自動検出，研究報告自然言語処理（NL）2013.7（2013）：1-8。
- [3] 朝日新聞社：前沢氏の1億円お年玉，法的OK？ 社会問題化すれば…，Yahoo!JAPAN ニュース（オンライン），入手先 <https://headlines.yahoo.co.jp/hl?a=20190109-00000065-asahi-soci>（参照 2019-1-10）。
- [4] 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。