

## Twitter に表出する感情生起要因推定手法の一検討 Preliminary experiment for cause estimation of emotions in Twitter

小松 恭子<sup>†</sup>      松本 一則<sup>†</sup>      帆足 啓一郎<sup>†</sup>  
Kyoko Komatsu   Kazunori Matsumoto   Keiichiro Hoashi

### 1. はじめに

近年ブログや SNS を通した一般ユーザによる自己の情報開示が増え、開示情報からユーザの生活を垣間見ることができるようになった。ニュース番組においても、事件当事者の人間性を理解する手段として、周辺への聞き込みに加え、当人のブログ、Facebook、Twitter への投稿内容を利用するケースが増えてきた。我々は SNS への投稿を解析し、心的傾向やもの見方といった個人の内面理解を目指す。

個人を理解する研究としては、人とコンピュータの対話において正確な感情把握が重要であることから、高度な感情推定を行う研究が数多く行われている[1][2]。文献[1]では 1 人暮らしの高齢者との対話に必要な 11 感情 {嬉しい, 楽しい, 安心, 怖い, 悲しい, 残念, 嫌, 寂しい, 心配, 腹立たしい, neutral} を対象に、Web 上のテキストから獲得した感情生起要因を利用し、ユーザ発話の感情を推定している。文献[2]では属性別の感情推定モデルにより、ツイートに多く含まれていた 6 感情 {嬉しい, 希望, 愛, 悲しみ, 驚き, 不安} の推定を行っている。

しかしながら、これら是对話システムへの発話やツイートからリアルタイムな感情を推定しただけに過ぎず、その感情が想起された要因については言及していない。文献[1]で用いられる感情生起要因においては、{誕生日を迎えた} ことにより {嬉しい} という感情が想起されるように、一般的な事象は獲得できるが必ずしも個人の真意に適合しない場合がある。一方で、個人の感情とその生起要因を一定数集めることができれば、各人への適切な話題提供や接客対応に加え、オンラインコミュニティ作りを活性化することが可能になると考える。そこで本研究では、冒頭で述べた個人の内面を理解することを目的とし、内面の一部として感情とその生起要因に着目する。

本稿では、ユーザの感情が表出する投稿ツイート内には感情生起要因が含まれると仮定し、感情生起要因の推定手法を提案する。本手法を用いて評価実験を行い、文献[1]の感情推定に用いられた生起要因獲得手法との比較を行う。

### 2. 提案手法

本稿では、個人が投稿したツイートを対象とすることで、個人に特化した感情とその生起要因抽出を行う。ツイートを対象に SVM による感情推定を行い、感情推定されたツイートに対して文献[1]で利用されている接続表現 8 種類 {ので, から, ため, て, のは, のが, ことは, ことが} を用いて要因抽出を行う。

感情の種類には、プルチックの感情モデル[3]を用いる。プルチックの感情では、強度を強・基本・弱に 3 分類した感情が定義されているが、本稿では強度について言及せず {恍惚, 敬愛, 恐怖, 驚愕, 悲嘆, 憎悪, 激怒, 警戒} の 8 種類と {感情無し} の計 9 種類を用いる。

SVM による感情推定に使用する素性は、以下の手順により作成した。まず、感情が表出するツイートの抽出に適した特徴語を集めた感情種別ごとの辞書を作成する。感情別辞書に含む特徴語の品詞は 8 種類 {形容詞 (自立), 動詞 (自立), 副詞 (一般), 名詞 (サ変接続, ナイ形容詞語幹, 一般, 形容動詞語幹), 感動詞} とし、感情別ツイートに現れる対象品詞の文字列について、単語の重要度を測るために AIC (赤池情報量基準) を用いて文字列の出現度合を算出する。次に、文献[4]の定義に従って算出される E(s)値が 0 以上となる語を抽出し、SVM の素性とした。

### 3. 実験, 結果

本稿では、提案方式による生起要因抽出手法の有効性を検証するため、ベースラインと提案手法との精度比較を行う。具体的にはツイートから感情推定する際の推定精度と、推定された感情を想起する要因の抽出件数をそれぞれ評価する。ベースラインには、文献[1]で利用された感情生起要因の獲得手法を用いる。

#### 3.1 実験条件

TwitterAPI を用いて、属性を判断できるユーザアカウント 9 名 (男性 4 名, 女性 5 名) を対象とし、ツイートデータを取得する。各ユーザにおいて平均約 850 ツイート取得し、それらのリプライツイートを合わせた 9831 ツイートを実験対象とした。

本稿では、1 名の作業員によるラベル付けを行い、データに感情を付与した。複数の感情が存在する場合は、複数の感情ラベルを付与する。ラベル付けの例を表 1 に示す。次に、{感情無し} を除くツイートには、感情生起要因の有無のラベル付けおよび感情生起要因語句の選択を行った。感情生起要因の有無および語句の選択は、当該ツイートおよび当該ツイートに対して他ユーザから受けたリプライ投稿を対象に実施した。表 2 に各感情ラベルのツイート数、および各感情ラベル数のうち感情生起要因が存在する割合を示す。

#### 3.2 比較手法

本稿では、比較手法として文献[1]を用いる。文献[1]は、Web 上のテキスト文章から、文構造が「感情生起要因+接続表現+感情表現」となる文章を抽出し、感情生起要因を獲得する。接続表現は前述の 8 種類を用い、感情表現は意見や評判分析に用いられる評価値表現辞書に含まれる語から、{neutral} を省く感情 10 種を表現する下位語を手で収集している。

本稿では、文献[1]の手法を実現するために、{感情無し} を省く感情 8 種を表現する下位語を限定せず、文献[1]利用の評価値表現辞書に含まれる語全てを感情表現とし、感情表現を含むツイートを抽出する。感情表現を含むツイートを感情ありツイートとみなし、接続表現を含むツイート抽出により感情生起要因を抽出する。

<sup>†</sup> (株) KDDI 研究所 KDDI R&D Laboratories, Inc.

表 1: ラベル付けの例

ツイート本文	感情	有無	感情生起要因語句
っしやああ!!ありがとうございます!!	恍惚	無	-
@XXX これ分かります!!	敬愛	無	-
南国女子、初めての長風呂でのぼせて目眩で動けなくなって途中ヤバかった。	恐怖	有	目眩で動けなくなって
知らない間にブロックしてる人が何人かいてビックリした。。。	驚愕	有	知らない間にブロック
仕事で痛恨のミス\(^o^)/わかつちやいたけど、だろうで済ますといいことない!	悲嘆	有	仕事で痛恨のミス
新年早々マジでやめて欲しいなー。	憎悪	無	-
埼京線がエアセクションに止まった様子。	激怒	有	埼京線が止まった
キッチンがやたら広い部屋に引っ越したいなあ(´Д`)yー…	警戒	有	アイランドキッチンが夢

表 2: 各感情のツイート数と要因含有率

感情	恍惚	敬愛	恐怖	驚愕	悲嘆	憎悪	激怒	警戒	感情無し	計
ツイート数	2535	870	973	970	1433	237	250	1841	2728	9831
要因数	1971	650	655	804	1040	150	157	1238	-	6665
要因含有率(%)	78	75	67	83	73	63	63	67	-	-

表 3: 感情推定精度

	恍惚	敬愛	恐怖	驚愕	悲嘆	憎悪	激怒	警戒	感情無し
Precision	0.736	0.579	0.716	0.746	0.738	0.806	0.773	0.765	0.876
Recall	0.387	0.237	0.342	0.361	0.427	0.422	0.436	0.327	0.435
F 値	0.507	0.336	0.463	0.486	0.541	0.554	0.558	0.458	0.581
Accuracy	0.807	0.918	0.922	0.925	0.895	0.984	0.982	0.855	0.826

表 4: ベースラインと提案手法の感情抽出数・  
要因抽出数・要因抽出精度の比較

手法	要因抽出数	精度
ベースライン	267	4.00%
提案	302	4.53%

表 5: 提案手法により抽出した感情生起要因一例

1	おんのじとか懐かしすぎて泣いた
2	年末、プリンタ衝動買いしておいてよかった(*_*)
3	れりごーは公開処刑すぎて可哀想…と思つたらすでに…
4	落合陽一が普通にあるいててびびった

### 3.3 実験

本稿では、10-fold cross validation により訓練データとテストデータを用意し、他クラス SVM により各ツイートの感情推定を行う。感情推定精度は 10 回の平均とする。次に、感情推定で正答したすべてのツイートを対象とし、接続表現を含むツイートの抽出を行い、感情生起要因語の抽出とする。要因抽出精度は、正答ツイート数/感情生起要因を含むツイート数により算出した。

### 3.4 結果

提案手法による感情別の感情推定精度を表 3 に示す。網羅性に偏りがあるが、全感情平均 75%以上の精度で抽出できており、プルチックの感情モデルに基づく感情分類が実現できていることが分かる。

表 4 にベースラインおよび提案手法による要因抽出数および感情生起要因の抽出精度を示す。感情推定されたツイート数は、ベースラインで 1742 件 (ユニークな感情表現数は 1233)、提案手法で 4703 件、うち要因抽出できた数はそれぞれ 267 件、302 件であり、提案手法の方がより多くの要因を抽出できた。提案手法によってのみ抽出された生起要因例を表 5 に示す。文献[1]で用いた評価値表現辞書には評価対象に対する評価者の感情や心的な態度を表す表現が含まれるとはいえ、「泣く」や「可哀想」など表 5 に示すような感情語に関する網羅性が欠けている。一方提案手法では感情別に辞書を構築したことでベースラインで

は抽出できなかった感情文を抽出でき、新たな要因の抽出に繋がった。以上より、提案手法により個人に特化した感情生起要因をより多く抽出できた。

### 4. おわりに

本稿では、Twitter に投稿されたツイートを対象に感情推定を行い、接続表現による感情生起要因抽出手法を提案した。結果として、すべての感情において高精度に感情抽出でき、要因抽出数については比較方式よりも多く抽出できた。

一方で、現状で抽出可能な感情生起要因を一定量収集するだけでは、要因事象の粒度が細かくなり、個人の感情生起要因を感情別に把握するといった実用性を考慮すると不十分であることから、今後は要因を抽象化する手法について検討を進める。

### 参考文献

- [1] 徳久 良子, 乾 健太郎, 松本 裕治, “Web から獲得した感情生起要因コーパスに基づく感情推定”, 情報処理学会論文誌, Vol.50, No.4 (2009).
- [2] 岩朝 史展, 松本 和幸, 吉田 稔, 北 研二, “Twitter ユーザの属性別感情推定の検討”, 言語処理学会 第 22 回年次大会 発表論文集, pp.389-392 (2016).
- [3] R. Plutchik, “The nature of emotions, American Scientist”, Vol.89, pp.344-355 (2001).
- [4] Matsumoto, K and Hashimoto, K, “Schema Design for Causal Law Mining from incomplete Database”, Proc. Discovery Science: 2nd International Conference(DS'99), pp.92-102(1999).