

# 顔文字に含まれる感情成分の分析と感情極性辞書の構築

奥村 紀之      大西 智佳

香川高等専門学校 情報工学科

okumura@di.kagawa-nct.ac.jp, chark310@yahoo.co.jp

## 1 はじめに

インターネットの普及に伴い、文字によるコミュニケーションが盛んに行われ、アメブロやFC2といったブログ記事や、Twitterのようなマイクロブログ、FacebookなどのSNS(Social Networking Service)を言語資源とした研究が盛んに行われている。SNSやブログでは、ユーザが読み手に対して、文章のみでは伝達が困難である感情を適切に表現する手段として、顔文字が広く利用されている。

顔文字には:-)やXDのように左側に目を配置し、右側に口を配置して横に寝かせた顔文字(Western Style)と、(^\_^)や(>\_<)のように横に寝かせず正面を向いている顔文字(Eastern Style, Japanese Style)がある。Western Styleの顔文字がおよそ260種<sup>1</sup>であるのに対し、Eastern Styleの顔文字は約60,000種<sup>2</sup>と極めて多くの種類の顔文字が使用されている。そのため、顔文字を適切に解釈することができれば、より詳細にブログやSNSの記事から評判分析などの処理を行うことが出来ると考えられる。

Sem Eval 2013<sup>3</sup>のTask 2[1]では、Twitterの投稿から評判分析を行うタスクが設定されている。一般に評判分析ではPositive/Negative/Neutralの極性に分類する。Reynierらは、Sem Eval 2013のTask 2において、顔文字の情報を利用することによって、Happy/Sad等の感情を抽出し、Positive/Negativeの分類をさらに強いPositive/強いNegativeに分類するSSA-UOシステム[2]を開発しており、顔文字の有用性が検証されている。

本研究では、顔文字に関するアンケート調査に基づき、顔文字に含まれる感情成分の抽出結果から、顔文字の感情極性辞書の構築を試みる。顔文字は、単一の感情成分のみを保持して居らず、一般に複数の感情が重複して含有される。そのため、顔文字の感情成分の

分布を定義することが重要となる。また、顔文字は基本的な顔文字にいくつかのパーツが付与されることにより別の顔文字に派生している。そこで、顔文字の基本形を定義し、派生した顔文字がどのような感情成分を保持しているかを検討している。さらに、文章と顔文字の組み合わせにより基本的な顔文字には存在しなかった感情成分が抽出されることを確認している。

## 2 関連研究

顔文字に関する研究は盛んに行われており、既に実用レベルのシステムも考案されている。Yamadaらの研究では、顔文字の構成パーツをNグラム解析し、693種の顔文字に対して自動分類するシステムを作成している[3]。このシステムでは、笑い・泣き・怒り・驚き・困惑・不満・不安・感情なしの8種類の感情に顔文字を分類し、90%以上の非常に高い精度が出ていることを報告している。また、Tanakaらの研究では、1075種類の顔文字をk-平均法、SVMによって喜び・悲しみ・怒り・驚き・行動・苦笑いの6種の感情に分類している[4]。

Michalらの研究では、では、顔文字をパーツ単位で分析するCAOシステムを提案している[5]。また、ト部らはCAOシステムの研究成果より、顔文字の自動推薦システムを構築している[6]。この推薦システムではiOSのアプリケーションとして実験を行っており59種類の顔文字について検証している。Yamadaらの研究やMichal、ト部らの研究では、1000種類未満の顔文字についてしか検証しておらず、現在確認されている60,000種規模には至っていない。CAOシステムでは腕やボールなどの顔を構成するパーツ以外のパーツは取り扱っていないため、多様な顔文字の派生への対応が難しい。村上らの研究でも同様に、873種の顔文字辞書を構築するに留まっている[7]。

一方、大規模な顔文字辞書に関する研究として、渡邊らの研究では、56,000種ほどの顔文字をTwitterの

<sup>1</sup><http://www.techdictionary.com/emoticoncont3.html>

<sup>2</sup><http://www.kaomoji.sakura.ne.jp/>

<sup>3</sup><http://www.cs.york.ac.uk/semeval-2013/>

ログから自動抽出し、46,000 種程度の顔文字が確からしい顔文字であると判定している [8]。しかし、渡邊らが構築した顔文字辞書のカバー率は 30%程度に留まっている。

また、Twitter からの頑健な顔文字の抽出手法として、Steven らの研究がある [9]。Steven らは、顔文字のパーツの対称性に着目し、左右でペアになるパーツを抽出し、顔文字の自動抽出を試みている。

このように顔文字に関する報告は多数なされているが、本稿で着目するような顔文字の基本形と派生形といった分類はほとんどの研究で行っていない。

### 3 顔文字の感情極性辞書の構築

本稿では、顔文字と感情に関するアンケート調査を実施し、その結果に基づき顔文字の感情極性辞書を構築する。以下、アンケート調査、顔文字の基本形と派生形の定義、感情極性辞書について述べる。

#### 3.1 顔文字の使用傾向に関する調査

本研究は、顔文字に関する予備調査として行った Onishi らの研究に基づいて行っている [10]。Onishi らの研究では、Tsuchiya らが開発した感情判断システム [11] の性能向上のため、顔文字を考慮した感情推定方式を検討している。

このアンケートでは、顔文字を付与した発話文を 49 名の学生から抽出し、2,365 種の顔文字を抽出している。Tsuchiya らのシステムでは、ユーザの発話文に含まれる感情成分を、喜び・悲しみ・怒り・恐れ・恥・安心・落胆・嫌悪・罪悪感・感情なしの 10 通りに分類している。しかしながら、アンケートにより抽出した 2,365 種の顔文字と、それに付与される文章を精査したところ、感情成分はこの 10 種では十分とは言えず、基本感情の設定が必要であることが分かった。

そこで、本稿では、顔文字の基本形を定義し、そこから派生する顔文字とその顔文字が持つ感情成分の変化を感情極性辞書として構築する。また、基本感情を情緒の系図に記載されている 72 種の感情<sup>4</sup>のうち、アンケートでは確認されなかった感情や、一般的でない感情を排除し抽出した 48 の基本感情に基づき、感情極性辞書を構築している (表 1)。例えば、「珍しい」などは感情として扱うよりももの状態として扱った方がよく、「悔い」などは他の感情にある「後悔」と統合しても問題がないと判断し、除外している。

<sup>4</sup>[http://technoscenario.web.fc2.com/jyoutyo/\\_body.html](http://technoscenario.web.fc2.com/jyoutyo/_body.html)

表 1: 情緒の系図より抽出した感情 (48 種)

喜び	嬉しい	幸せ	満足	照れる
楽しい	笑み	好奇心	笑い	安心
親しみ	懐かしい	労り	恩	誇り
愛	希望	美しい	驚き	怒り
妬み	煩わしい	いい気味	厭う	不満
悲しみ	落胆	罪悪感	後悔	寂しい
絶望	嘆き	憂い	卑屈	恐れ
蔑み	羨ましい	憐れみ	諦め	恥
惑	心配	可笑しい	不安	苦しさ
高慢	欲	疑	-	-

#### 3.2 顔文字の基本形と派生形の定義

顔文字は現在およそ 60,000 種が確認されている。一方、顔文字には (((; ° °))) のように、驚いている表情を表す (° °) という顔文字から派生して構成されるものも多数存在している。Steven らの研究のように対称性に着目したものはあるが、基本形と派生形での分類はまだなされていない。顔文字をブログなどから自動抽出し、感情成分を推定することを目的とする場合、抽出した顔文字がどの顔文字の派生形であるのかを推定できれば、感情成分の抽出が容易になると考えている。

そこで、本研究では顔文字のパーツの中でも、「頬、目、口」を基本パーツとして捉え、顔文字の分類と感情成分の分布を調査する。顔文字の分類は多岐にわたるため、表 2 に示す 10 種の顔文字を基本形とし、基本形の感情成分と派生形の感情成分を比較検討する。

表 2: 基本形の顔文字 (10 種)

(° °)	(´ `)	(° °)	(´ `)
( )	(´ ´)	(・ ・)	(´ - `)
(・ ・)	(^ ^)	-	-

これら 10 種類の顔文字に対して、平均 54 個の派生形の顔文字が存在することが分かった。表 3 に (° °) から派生した顔文字の例を示す。

派生した顔文字には「\*」や「#」といった共通の付加文字が存在し、それぞれの付加文字は、ベースの顔文字が持つ感情を協調する作用があることが分かった。例えば、「\*」の場合は基本形の持つ感情が喜びである場合喜びの感情が、「#」の場合は基本形の持つ感

表 3: (° °) から派生した顔文字

(° °) b	(° ° ;)	(° ° )
(*° °*)	(° °) /	- (° °)
(° ° ;)	^ (° ° ^)	(° ° °)

情が怒りである場合は怒りの感情がそれぞれ強調されて出現していることが確認された。

### 3.3 顔文字の感情極性辞書

表 4 に構築した顔文字の感情極性辞書の一部 (基本形 (° °) に対し, 派生形 (\*° °\*) の感情極性の変化) を示す。

表 4 によれば, 基本形の (° °) に対して喜び, 嬉しいといった正の感情が多く取得されていることが分かる。これに対し, 派生形の (\*° °\*) では, 「\*」が付与されることにより, 基本形に対して喜びの感情が 3 倍近くの値を示しており, 感情が強調されていることが分かる。一方, 基本形には存在しなかった照れるといった感情が新たに取得されていることが分かる。

本稿で構築した 10 種の基本形の顔文字に対する感情極性辞書は, どの基本形に対しても同様の傾向を示しており, 感情を強調するような派生をする場合と, 全く異なる感情が生まれる場合が確認されている。

## 4 考察

本研究で構築した小規模な顔文字の感情極性辞書では, 基本形の顔文字に対しいくつかのパーツが組み合わせられることによって, 基本形の持つ顔文字の感情成分に大きな変化が生まれることが確認された。特に, 「\*」が付与されると同種の感情が強調される傾向にあることが分かった。

一方, 表 4 によれば, 基本形の感情成分としては正の感情を多く持つ顔文字であるにも関わらず, 「\*」が付与されることにより, 諦めや悲しみ, 落胆といった感情が新たに少数ではあるが取得されていることが分かる。これは, アンケート調査の方法として, 文章とそれに付与される顔文字のセットにおける感情成分を調査していることに起因する。諦めという感情が取得された基となった文章は, 「応用物理がさっぱりわからない (\*° °\*)」というものであり, もうここまで来たらどうにもならない, という心境を表していると言った被験者は答えている。

このように, 顔文字は顔文字単体から想起されるべき感情の他に, 文章との組み合わせによって生じる特異な感情が存在している。そのため, 一般的に顔文字から取得される感情成分の他に, 文章の持つ感情成分との対比が必須であることが分かる。なお, 文章と顔文字の組み合わせによって新たな感情が想起される場合については, すでに報告済みであり, 本稿で構築している感情極性辞書との統合が重要となる [12]。

## 5 おわりに

本稿では, アンケート調査に基づき顔文字の感情成分を抽出し, 顔文字の感情極性辞書の構築を行った。顔文字にはベースとなる基本形の顔文字と, 基本形の顔文字に対していくつかのパーツが追加されることで派生形の顔文字が数多く生成されていることを確認した。また, 特徴的なパーツが付与されることで顔文字の感情極性に大きな変化が生まれることを確認した。さらに, 基本となる顔文字が持つ感情成分とは異なる感情成分が抽出されていることが確認され, 文章と顔文字の組み合わせによる新たな感情想起が考えられ, 感情極性辞書への組み込みが重要であることが分かった。

本研究は, 49 名の 18 歳, 19 歳の学生から収集したアンケートのみに基づいた局所的な調査によるものであるため, 今後は被験者の年齢層を考慮し, より一般的な顔文字の感情極性の抽出を行う。また, 構築した感情極性辞書に基づき, 顔文字の自動抽出と顔文字の持つ感情成分の推定手法を構築していく。

## 参考文献

- [1] Theresa Wilson, Zornitsa Kozareva, Preslav Nakov, Sara Rosenthal, Veselin Stoyanov, Alan Ritter. SemEval-2013 task 2: Sentiment analysis in twitter. In Proceedings of the International Workshop on Semantic Evaluation, SemEval '13. 2013.
- [2] Reynier Ortega Bueno, Adrian Fonseca Bruzón, Yoan Gutiérrez, Andres Montoyo. SSA-UO: Unsupervised Sentiment Analysis in Twitter. Second Joint Conference on Lexical and Computational Semantics (\*SEM), Volume 2: Proceedings of the Seventh International Workshop on Se-

表 4: 感情極性辞書の一部

顔文字	喜び	嬉しい	幸せ	満足	照れる	楽しい	笑み	笑い	好奇心	安心
(° °)	18	19	2	-	-	5	-	-	5	1
(*° °*)	49	4	-	-	12	4	-	-	-	6
顔文字	親しみ	懐かしい	労り	恩	誇り	愛	希望	美しい	怒り	煩わしい
(° °)	-	-	1	1	-	-	-	4	-	-
(*° °*)	-	-	-	-	-	-	-	-	-	-
顔文字	妬み	いい気味	厭う	不満	悲しみ	落胆	後悔	罪悪感	寂しい	絶望
(° °)	-	-	-	-	-	-	-	-	-	-
(*° °*)	-	-	-	-	2	1	-	-	-	-
顔文字	嘆き	憂い	卑屈	恐れ	羨ましい	蔑み	憐れみ	諦め	恥	欲
(° °)	-	-	-	-	3	-	-	-	-	-
(*° °*)	-	-	-	-	-	-	-	2	-	-
顔文字	疑	惑	心配	不安	可笑しい	苦しさ	高慢	驚き	感情なし	-
(° °)	-	-	-	1	1	-	1	18	2	-
(*° °*)	-	-	-	-	-	-	-	-	-	-

mantic Evaluation (SemEval 2013), Association for Computational Linguistics. pp.501-507, 2013

[3] Taichi Yamada, Seiji Tsuchiya, Shingo Kuroiwa, Fuji Ren. Classification of Facemarks Using N-gram. International Conference on NLP-KE 2007. pp.322-327, 2007

[4] Yuki Tanaka, Hiroya Takamura, Manabu Okumura. Extraction and classification of facemarks with kernel methods. In Proc. 10th Int '1 Conf. Intelligent User Interfaces. pp.28-34, 2005.

[5] Michal Ptaszynski, Jacek Maciejewski, Pawel Dybala, Rafal Rzepka, Kenji Araki CAO: A Fully Automatic Emoticon Analysis System. Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence (AAAI-10). pp.1026-1032, 2010.

[6] ト部有記, ジェブカラファウ, 荒木健治. 顔文字の表す感情を用いた顔文字推薦システムの構築言語処理学会 第 19 回年次大会 発表論文集. P3-20, pp.648-651, 2013.

[7] 村上浩司, 山田薫, 萩原正人. 顔文字情報と文の評価表現の関連性についての一考察. 言語処理学会 第 17 回年次大会 発表論文集. E5-5, pp.1155-1158, 2011.

[8] 渡邊謙一, 高橋寛幸, 但馬康宏, 菊井玄一郎. 系列ラベリングによる顔文字の自動抽出と顔文字辞書の構築. 言語処理学会 第 19 回年次大会 発表論文集. P6-3, pp.866-869, 2013.

[9] Steven Bedrick, Russell Beckley, Brian Roark, Richard Sproat. Robust kaomoji detection in Twitter. In Proceedings of the Second Workshop on Language in Social Media (LSM '12), Association for Computational Linguistics, pp.56-64, 2012.

[10] Chika Onishi, Noriyuki Okumura. An Investigation of the Usage of Kaomoji for Emotions Judgment and Kaomoji Recommendation. The 13th IASTED International Conference on Artificial Intelligence and Applications AIA2014. #816-014, 2014.

[11] Seiji Tsuchiya, Eriko Yoshimura, Hirokazu Watabe. Emotion Judgment Method from an Utterance Sentence. KES 4, volume 6279 of Lecture Notes in Computer Science. pp.1-10, 2010.

[12] 奥村紀之, 大西智佳. 文字情報と顔文字からの話者感情推定. 信学技報, NLC2012-30. Vol.112, No.268. pp.31-33, 2012.