

E-057

話し言葉文章からの感情抽出と文章修飾による感情伝達の効果について  
 A study about the extraction of feelings from colloquialism sentences and the  
 modification of the sentences.

鈴木 聡†

Satoshi Suzuki

平山 正治†

Masaharu Hirayama

## 1. まえがき

本研究は、文章だけでは書き手の意図を伝えにくい話し言葉を対象に、顔文字や特殊記号の付与などの文章修飾を自動的に行うことを目的とした研究である。ブログやSNSなどで実際に書かれている話し言葉文章を題材にし、書き手の意図する感情の抽出方法、及びその感情に応じて修飾された文章の感情伝達の効果について述べる。

## 2. 背景

インターネット利用者の増加に伴い、ブログやSNSといった文章による交流スペースの利用者人口も急増している。総務省の調べでは、SNSの利用者は2005年3月末には111万人だったものが、2006年3月末には700万人を超えるといった急激な増加を見せている。[1]

これらの場で書かれる文章では、書き手が意図している感情をよりわかりやすく伝えるために、フォントの変更や特殊記号・顔文字の付与といった文章修飾を行うことがある。これによって書き手の意図を視覚的に訴えることができるため、本来伝えようとしていた内容を正確かつ強調して読み手に伝えることが可能となる。しかし、そのような文章修飾を行うのは基本的に手作業であり、非常に煩わしいという問題がある。

本研究は、上記問題を解決するために、入力された文章から書き手の感情抽出を行い、その結果から文章修飾を行うシステムを試作した。本論文では、この試作システムにおける感情抽出と感情伝達の効果について述べる。

## 3. 試作システム

## 3.1 概要

本研究では、単語と感情との関係性を数値化した感情データベースを用いることで、入力された文章から書き手の感情(笑い、泣き、怒り等)を自動的に抽出する。その後、抽出された感情から、その文章に対して適切な修飾を自動的に行って出力する。この試作システムの概要を図1に示す。

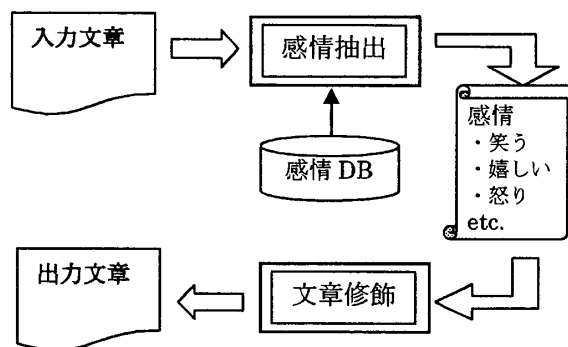


図1 試作システム概要

## 3.2 感情抽出

書き手によって入力された文章からの感情抽出の手順は図2の通りであり、詳細を以下に述べる。

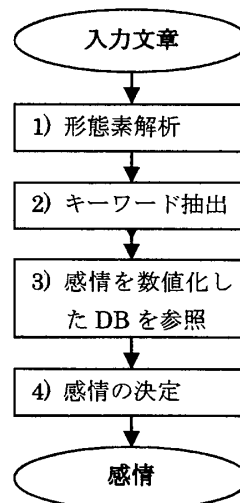


図2 感情抽出の手順

## 1) 形態素解析

形態素解析器であるMeCab[2]を用いて、入力文章を単語単位で切り出し、各単語の品詞やその活用形を得る。

## 2) キーワード抽出

形態素解析によって得られた単語から、感情表現に影響があると考えられる単語をキーワードとして抽出する。

## 3) 感情を数値化したデータベースの参照

単語ごとに感情に対する重み付けがされている感情データベースを使用し、抽出したキーワードに対する感情の重みを出力する。各単語に重み付けされる感情の種類を表1に示す。

†大阪工業大学大学院 情報科学研究科

表1 感情の種類

笑い	喜び	怒り	泣き
落胆	困惑	失敗	驚き
怯え	余裕	別れ	謝罪/依頼

## 4) 感情の決定

入力文章中の複数のキーワードから感情の重みを加えていき、この最も大きな値から、入力した文章の表現している感情を1つに絞り込む。

## 3.3 文章修飾

今回の試作システムでは、文章の修飾方法として特殊記号や顔文字の付与を行う。すなわち、表1で示される感情ごとに特殊記号(例:!,♪)や顔文字(例:(・ω・)、(>\_<))を用意しておき、抽出された感情の結果によって修飾が行われた文章を出力する。

## 4. 評価実験

## 4.1 実験結果

実際にSNSに書かれている文章を対象として、上記の試作システムにおいて感情抽出及び文章修飾を行った結果の例を表2に示す。この結果を無作為に選んだ学生20人に見てもらい、感情伝達の効果を4段階(しっかり伝わった、だいたい伝わった、特に変わらなかった、分かりにくくなった)で評価した結果を表3に示す。

表2 感情抽出及び文章修飾の結果の例

文章 1	入力	帰りは帰りで大雨に会うし、横山典弘騎手は2週間騎乗停止だし、散々でしたとき
	感情	泣き
	出力	帰りは帰りで大雨に会うし、横山典弘騎手は2週間騎乗停止だし、散々でしたとき(つゝ)

文章 2	文章 2	せめてまともにラリーができるようにならんと試合にすらならんよ
	感情	困惑
	文章 2	せめてまともにラリーができるようにならんと試合にすらならんよ(´・ω・`)

表3 感情伝達の効果

感情伝達の効果	文章1	文章2
しっかり伝わった	4人	8人
だいたい伝わった	3人	7人
特に変わらなかった	11人	3人
分かりにくくなった	2人	2人

表2の入力文章1は、話し言葉をあまり使用せず、曖昧な記述が少ないため、「大雨」や「散々」という言葉から、あらかじめ著者が予想していた感情と同様の「泣き」という結果を得ることができた。入力文章2は曖昧な記述を

含んでいるため、「ラリーができない」、「試合にならない」という文意から、著者は「落胆」と予想していたが、実際の出力結果は「困惑」となり、異なる結果が出た。

また表3の感情伝達の効果に関して、文章1では「大雨」や「散々」のような感情を連想させるキーワードを含んでおり、このような場合には効果が感じにくいという傾向が見られた。しかし、文章2では感情を連想させるキーワードが含まれておらず、顔文字を付与することによって、感情抽出における結果の正誤に関わらず、感情伝達の効果が大きいという評価を得た。

## 4.2 考察

今回の実験において文章2はくだけた表現で、より口語に近い文章となっており、「ならん」という語が「並ぶ」の活用形として抽出されてしまったため、感情抽出に悪影響が出たと考えられる。この問題を解決するためには、曖昧な記述を含む文章でのキーワード抽出において、口語にチューニングした辞書の利用が考えられる。

また、文章中に感情を連想させる単語が含まれている場合、その単語自体が直接感情伝達の効果を出すため、文章修飾による効果は得にくかった。しかし、そのような単語がほとんど含まれていない文章においては、文章修飾が感情伝達に大きな役割を果たしていると言えるため、自動的に文章修飾を行うことは有意義であると考えられる。

## 5. まとめ

本論文では、話し言葉で記述された文章を対象に、書き手が意図している感情の抽出方法、及びそれによる文章修飾の効果について述べた。今後は、より正確な感情抽出の手法と、感情伝達の効果をより大きくするために効果的な文章修飾について考えていく予定である。

## 謝辞

本研究に際して、形態素解析器「MeCab」をご提供して下さった京都大学情報学研究科-日本電信電話株式会社コミュニケーション科学基礎研究所共同研究ユニットプロジェクト、及び対象データを提供して下さった[mixi]に感謝します。

## 参考文献

[1] 総務省, 報道資料「ブログ・SNSの現状分析及び将来予測」

[http://www.soumu.go.jp/s-news/2005/050517\\_3.html](http://www.soumu.go.jp/s-news/2005/050517_3.html)

[2] 京都大学情報学研究科-日本電信電話株式会社コミュニケーション科学基礎研究所共同研究ユニットプロジェクト: 形態素解析システム「MeCab 0.97」