



Title	副詞の影響度を考慮した感情推定
Author(s)	滝澤, 満; ジェプカ, ラファウ; 荒木, 健治
Citation	ことば工学研究会 : 人工知能学会第2種研究会ことば工学研究会資料, 44, 13-18
Issue Date	2013-11-29
Doc URL	http://hdl.handle.net/2115/63964
Type	article
File Information	Fukushi.pdf



[Instructions for use](#)

副詞の影響度を考慮した感情推定

○滝澤 満 , ラファウ ジェプカ, 荒木 健治
北海道大学大学院情報科学研究科

1. はじめに

現在, PC やモバイル, タブレットなどの急速な普及に伴い, その利用目的も多様な広がりを見せている. 具体的な例として, SNS や電子メール, システムとの対話による情報検索, WWW 上の言語資源を活用したマーケティングなどが挙げられる. こういった中で, コンピュータが人の言語を理解することが必要となってきている. しかし, コンピュータが人と同等の理解を行うためには様々な課題が残っている. 実際に人間が言語を理解する場合, 感情や文法, 一般常識など様々なことを考慮している. このような扱いにくい情報の数値化を行うことは, コンピュータが人間と同等に理解を行うために非常に有効であると考えられる.

近年, 人間の感情を推定する研究は盛んに行われている. 感情を推定する手法は大きく 2 つに分けられる. 一つは, 文頭・文末表現などの表層的な表現に着目し, 構文構造を用いて感情を推定するものである[1][2]. もう一つは, Web マイニングを用いた手法である. これは, Web 上の大量の言語資源との因果関係により感情を推定する手法である[3]. また, それらの手法を組み合わせる感情を推定する研究もなされている[4][5]. しかし, これらの手法では感情がどの程度の強度なのかを判断することはできない. 感情強度は文頭・文末表現や「好き」などのような感情表現, 副詞などの影響を受ける. 感情推定に副詞を用いた研究には, 副詞および形容詞のイメージ値を用いた研究がある[6]. しかし, 感情の値をポジティブ, ネガティブのみで判定しており, 感情の強度までの推定は行われていない. そこで, 本稿では, 副詞を対象とし感情強度の推定を行う手法を提案する.

2. 副詞のイメージ値

2.1. 対象とする副詞

副詞による影響度の調査を行う前に, 文の感情強度を推定するために, 副詞のイメージ値の調査を行った. 対象とする副詞は, 感情語を伴う程度副詞を対象とした. 具体的には, 川端において考察されている程度評価用語・尺度感覚を表す語彙群(93 語) [7]のうち, Google¹検索により感情語と共起する件数が 1, 000 件未満のものを除いた 87 語を対象とした.

2.2. 副詞アンケート調査

副詞の直後にポジティブな感情語「好き, 嬉しい, 楽しい」とネガティブな感情語「嫌い, 悲しい, 辛い」それぞれを付与した. 「とても」の直後に上記の 6 語の感情語を付与し, Google¹ 検索により完全一致検索を行った結果, 「とても」のヒット件数の 4.97% に存在することが確認された. 徳久らによって得られた, イメージ値を持つ副詞を含んだ文の出現頻度が 5.39% という結果から, 概ね網羅していると考えられるため, 6 語の感情語に設定した.

各副詞にポジティブ, ネガティブの感情語を付与した 2 文を被験者に判断してもらった. 被験者に判断してもらうのは感情の強度を-5~5 の値で判断してもらった. 負の値の場合はネガティブ, 正の値の場合はポジティブ, 0 の場合はニュートラルで判断を行ってもらった. また, 一般的にその副詞を使うかどうかとも判定してもらった. 被験者は 20 代理系男子 7 名, 20 代理系女子 3 名の計 10 名である. 以下に得られた結果の例を載せる. 7 名以上が使用しないと判断した副詞, 「存外」は除外することにした.

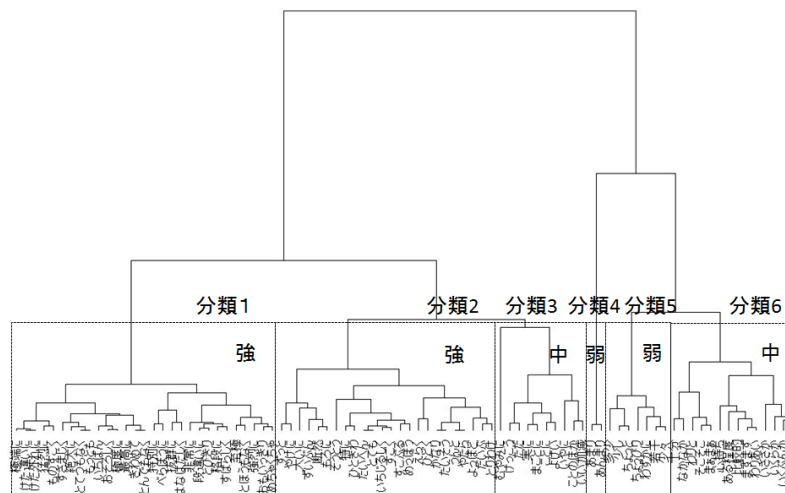


図 1 副詞のクラスタリング分類

¹ <http://google.co.jp>

2.3. 副詞の分類

アンケートで得られた各副詞に対してクラスタリングを行い、分類を行った。類似度計算は最遠隣法で行った[8]。クラスタリングを行った結果を図 1 に示す。クラスタリングにより、6 分類の副詞が得られた。得られた 6 分類の結果から、各分類ごとに平均値を求めた。その結果を表 1 に示す。これらの平均値の結果からさらに「強、中、弱」の 3 分類にした。分類 1 と分類 2 を「強」の分類とし、分類 3 と分類 6 を「中」の分類、分類 4 と分類 5 を「弱」の分類とした。その分類の例を表 2 に示す。

表 1 副詞の各分類における平均値

	1	2	3	4	5	6
平均値 (POS)	4.7	3.9	3.0	1.9	1.5	2.1
平均値 (NEG)	-4.7	-3.9	-3.3	-0.9	-1.5	-2.3

表 2 副詞の分類例

分類	例
強	いちばん、強烈に、うんと、大いに
中	いやに、ただ、ある程度、いくらか
弱	若干、少々、わずかに、あまり

3. 提案手法

システムの概要を図 3 に示す。提案手法では感情認識システム[5]を用い、感情語を伴う副詞を抽出した場合、アンケートで得られた感情強度を出力する。出力する感情強度を 2 タイプ用意した。タイプ 1 は、アンケートで得られた副詞の分類「強、中、弱」を出力する。タイプ 2 では感情語がポジティブの場合はポジティブの平均値を出力し、感情語がネガティブの場合はネガティブの平均値を出力する。ポジティブ、ネガティブを感情語で判断できない場合は、ポジティブとネガティブの平均値を出力する。

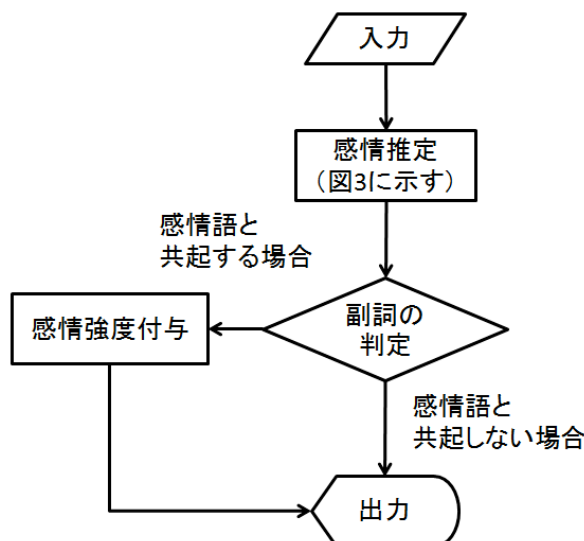


図 2 提案手法の流れ

4. 感情推定手法

ここでは感情推定手法について述べる。システムの概要を図 3 に示す。感情推定には、従来手法である、重文と否定形を考慮した、文の表層的表現とウェブ上の因果関係とを組み合わせたハイブリッド手法を用いた[5]。

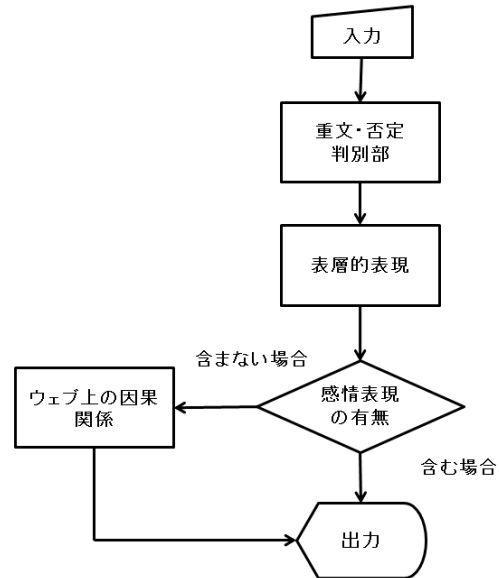


図 3 感情推定部の流れ

4.1. 感情の分類

本研究では、感情表現が網羅されている『感情表現辞典』[9]を利用し、人間の感情のカテゴリ化を行ったものを使用する[1]されている感情語は、著者である中村が{喜、怒、哀、怖、恥、好、厭、昂、安、驚}の 10 種類に分類し収集を行ったものである。小説などから収集した 2, 167 語の感情語が掲載されている。これらの感情分類を本システムでは出力する。

4.2. 重文否定の処理

入力された文が重文であった場合、二つの文に分割しそれぞれで感情推定を行う。「テニスは好きだけど、サッカーは嫌い」の場合、「テニスは好き」と「サッカーは嫌い」の 2 文に分割しそれぞれ感情認識を行う。また、二重否定を含んでいるかどうかの判定もここで行う。「ないわけがない」などのように二重否定を含んでいるかどうかの判定を行い、表層的表現での感情の遷移を行う際に用いる。

4.3. 表層的表現を用いた感情推定

表層的表現の抽出には ML-Ask[4]を用いた。「テスト終わったあああ！」などのように感情的な表現をする場合、普段とは異なる表現が文末に表出することがある。この場合、感情的な文末表現が、「あああ」と「！」のように表出する。このような感情的な文末の表現を感情要素として抽出を行う。また、「スープカレーが好きだ」などのように「好き」などの直接的な感情が文自体に含まれている場合にも、感情表現として抽出を行う。その際、中村の感情分

類を用いて分類を行い、出力する。つまり、表層的表現では文自体に含まれる感情表現および文末の感情要素を抽出し文自体が感情を含んでいるのかどうかを判断し、感情を推定する。

4.4. 表層的表現の否定形処理

また、表層的表現では否定形の処理を行う。「スープカレーは好きじゃない」のように否定形が感情語の後ろに現れた場合、Ptaszynskiらによる活発的か非活発的、肯定的か否定的かの2次元空間(ラッセル空間[10])にマッピングされた中村の感情分類(図4に示す)により、感情分類を遷移させることで感情抽出を行う[4]。この例の場合、「好き」という感情語は「好」という感情だが、2次元空間にマッピングされた反対である「厭」という感情を抽出することになる。また、二重否定が抽出された場合は2次元空間にマッピングされた感情を遷移させずに、抽出された感情語の感情を抽出する。「スープカレーが好きじゃないとは言えない」の場合、「好」という感情を抽出する。

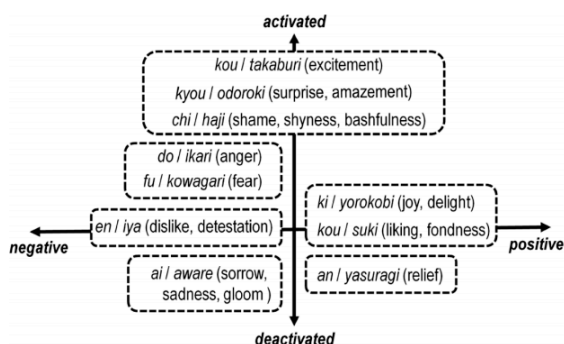


図4 ラッセル空間における中村感情分類

4.5. Web上の因果関係を用いた感情推定

Web上の因果関係の抽出には施らのシステム[3]を用いる。入力文に対して、接続詞を付与し検索クエリとしてWeb上からスニペットの取得を行う。付与する接続詞は「て、と、ので、から、たら」の5種類である。入力文に接続詞を付与したクエリより得られたスニペットのうち、クエリ直後に現れる文を因果関係を保持する文とし、感情抽出に用いる。また、因果関係のある文中に逆接を表す接続詞が現れる場合や否定を表す単語が現れた場合は感情推定には用いない。得られた因果関係を保持する文中より抽出された、各感情分類のヒット件数のうち上位60%を超える感情分類を感情推定結果として出力を行う。図5に感情推定の例を示す。

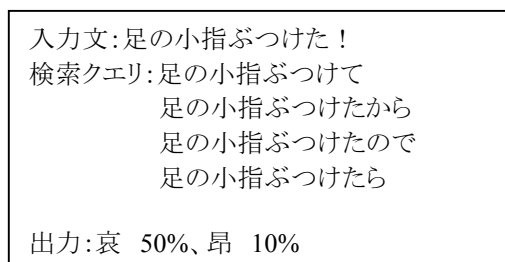


図5 Web上の因果関係の感情推定例

5. 評価実験

5.1. 事前アンケート

今回の実験に用いるためのデータの作成を行った。ツイッター²から各副詞につき感情語を伴う文を500件のうちから2,3文第一著者が抽出を行った。得られた文数が2文の場合は2文とした。得られた文から、各副詞ごとに1文ずつランダムに抽出を行った。ツイッターから抽出する際には顔文字を対象としていないため、文自体の感情抽出を行うために顔文字を含まない文に限定し抽出を行った。したがって、各副詞につき1文ずつ計86文に対し評価を行ってもらった。

また、副詞以外の影響度を考慮するために各86文に対し、副詞を除去した文86文に対しても評価をもらう。被験者は20代理系男子6名、20代理系女子2名の計8名である。それぞれの文に対し、感情の分類と感情強度の判断を行ってもらった。感情の分類は10種類の感情分類を判断してもらい、感情強度は「強、中、弱」の3種類の場合および-5~5までの値で判断してもらった。負の値の場合はネガティブ、正の値の場合はポジティブで判断を行った。

5.2. 事前アンケートの考察

事前アンケートにより得られた副詞を含まない各文の平均値の絶対値を用い、86文全体の平均値を算出した。各文の平均値は2.80であった。この平均値との差が大きかった文には、文末に「!、…」などの記号や、「ウオオオオオ、なう、ヨネ」などの文末表現が見られる場合が16文(18.6%)見られた。感情強度が平均値より低いものには「…」が7文(8.1%)見られた。感情強度を平均値よりも強めているものとしては、「!」が13文(15.1%)、「♪」が2文(2.3%)見られた。また、文末表現に記号の繰り返しや「ウオオオオオ」のように繰り返し表現されるものが表出した場合、その文の感情強度に影響を与えているものが10文(11.6%)存在した。しかし、「ね、よ」などのような文末による影響は他のものの平均0.52に比べ平均が0.06と影響が少ないことが確認された。これらのことから、これらの文末要素も感情強度に影響を及ぼすことが考えられる。

5.3. 実験

事前アンケートで収集した86文に対し、提案手法の精度の評価を行った。感情強度を「強、中、弱」の3種類で出力するタイプを提案手法(強中弱)とした。また、アンケート調査のポジティブ、ネガティブのそれぞれの平均値を出力するタイプを提案手法(数値)とした。提案手法(数値)において、ポジティブ、ネガティブの値を絶対値で出力するタイプを提案手法(絶対値)とした。提案手法(数値)と提案手法(絶対値)については、-5~5の範囲でアンケート調査を行ったため、前後1以内、つまり、差が0.5以内のものを正答とした。ただし、提案手法(絶対値)においては、アンケートで得られた数値の絶対値との比較を行った。感情分類の精度を表3に示し、感情強度の抽出結果を表4

² <http://twitter.com>

に示す。また、副詞だけの影響度を考慮するために、感情分類に失敗した 16 文を除いた場合の感情強度の精度を表 5 に示す。それぞれの提案手法で、成功した例と失敗した例をそれぞれ表 6, 表 7, 表 8, 表 9, 表 10 に示す。

表 3 感情分類の精度

	成功	失敗	精度[%]
提案手法	71 文	15 文	82. 3

表 4 感情強度の精度

	成功	失敗	精度[%]
提案手法 (弱中強)	68 文	18 文	79. 1
提案手法 (数値)	51 文	35 文	59. 3
提案手法 3	57 文	29 文	66. 3

表 5 感情強度の精度 (感情分類成功のみ)

	成功	失敗	精度[%]
提案手法 (弱中強)	58 文	13 文	80. 3
提案手法 (数値)	47 文	34 文	66. 2
提案手法 (絶対値)	52 文	19 文	73. 2

表 6 感情推定失敗例

	入力文感情	提案手法
けた外れにオモロい映画やった	無感情	喜
実に幸せな小学校時代だったわ	喜	安
今日は久々に人目も気にせずにおもいきり泣いた	昂	哀

表 7 提案手法 (弱中強) の成功例

	入力文感情強度	提案手法
今回けた違いに面白いですよ	強	強
今となってはもう、あんまり楽しくはないけれど	弱	弱
まだちよっぴり痛いけどナ...	厭	厭

表 8 提案手法 (弱中強) の失敗例

	入力文感情強度	提案手法 (弱中強)
実に幸せな小学校生活だったわ	中	強

黒バス関連はもうまじ目に入れてくねってほど割と怒ってるゾ	中	強
桜も紅葉も無いからやや物足りない...	中	弱

表 9 提案手法 (数値, 絶対値) の成功例

	入力文感情強度	提案手法 (数値)	提案手法 (絶対値)
幽霊なんかより生きてる人間の狂気のがよっぽど怖いわ...	-4. 25	-4. 11	4. 11
眼鏡っ子が特別に好き!!	4. 50	4. 25	4. 25
まだちよっぴり痛いけどナ...	-1. 38	-1. 20	1. 20

表 10 提案手法 (数値, 絶対値) の失敗例

	入力文感情強度	提案手法 (数値)	提案手法 (絶対値)
実に幸せな小学校時代だったわ	3. 88	3. 30	3. 30
ちなみに私は花粉症なのでお花はあまり好きじゃありません。	-2. 00	-0. 90	0. 90
お出迎えまことに嬉しい限りでございます	3. 88	3. 33	3. 33

6. 考察

実験結果より、感情の抽出精度は 82. 3%で行うことができた。感情抽出に失敗した例は 3 文見られた。「オモロい」などのように「面白い」などの表現が変化したものや「好き」がひらがなで表記された「すき」になった場合などである。したがって、感情語の言い換えに対する改良を行う必要があると考えられる。また、言い換え表現を感情分類に登録するか、Web 上の因果関係からの抽出の精度を上げる必要もあると考えられる。

「強, 中, 弱」の感情強度の推定の結果では 79. 1%の精度での抽出が確認でき、感情推定が成功したもののみを対象とした場合、80. 3%の精度で抽出を行うことができた。失敗した原因としては「強, 中, 弱」の「中」で分類した副詞での抽出失敗が 13 文 (15.1%) 見られた。「中」のクラスをさらに細分化する必要があると考えられる。

提案手法 (数値) の感情強度をアンケートのポジティブ, ネガティブの平均値を用いた場合では 59. 3%の精度の確認ができた。また、感情推定が成功した場合を考慮した場合、66. 2%の精度で感情強

度を抽出することができた。失敗した原因として、ポジティブ・ネガティブを考慮し、感情強度を出力した場合、「恥ずかしい」のようなどちらの値にもなりうる語による失敗が 15 文 (17.4%) 見られた。

本手法ではポジティブ、ネガティブを判断できない場合は、それぞれの絶対値の平均値として出力を行っている。このため出力される数値が正の値となり、差が大きくなってしまったと考えられる。「恥ずかしい」のような文脈に依存し、ポジティブ、ネガティブが変化する語に対しては、1 文のみの判定だけではなく、前後の文脈から推定することを考慮する必要がありと考えられる。

また、「羨ましい」と言う感情語は、感情分類では「厭」に登録されており、ネガティブな値として出力しているが、被験者はポジティブな値として判断を行っていた。文脈により単語のイメージが変わってしまうものも考えられるため、そういった語に対する調査も必要である。

また、文末表現による影響による失敗も見られた。「つーの、ゾ」のような文末表現が表出した場合や、「!、♪、…」が表出した場合にシステムと被験者との差が表れている。

提案手法 (絶対値) の絶対値の比較では、66.3% の精度が得られ、感情推定成功のみを対象とした場合では、73.2% の精度が得られた。失敗した原因は提案手法 (数値) と同様に文末表現による影響による失敗が見られる。記号による影響だと考えられるものが 6 文、「つーの、ゾ」などによる影響だと考えられるものが 5 文存在した。

7. 追加実験

7.1. 文末を考慮した感情強度抽出実験

ここでは、実験結果に対しさらなる分析を行うために行った追加実験について述べる。文末表現を考慮した提案手法 (数値) と提案手法 (絶対値) での実験を行った。文末表現は ML-Ask により抽出する。文末の影響は記号の場合は重みを 0.20 に設定し、記号以外の「ね、よ、あああ」などの表現には 0.10 の重みづけを行った。重みづけの値は、副詞アンケートの結果から実験的に第一筆者が設定したものである。文末や副詞を考慮した感情強度が -5 を下回った場合や、5 を上回った場合は、-5, 5 として出力する。それぞれ実験結果を表 8, 表 9 に載せる。

7.2. 考察

文末を考慮した提案手法 (数値) では、64.0% と考慮しない場合よりも 4.7 ポイントの上昇が確認できた。また、文末を考慮した提案手法 (絶対値) では 70.9% の精度が得られ、考慮しない場合よりも 4.6 ポイントの上昇が確認できた。これらのことから文末

表現による影響度を考慮することで感情強度の抽出精度の向上が得られることが確認された。

今回対象とした文末表現は記号や文末に表出する「ね、や、あああ」などだけであったが、「嬉しゅう限りでございます。」のように「ございます」などの特有な表現の調査を行う必要がある。そういった文末表現を収集し、感情値ごとに収集を行うことでさらなる精度の改善が得られると考えられる。

表 8 文末を考慮した提案手法の精度

	成功	失敗	精度[%]
提案手法 (弱中強)	55 文	31 文	64.0
提案手法 (数値)	61 文	15 文	70.9

表 9 文末を考慮した提案手法の成功例

	入力文感情強度	提案手法 (数値)	提案手法 (絶対値)
一人の時のこの部屋はやけに寂しいですな	-4.00	-3.50	3.50
実に幸せな小学校時代だったわ	3.88	3.40	3.40
色のセンスはない私だけど初挑戦のパッチワークはまずまず満足の出来♪	2.88	2.40	2.40

8. まとめと今後

本稿では、副詞の影響度を考慮した感情推定手法の提案を行い、アンケート調査により得られた程度副詞のイメージ値を用いて、副詞の分類を行った。感情強度を「強、中、弱」で推定する提案手法 (強中弱) では 79.1% の精度が得られた。「中」の分類における感情強度推定が課題として残されている。感情強度を数値で推定する提案手法 (数値) では 59.3% の精度が得られ、ポジティブ、ネガティブを考慮しない提案手法 (絶対値) では 66.2% の精度が得られた。感情抽出に失敗した例を除いた場合、提案手法 (数値) では 66.2%、提案手法 (絶対値) では 73.2% の精度が得られた。また、文末に現れる記号や語句を考慮した提案手法 (数値) と提案手法 (絶対値) の追加実験を行った。それぞれの精度は 64.0% と 70.9%、となり、文末を考慮しない場合よりもそれぞれ 4.7 ポイントと

4.6 ポイントの上昇が確認された。これらの結果から、副詞を考慮することで感情強度の推定が可能であることが確認できた。

しかし、文末や文中に現れる感情的な語の影響や文脈を考慮するために複数の文中での感情推定など、今後考慮していかなければならない課題が残っている。また、感情的な文に伴う特有な文末表現を収集しリスト化することでさらなる精度向上が考えられる。

参考文献

- [1] 竹本 翔, 徳久 雅人, 村田 真樹, “SO-score とパターン辞書を用いた情緒推定”, 言語処理学会 第 18 回年次大会発表論文集, Vol. 0, No. 0, pp. 771~774, 2012.
- [2] 遠藤 大介, 斎藤 真実, 山本 和英, “係り受け関係を利用した感情席気表現の抽出”, 言語処理学会年次大会発表論文集, pp. 947-950, 2006.
- [3] 施 文翰, ジェプカ ラファウ, 荒木 健治, “ウェブ上の因果関係を用いたユーザ入力文からの感情情報の推測”, 情報科学技術フォーラム講演論文集, pp. 267-268, 2008.
- [4] Michal Ptaszynski, Pawel Dybala, Wenhan Shi, Rafal Rzepka and Kenji Araki, “Contextual Affect Analysis: A System for Verification of Emotion Appropriateness Supported with Contextual Valence Shifters”, International Journal of Biometrics, Vol. 2, No. 2, pp. 134-154, 2010.
- [5] 滝澤 満, ジェプカ ラファウ, 荒木 健治, “重文および否定形を考慮した感情認識システムの改善”, 第 29 回ファジィシステムシンポジウム講演論文集, pp. 774-777, 2013.
- [6] 佐伯 美香, 徳久 雅人, 村上 仁一, 池原 悟, “副詞および形容詞による感情表現性の判定”, 科学技術フォーラム, Vol. 2, No0, pp117-118, 2003.
- [7] 川端 元子, “程度の副詞を分類する視点の考察”, 愛知工業大学研究報告, Vol. 47, pp. 115-124, 2011.
- [8] 兼子 毅, 『Rで学ぶ多変量解析』, 日科技連 2011.
- [9] 中村 明, 『感情表現辞典』, 東京堂出版 1993.
- [10] Russel, J. A. ”A circumplex model of affect”, Journal Psychology, Vol. 39, No6, pp. 1161-1178

連絡先

〒060-0814

札幌市北区北 14 条西 9 丁目

北海道大学大学院情報科学研究科

言語メディア学研究室

滝澤 満

TEL: 011-706-7389

Mail:taki_twilight@media.eng.hokudai.ac.jp