

多値ラベルによるニュース記事のコメントに対する感情推定

但馬 康宏^{1,a)}

概要: Twitter や Facebook に代表される他のウェブサービスとの関連が手軽にできる API を備えたソーシャルコミュニケーションツールによって、ウェブページに対する利用者からのコメントを容易に収集できるようになった。これらのコメントは短い、その人の感情的な意見であることが多いため、その分類、分析手法はテキストマイニングにおいて重要である。本研究では、ウェブのニュース記事に対して付けられた読者のコメントについて、その書き手がどのような感情を持って書き込んでいるかを自動分類する手法を提案し、その性能評価を行う。第一の手法は、単純なナイーブベイズによる分類であり、分類対象を形態素に分けた後に感情を表す確率変数の分布を求めることにより推定を行う。第二に、推定先の感情を「悲しみ」と「喜び」など対を成すものに分類し、その対の出現率で重みを付けた手法で推定を行う。第三に、そのような対となる感情の出現率で一度推定を行い、その後、対となっている感情のどちらであるかを推定する二段階の推定を行う。その結果、単純なナイーブベイズの適用よりも、対となる感情をもとに重み付けを行った後に分類を行う手法が高性能であることが明らかとなった。

キーワード: 感情推定, マルチラベル, マイクロブログ, ナイーブベイズ

Emotion evaluation of multilabeled comments on web news sites

TAJIMA YASUHIRO^{1,a)}

Abstract: We propose emotion estimation methods for multilabeled short comments of web news sites. Our methods can be applied to sentiment analysis and opinion mining. At first, we show the performance evaluation of a Naive Bayes classifier. Then, we propose two improved methods which focused on opposite emotions. The first improved method classifies each emotion label into two opposite emotions and uses their weights. We call this the weighting method. The second method consists of two stages of classifiers. The first stage distinguishes these oppositely classes, and the second stage selects one emotion from the opposite emotions. From our evaluation, we conclude that the weighting method is better than the simple Naive Bayes method.

Keywords: emotion estimation, multilabel, microblog, Naive Bayes

1. はじめに

Twitter, Facebook などに代表されるソーシャルコミュニケーションツールは、コメントの書き込み、およびリアクションを容易に取得できるような API が準備されている。この機能を利用し、最近のウェブのニュースページでは、読者がコメントを付けられるものが多い。また、API

の利用ではないが、テレビのニュース番組の中でも視聴者のコメントとして紹介されたり、画面の一部に表示されることもある。そのような双方向性の情報を分析、分類することは元の記事作成者にとってもその記事を利用する者にとっても有益である。

本研究では、ウェブニュース記事に読者のコメントとして書き込まれた短い文に対して、その書き手の感情推定を行う手法を提案し、評価実験を行った。コメント機能をマッシュアップで実現しているウェブ上のニュースは、一般に柔らかめの記事が多く、読者もその記事に対する感

¹ 岡山県立大学情報システム工学科
111, Kuboki, Soja, Okayama 719-1197, Japan
^{a)} tajima@cse.oka-pu.ac.jp

想や、賛成、反対など感情のこもったコメントとなることが多い。したがって、Twitter 上に流れる他の短文よりも、より感情推定の需要があり、推定精度も高くなることが見込まれる。また、これらのコメントは短い文であるため一言で表せる感情で分類できる場面が多いと思われる。Twitter のようなマイクロブログ、短文の投稿システムでは、仲間うちの連絡やある特定の話題に関する情報提供、日記のような利用も多く行われているが、ブログやニュース記事に対するコメントでは、その場で複数回の投稿を行い議論が行われることは稀である。したがって、本研究の適用対象もそのようなコメントに絞っている。

本研究における推定手法は、形態素に対するナイーブなベクトルモデルに対する機械学習であり、機械学習手法にはナイーブベイズを用いた。さらに、感情を表す確率変数について、「悲しみ」と「喜び」など対になる感情に対する前段階の推定を行うことによる性能向上を試みた。これは、本研究で用いた感情ラベルが Plutchik[6] による 8 種類の感情ラベルであり、4 つの基本的な感情とその対から構成されているためである。その結果、単純に感情を表す確率変数を用いて推定を行った場合よりも、対になる感情に対する前段階推定を行う手法により性能が向上することが確認できた。

自然言語における感情推定は、感情生起要因コーパスによる方法 [2] や、単語の結合係に基づいた方法 [1] など、コーパスや辞書を作成して行なう方法が過去に行われている。また、辞書の作成において機械学習は有益な手段であり、その手法を用いて直接感情推定を行う方法 [3] なども研究されている。文献 [5] では、感情の極性と複数の軸の組合せに対して制約を設けることで推定性能の向上をはかっている。

2. ベクトルモデルによる感情推定

n 個の形態素からなる文 t を $t = w^{(1)}w^{(2)} \dots w^{(n)}$ と表し、学習データに現れる形態素すべての集合を W_l 、評価データに現れる形態素すべての集合を W_e とする。さらに、 $W = W_l \cup W_e$ に対する形態素の異なり数を $m = |W|$ とし、一般性を失うことなく W における形態素の全順序を 1 つ定め $W = \{w_1, w_2, \dots, w_m\}$ と表す。ナイーブなベクトルモデルは、文 t に対して m 次整数ベクトル (u_1, u_2, \dots, u_m) を対応させ、このベクトルをもとに文の分類を行うモデルである。本研究では、 u_i を文における形態素 w_i の出現数とした。すなわち、

$$u_i = |\{j | w_j = w^{(j)}, t = w^{(1)}w^{(2)} \dots w^{(n)}\}|$$

である。また、文 t に w_i が現れるか否かを δ_i で表す。すなわち、

$$\delta_i = \begin{cases} 1 & w_i \text{ appears in } t \\ 0 & \text{otherwise} \end{cases}$$

である。

分類対象とする感情は、文献 [1] の研究などでも用いられている Plutchik による 8 種類の感情（喜び、悲しみ、信頼、嫌悪、恐怖、怒り、驚き、予期）とした。推定対象の文に対しては、このいずれかの感情が 1 つ以上付与されるものとし、感情ラベルがつかない文は、学習データ、評価データともに存在しないものとした。この 8 種類の感情ラベルは、以下の 4 つの分類について、それぞれ対をなす感情である。

- 喜び ↔ 悲しみ
- 信頼 ↔ 嫌悪
- 恐怖 ↔ 怒り
- 驚き ↔ 予期

対になっている感情ラベルは、同じ文に対して同時に付く可能性は大きくないと予想される。しかし、評価実験において作成したデータでは、対となるものが同時に付与されている文もあった。

2.1 単純ナイーブベイズによる推定

ナイーブベイズは確率変数 X, Y についてベイズの定理

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)}$$

により、因果関係を反転させた学習データから条件付き確率を求める手法である。本研究では、推定対象を表す確率変数を Y とする。 Y の取り得る値は { 喜び, 悲しみ, 信頼, 嫌悪, 恐怖, 怒り, 驚き, 予期 } である。感情に対応する条件は、文に対応するベクトル (u_1, u_2, \dots, u_m) である。求めるべき値は $P(Y|(u_1, \dots, u_m))$ であり、これを

$$P(Y|(u_1, \dots, u_m)) = \frac{P((u_1, \dots, u_m)|Y)P(Y)}{P((u_1, \dots, u_m))}$$

より求める。これは、学習データの各形態素に対して、その出現数から得点を与え、未知の文章に対しては、その得点の線形和で感情を求める動作となる。一般に形態素の辞書 W 上の確率分布を用いて文を生成するモデルを用いて、

$$P((u_1, \dots, u_m)|Y) = \prod_{i=1, \dots, m} u_i P(w_i|Y)$$

$$P((u_1, \dots, u_m)) = \prod_{i=1, \dots, m} u_i P(w_i)$$

とする。

推定は、 Y に含まれるすべての感情ラベルごとに $P(Y|(u_1, \dots, u_m))$ を求め、その値がしきい値を超えていた場合にラベルを付与する。しきい値は、8 種類の感情ラベルから $1/8 = 0.125$ とした。

2.2 重み付け法による推定

本研究で用いる感情ラベルは、2 値づつ 4 つのクラスに分類することができる。これは、基本的な感情を表すと思われるもの「喜び」「信頼」「恐怖」「驚き」に対してその反

対となる感情との組み合わせで1つのクラスを作ることにより定義されている。すなわち、「喜び」と「悲しみ」のように対になっている感情ラベルについて、その両方が同時に1つの文に対して付与されることが少ないことが予想される。

感情ラベルの4つのクラスを C_1, C_2, C_3, C_4 と表し、クラス C_i には2つの感情ラベル y_i, n_i が含まれるとする。具体的には、

$$\begin{aligned} y_1 &= \text{“喜び”}, & n_1 &= \text{“悲しみ”}, \\ y_2 &= \text{“信頼”}, & n_2 &= \text{“嫌悪”}, \\ y_3 &= \text{“恐怖”}, & n_3 &= \text{“怒り”}, \\ y_4 &= \text{“驚き”}, & n_4 &= \text{“予期”} \end{aligned}$$

である。このとき、 $m_i \in \{y_i, n_i\}$ とすると、ある1つの感情は、 (C_i, m_i) と表せる。したがって、ある形態素 w に対する感情ラベルの分布は以下のように表せる。

$$\begin{aligned} P(Y|w) &= P(C_i, m_i|w) \\ &= P(m_i|w, C_i)P(C_i|w) \\ &= \frac{P(w|C_i, m_i)P(m_i|C_i)}{P(w|C_i)}P(C_i|w) \end{aligned}$$

ここで $P(m_i|w, C_i)$ は、形態素 w とクラス C_i が定まった後の感情ラベルの出現分布であるが、2値であるため学習データの偏りによる影響が小さいものと思われる。また、 $P(C_i|w)$ は形態素 w が出現する文においてクラス C_i に含まれる感情ラベルが付与されている割合であり、直接感情ラベルを求めるよりも荒い分類であるため、こちらも少ない学習データで安定した値が得られるものと思われる。

実際の推定は、単純ナイーブベイズの場合と同じく、 $P(Y|(u_1, \dots, u_m))$ を上記の手法で求め、その値がしきい値を超えていた場合に感情ラベルを付与する。

この手法は、感情ラベルのクラスごとに重み $P(C_i|w)$ を用いて推定を行うため、以後重み付け法と呼ぶ。

2.3 二段階法による推定

重み付け法と同様に、感情ラベルを2値の4つのクラスに分け、 $P(C_i|w)$ を用いる。しかし、この手法では $P(C_i|w)$ の値がしきい値を超えているか否かでそのクラスに含まれる感情ラベルが付与される可能性があるかないかをあらかじめ決定する。すなわち、 $P(C_i|w)$ がしきい値以下ならば、 C_i に含まれる感情ラベル y_i および n_i は決して付与されないこととなる。 $P(C_i|w)$ の値がしきい値を超えていたならば、 $P(m_i|w, C_i)$ の値からラベルの付与を決定する(図1)。

この手法で利用される条件付確率の値は、重み付け法および単純ナイーブベイズと同様に学習データ内の形態素の出現頻度から求められる。しきい値は、第一段階のクラスの選択、第二段階のラベルの選択ともに0.1で実験を行っ

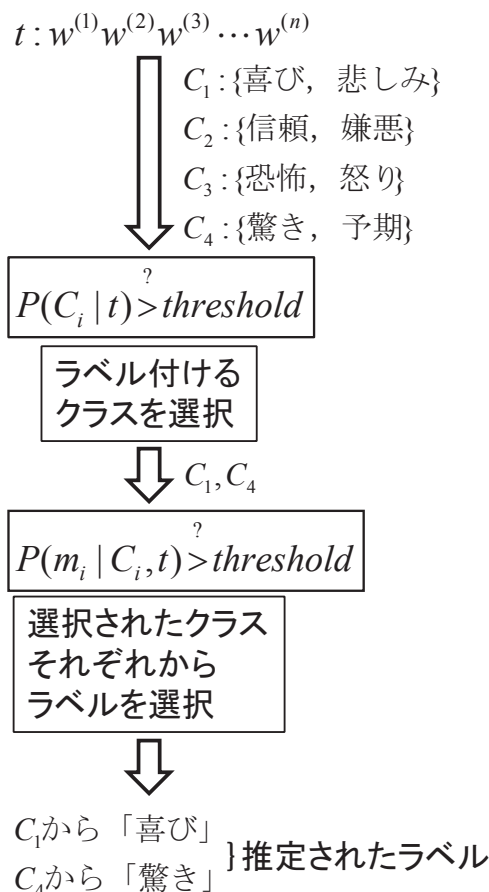


図1 二段階法による推定

た。第一段階のクラスの選択しきい値を高くすると推定で付与されるラベルの絶対数が減り、低くすると多く付くようになる。また第二段階のラベルの選択しきい値を高くすると、クラス内のどちらか一方のラベルのみが付与され、2つのラベルが同時に付与される確率が低くなる。逆に第二段階のしきい値を低くすると、対になる感情ラベルの両方が付与される確率が高くなる。

3. 評価実験

3.1 実験データ

評価実験として、ウェブニュースサイトにおける記事に対するコメントの感情推定を行った。対象としたニュースサイトはニコニコニュース (news.nicovideo.jp) とし、以下の条件のコメントを収集した。このデータは、以前の我々の研究 [4] で用いたものから、リツイートに関する文字列を削除し、前後の空白の削除を行ったデータである。

- 対象ニュース記事数: 28
- 対象コメント総数: 2075
- 1記事あたりの平均コメント数: 78.04
- 1記事あたりの最大コメント数: 100
- 1記事あたりの最小コメント数: 12
- 1コメントあたりの平均形態素数: 19.34

- 1 コメントあたりの最大形態素数：65
- 1 コメントあたりの最小形態素数：1
- データ全体の形態素の異なり数：4987

学習データは収集したコメントのみであり、ニュース記事本体は学習、評価双方において利用していない。各コメントは、形態素解析器 mecab を用いて形態素に分け、それぞれの原形と品詞のペアで形態素を区別した。

感情ラベルは、それぞれのコメントに対して手作業で付与した。この方法では作業者の考えにより偏ったラベル付けがなされる可能性がある。この意見の偏りを少なくするため、各コメントに対して最低2名の作業者が割り当てられるようにし、どのコメントも重複を含めて2つ以上の感情ラベルが付くようにした。以後、ひとりの作業者により付与された、1つの感情ラベルを1票と呼ぶ。

したがって、1つのコメントに複数の感情ラベルが付くこともあり、また作業者全員の意見が一致すれば、1つの感情ラベルに複数の票となることもある。本研究では、感情ラベルに複数の票が入ることに関しては学習を行わず、2人以上が同じ感情ラベルをつけた場合でも、その感情ラベルがついていることのみを学習データとして利用している。

感情ラベル付け作業は、のべ12人により行われ、ひとりの作業者は1つのコメントに対していずれかの感情ラベル1票を付与できる。1コメントについて付けられた感情ラベルは、平均2.52票、最大9票、最小1票である。各感情ごとの票数は以下の通りである。

- 喜び：516票
- 悲しみ：756票
- 信頼：147票
- 嫌悪：1347票
- 恐怖：291票
- 怒り：936票
- 驚き：580票
- 予期：656票

感情ラベルが1票以上付いたものを、そのコメントのもつ感情とする。収集したデータにおける、それぞれの感情をもつコメント数は以下の通りである。

- 喜び：326コメント
- 悲しみ：583コメント
- 信頼：130コメント
- 嫌悪：954コメント
- 恐怖：217コメント
- 怒り：621コメント
- 驚き：405コメント
- 予期：474コメント

前記の票数と食い違う部分が、複数の作業者により重複してラベル付けられた部分である。1つのコメントがもつ感情、すなわち1票以上の感情ラベルの数は、均1.79、最大5、最小1である。

ウェブニュースに対するコメントである点から、否定的な感情、文句を言う場面が多く、最も多く付けられたラベルは「嫌悪」である。すべてのコメントに対して「嫌悪」だけをラベル付けた仮説を提示した場合、1コメントあたりの平均性能は、適合率0.46、再現率0.25、F値0.32である。この値がベースラインとなる。

3.2 実験方法と結果

単純ナイーブベイズ、重み付け法、二段階法のそれぞれについて、以下の実験を行った。

3.2.1 単純5分割交差検定

各ニュース記事に対するコメントの1/5を評価データとし、残りの4/5を学習データとした。それぞれのニュース記事に対するコメントが必ず含まれているため、評価は高めになると見込まれる。図2および表1にそれぞれの

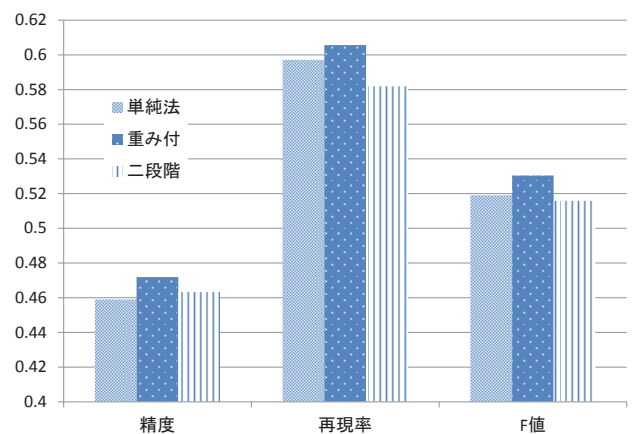


図2 5分割交差検定のグラフ

表1 5分割交差検定の値

推定手法	単純法	重み付	二段階
精度	0.4590	0.4718	0.4632
再現率	0.5970	0.6056	0.5818
F値	0.5190	0.5304	0.5158

推定手法による精度、再現率、F値のグラフおよび値を示す。いずれの値でも重み付け法が最も性能が高い。二段階法は、精度では単純ナイーブベイズを上回っているが、再現率およびF値では劣っている。

3.2.2 1つの記事のコメント除外

あるニュース記事に関するコメントをすべて評価データとし、それ以外の記事のコメントを学習データとする方法。同じ話題のコメントが含まれていないため、評価は低くなると見込まれる。図3および表2に結果のグラフおよび値を示す。5分割交差検定に比べ、それぞれの推定手法においてすべての性能値が下がっている。1つの記事に関するコメントをすべて学習データから除外し、評価データと

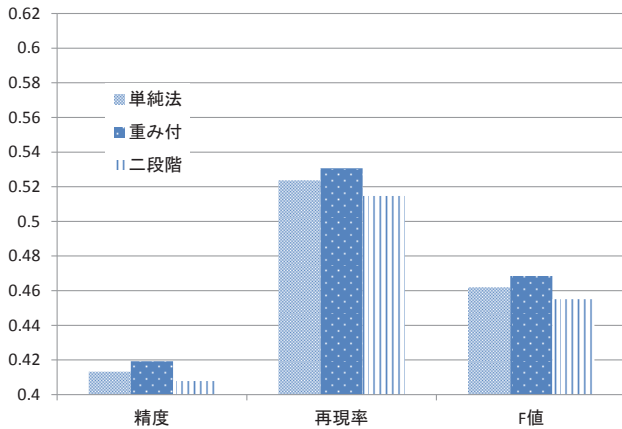


図 3 1つの記事のコメント除外のグラフ

推定手法	単純法	重み付	二段階
精度	0.4133	0.4193	0.4079
再現率	0.5238	0.5307	0.5147
F 値	0.4620	0.4685	0.4551

した影響であることがわかる。推定方法の違いによる性能は、5分割交差検定のとおり同じく、重みづけ法、単純ナイーブベイズ、二段階法の順となっている。またこの場合も、すべての推定手法において再現率が高く精度が低いことが読み取れる。これは、感情ラベルを多めに付与する傾向であり、しきい値の設定の変化による性能変化を調べることが今後の課題となる。

3.2.3 Leave one out 法

1つのコメントのみを評価データとし、残りすべてを学習データとする方法。最も学習データ数が増える評価方法。図4および表3にそれぞれの手法の精度、再現率お

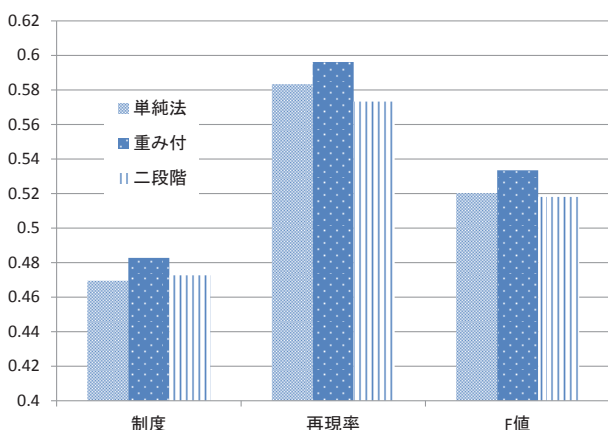


図 4 Leave one out のグラフ

よび F 値のグラフと値を示す。1つの記事除外よりも精度、再現率、F 値ともに高性能であり、単純交差検定よりも精度と再現率の離れ方が少ないことがわかる。また、3

表 3 Leave one out の値

推定手法	単純法	重み付	二段階
精度	0.4695	0.4827	0.4726
再現率	0.5833	0.5961	0.5732
F 値	0.5202	0.5334	0.5180

つの実験方法のうち、最も学習データが多いため、F 値はどの推定手法でも最も高い値となっている。

3.2.4 クローズドテスト

図5および表4に今回作成したデータ全てに対するクローズドテストの結果を示す。オープンデータの実験結

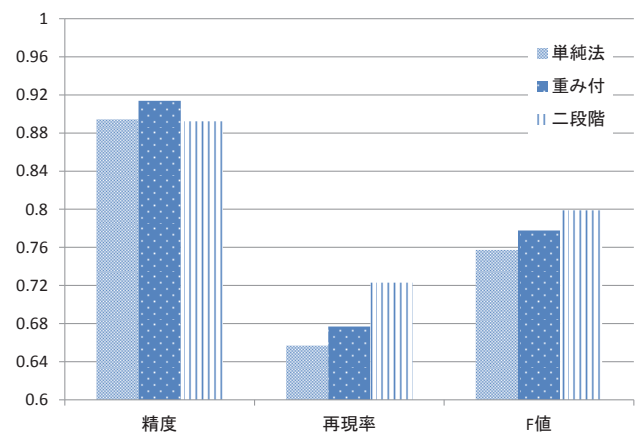


図 5 クローズドテストのグラフ

表 4 クローズドテストの値

推定手法	単純法	重み付	二段階
精度	0.8944	0.9139	0.8924
再現率	0.6569	0.6771	0.7231
F 値	0.7575	0.7779	0.7989

果とは対照的に、二段階法の F 値が最も高い値となっている。特に精度は、オープンデータに対する結果と同様の性能であるが、再現率が単純法、重み付け法、二段階法の順に高くなっている。これより、学習データに対する再現性は二段階法が優れていることがわかる。

4. おわりに

本研究において、マルチラベルの付いたコメントの感情推定に対してナイーブベイズを基礎とした2つの新たな手法、重み付け法と二段階法を提案し、重み付け法は従来のナイーブベイズよりも高性能であることを3タイプの実験環境で示した。2つの新たな手法はともに、感情ラベルがいくつかのサブクラスに分けることができかつ、そのクラスにおいてはある程度排他的なラベル付けが行われる場合に適用することができる。本研究で用いた8種類の感情ラベルは、対になる4つのクラスに分けられるものであったが、これがすべて似たような感情であっても、各クラス内

のラベルが排他的に付与されるものならば、本研究による提案手法はいずれも適用可能である。このような場合の性能評価は、今後の課題である。

さらに、今後の課題として、しきい値の効率的な定め方が挙げられる。特に二段階法では、最終的なラベル付与に関わる前にクラス選択のしきい値もあるため、その効率的な見つけ方は本手法の発展において重要である。また、二段階法ではクラスの選択において、ナイーブベイズ以外の選択手法も利用可能であり、異なる機械学習法を取り入れることも今後の課題として挙げられる。

ニュースの記事本体を利用する方法も今後の課題である。本研究のような形態素に対する確率値の付与では、コメントのような短い文に対する大きな性能向上は見込めない。コメントは、本文に対する感想、意見であることを考えると、今後の性能向上では記事本文の傾向を判断し、コメントの感情推定に反映させる仕組みが不可欠であると思われる。

参考文献

- [1] 情緒生起情報付き結合価パターン辞書の開発, 田中努, 徳久雅久, 村上仁一, 池原悟, 言語処理学会第12回年次大会 発表論文集, pp.1151-1154, 2006.
- [2] 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, 徳久良子, 乾健太郎, 松本裕治, 情報処理学会論文誌, vol.50, no.4, pp.1365-1374, 2009.
- [3] "えもにゅ"における短文の感情推定について, 小川拓貴, 松本和幸, 任福継, 情報処理学会研究報告 自然言語処理 2010-NL-195(2), pp.1-6, 2010.
- [4] コメントとしてつぶやかれた短文の感情推定, 但馬康宏, 信学技報, 言語理解とコミュニケーション研究会, vol.112, no.196, pp.37-40, 2012.
- [5] 感情軸における感情極性制約を用いたマルチラベル感情推定, 江崎大嗣, 小町守, 松本裕治, 言語処理学会第19回年次大会 発表論文集, pp.244-247, 2013.
- [6] The emotions, R. Plutchik, University Press of America, Lanham MD, 1962.