

平成 27 年度 卒業論文

# Web ニュースデータを用いた 感情極性値推定

法政大学 理工学部 経営システム工学科

経営数理工学研究室

2016年2月

学籍番号:12X4102

志村 千尋

指導教員:五島洋行 教授

学科名	経営システム工	学籍番号	12x4102
申請者氏名		志村 千尋	
指導教員氏名		五島 洋行	

## 論文要旨

論文題目	Web ニュースデータを用いた感情極性値推定
------	------------------------

本研究では、Web ニュースデータと単語感情極性対応表を用いて新規単語の感情極性値の推定を行う。7年間のWeb ニュースデータを形態素解析し、単語ごとに分割し、使用頻度を算出する。極性値が与えられていない単語を、今回極性値の推定を行う対象の単語とし、指数平滑法にて計算を行う。Web ニュースデータの本文中で使われている、対象の単語を含めた前後3単語に対して、極性値を与え、時系列に計算を行い比較、考察を行う。

単語感情極性対応表にない単語に対して、極性値を与え、極性値を持つ単語を増やす。本研究の手法では、ニュースデータの期間を変える事によって、期間内での単語感情極性値を算出する事が可能である点が他と異なる。これによって、テキストの感情分析を行うことができ、マーケティング等の分野に活用できると考えられる。

# 目次

第1章 はじめに.....	1
1.1 研究背景.....	1
1.2 研究目的.....	4
第2章 関連研究.....	5
2.1 菅原らの研究.....	5
2.2 高村らの研究.....	5
2.3 研究の実用性.....	6
第3章 関連知識.....	8
3.1 Web ニュースデータ.....	8
3.2 形態素解析.....	8
3.3 単語感情極性対応表.....	9
第4章 分析手順.....	11
4.1 前提条件.....	11
4.2 出現頻度分析.....	11
4.3 指数平滑法.....	11
第5章 分析結果.....	13
5.1 頻度分析結果.....	13
5.2 指数平滑計算結果.....	15
第6章 おわりに.....	17
参考文献.....	18
謝辞.....	19
付録.....	20

## 第1章 はじめに

本論文では、ニュースデータを用い新たな単語の極性値を算出し、比較を行う。これにより、既に存在している単語感情極性対応表よりも多くの単語に値を与える事が可能となり、単語の感情評価がわかる。本研究を行うことに伴い、必要なインターネット上のニュースに関する背景、単語感情極性値、本研究を行う動機を説明する。

### 1.1 研究背景

近年、インターネットの普及やIT技術の進歩により、誰もが膨大な情報量を手に入れる事が可能になっている。誰もが入手できるという点において、図1のインターネット利用者数の推移[1]のデータがある。平成13年度末から平成25年度末までの13年間において、利用者数が約2倍に増えている事がわかり、平成25年末ではインターネット利用者数は1億人を超えている。膨大な情報量という点において、例えばWebニュースが挙げられる。元々、TVや新聞でニュースが報道される機会が多かった。しかし、図2の新聞発行部数の推移[2]によると、2007年では約5,200万部を発行していたが、2015年では約4,400万部と8年間で約800万部の減少傾向を示しており、新聞発行部数の復調の兆しが見えないということが現状である。一方、近年ではスマートフォンやタブレットに媒体を変える人が増えてきている。図3はモバイルニュースアプリとモバイルWebニュース利用者数の推移[3]を表している。2012年度から毎年モバイルニュースアプリの利用者数は増え続けており、2017年度には2012年度の約15倍になることが予想されている。具体的には、Yahoo!JapanのYahoo!ニュースやグノシー、SmartNews、LINENEWSなどのアプリが存在する。また、モバイルニュースデータアプリの利用者数と共に、モバイルWebニュース利用者数も増えていることがわかる。こちらは、同期間で約1.5倍の利用者になることが予想されている。以上から、ニュースを新聞からスマートフォンやタブレット等で、ニュースを閲覧する方の需要が増えているということがわかり、図2、図3を比較すると近年のニュース閲覧方法や媒体の変化がわかる。また、TwitterやFacebook等のSNSの普及により、記者だけでなく、我々もインターネット上に文書を書く機会が増えてきている。しかし、そのような文書内の感情を発見することや特定することに対しては、様々な場面の活用に重要な課題がある。例えば、ある会社の新規製品に関しての評価を、インターネット上での書き込みやコメントを通じて行うことが可能になれば、その後の施策や需要に見合った製品開発の手助けになると考えられる。

このような文書型テキストの感情分析を行うリソースとして、単語の感情極性というものが挙げられる。これは、ある単語が良い印象（ポジティブ）か、悪い印象（ネガティブ）かについて数値で表すものである。感情極性値を用いることで、文全体の評価を行うことが可能になり、ある対象の評価を自動的に算出することが可能になる。そのため、感情極性値を日本語すべてに与えることができれば、分析が容易になり、さらにその先の分析や実行、判断に活かすことができるだろう。

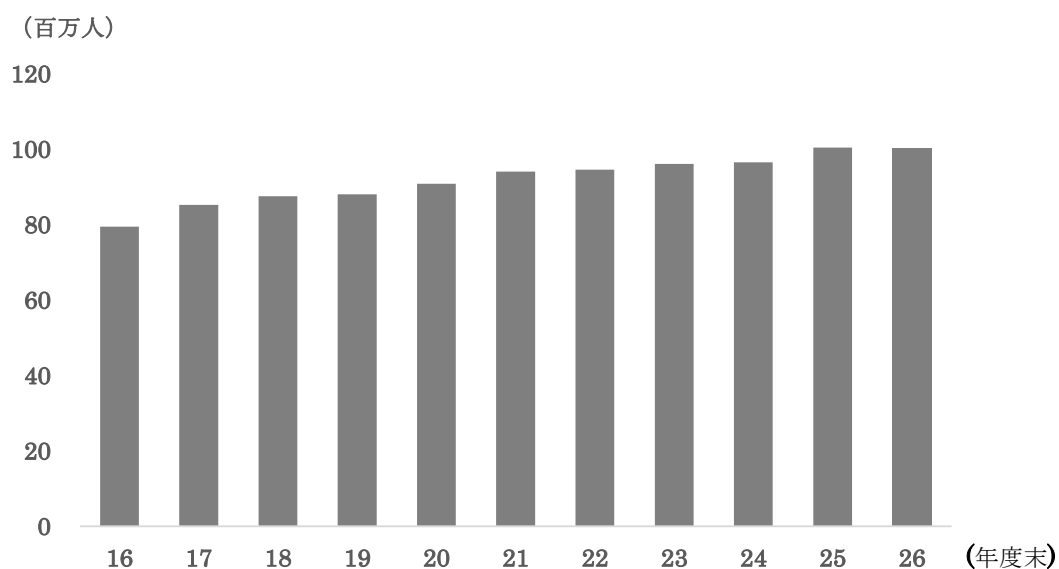


図1. インターネット利用者数の推移

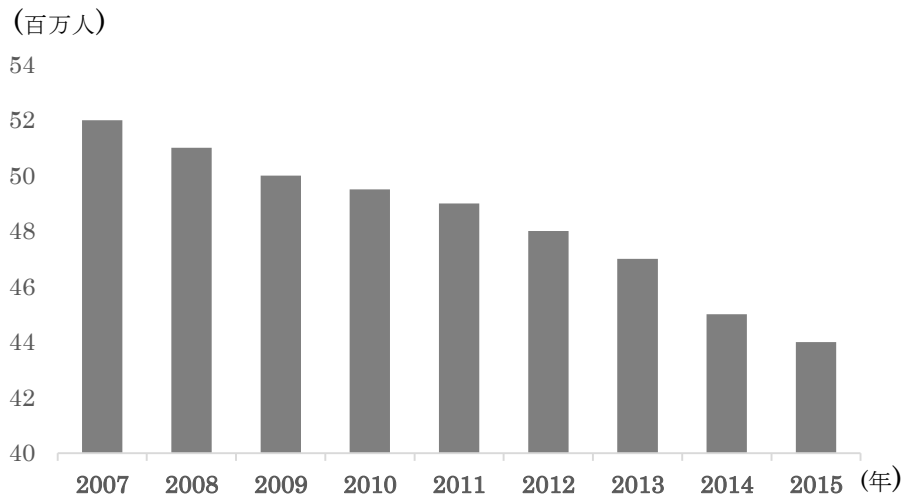


図 2. 新聞発行部数の推移

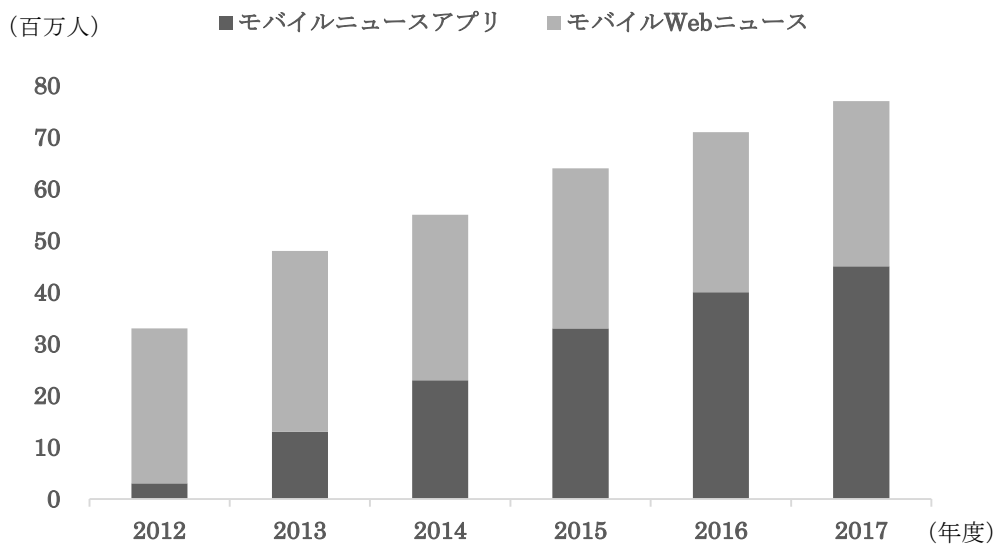


図 3. モバイルニュースアプリとモバイル Web ニュース利用者数の推移

## 1.2 研究目的

1.1 では、近年におけるインターネットと、モバイルニュースの普及を時間の経過とともに変化している事を表している。このデータからニュースをパソコンやスマートフォン、タブレットなどから見る人が増えている事がわかる。また、感情極性値という分析を行うリソースの可能性についても述べた。

そこで本研究では、Web 上にあるニュースの本文をデータとして扱うことで、ある特定の単語が時間の経過とともに、ポジティブもしくは、ネガティブのどちらの極性を表すのかを、値を算出し考察を行う。55,125 語の極性値を表している単語感情極性値対応表はすでに存在しているが、より多くの単語感情極性値を表す必要があると考える。

## 第2章 関連研究

感情極性とは、人の感情をテキストから判別する自然言語処理技術を応用した研究である[4]。言葉や表情に比べ、文書中に感情が直接表現されていることよりも、間接的に表現されている事が多い。そのため、感情を読み取る事が直接表現よりも困難で、複雑である。このような間接表現から正確な感情を読み取る手法の例を以下に示す。

感情極性に関する研究には、テキストの文単位で感情を判断する方法と、単語ごとに分け感情の値を算出する方法の2つがある。前者の文単位で判断する方法は、文章を人手で感情のタグ付けを行う方法や、辞書から感情に関係している語をそれぞれ分類し、半自動的に感情コーパスの作成を行う方法等がある。後者の単語ごとに判断する方法は、文章を単語ごとに分け、感情の尺度値を与えた感情語辞書を用いる方法等がある。

### 2.1 菅原らの研究[4]

菅原らの研究では、前に述べた感情極性に関する研究の2つの方法を使用している。感情がそれぞれに関係している語を感情関係語として、9つの感情カテゴリに分類し、データベース化を行う。その単語と分類されたカテゴリ内との結びつきを0~1の値で割り振る事で強度を示している。それぞれ分類された単語を文レベルで解析する。主語と目的語、述語の3点に絞り、スコアを算出し、文全体の感情スコアを算出する。文全体の感情スコアは喜怒哀楽の4つのカテゴリに分類され、最も強度の高い数値を示している。実験結果として、文全体の感情スコアを算出する事ができると同時に、感情語辞書を増やす事でより多くの単語に感情評価を行えることが可能になっている。

### 2.2 高村らの研究[5]

高村らの研究では、それぞれの単語を機械学習の分野などで研究がされているスピンモデルを用い、値の算出を行う。それぞれの単語に上向き、下向きに値を与える。また、隣り合う単語は同じ向きになりやすいという特性が知られている。これを用いて語彙ネットワークを作成し、実験を行う。結果として1,616のポジティブ、1,980のネガティブな英単語の極性値のネットワークが作成されている。少ない数の初期単語のみで高精度の感情極性分類を行うことが可能になっている。



## 2.3 研究の実用性

感情極性の研究はすでに実用化されているものがある。代表的なものとして、つぶやき感情分析がある。これは Twitter と Facebook ユーザーの反応をネガティブ、ポジティブ評価することを可能にしている。2014年11月12日に「Yahoo!リアルタイム検索」で正式に公開された。検索ワードに対して、ツイート数、感情の推移を円グラフで評価するものであり、ツイートした人の感情や反応を視覚的に判断することができる。また、複数の単語を検索ワードに入力することで、その複数の単語が含まれているツイートの感情の推移を判断することも可能である。

例として、以下に図を示す。1月15日時点で、検索ワードを「原発」にし、分析結果を得る。また、それぞれの図の説明をする。

図4は検索ワード「原発」に対するグラフである。グラフ上の「24時間」、「7日間」、「30日間」で表示期間を選択することができ、グラフ下のスライダーを動かすことで、検索ワードに対するツイート数の期間や時間帯を変えることが可能である。一目でツイート数の推移を確認することができ、人気ワードの情報も得ることが可能である。

図5は、検索ワード「原発」に対する推移をグラフで表している。原発という言葉とともにポジティブな言葉が使われている場合、緑色で表され、ネガティブな言葉とともに使われている場合、赤色で表される。どちらでもない場合は、灰色で表される。図5においては、2015年12月17日から2016年1月16日の期間内でネガティブな割合が37%を示し、ポジティブな割合が1%を示している。



図 4. 「原発」 分析グラフ



図 5. 「原発」 に対する感情の推移

## 第3章 関連知識

本章では、本研究を説明するための知識として、Web ニュースデータ、形態素解析、単語感情極性対応表について述べる。

### 3.1 Web ニュースデータ

本研究で用いる Web ニュースデータとは、インターネット上に公開される Yahoo!ニュースなどのテキスト形式のニュースデータを意味する。これらのデータには、ニュースサイト名、タイトル、本文、配信時刻の情報を持っている。この Web ニュースデータを MySQL 上に集める事で、過去から現在までの幅広いデータ解析を行う事ができるという特徴がある。今回使用するデータは、2008年から2014年の7年間に配信された Web ニュースデータの2,638,008件であり、記録されているうちの配信時刻と本文を利用する。表1は収集した Web ニュースデータの配信元一覧である。

### 3.2 形態素解析

言語学で、意味を担う最小の言語要素を形態素と呼ぶ。また、これに対応して自然言語処理では、文を単語ごとに分解し、どの語形変化なのかを解析する処理を形態素解析と呼ぶ[6]。

本研究では、形態素解析の特徴である単語分割、品詞付与、分かち書きの3つを用いる。近年、単語の組み合わせによって新たな単語が無限に生成されている。しかしながら、コンピュータ上の辞書に格納できる単語数は有限である。したがって、既に辞書に登録されている語の単位に入力された分を単語ごとに分解する方法を用いる。これを行うプログラムを単語分割と呼ぶ。単語分割によって分けられたそれぞれの単語に、品詞を与えるプログラムを品詞付与と呼ぶ[6]。

英語の正書法には、単語と単語の間に空白を入れる分かち書きという習慣がある。しかし、日本語の正書法では分かち書きをしない。そのため、日本語の解析を行う際には分かち書きを行うことよりも、単語分割を行うことが課題として挙げられる。本研究では単語ごとの解析を行うため、単語分割によって分けられた文章を分かち書きという手法を用いて、実験を行い易くしている。本研究では、R と MeCab を使用し、解析を行う。

### 3.3 単語感情極性対応表

感情極性値とは、-1 から 1 までの実数値を語彙ネットワークを利用することで計算されたものである。-1 に近いほどネガティブ表現の強い評価、1 に近いほどポジティブ評価の強い評価とここでは定義されている。単語感情極性対応表とは、単語とその単語に対するふりがな、品詞、感情極性値をまとめている表のことである。この表には全 55,125 語の極性値が与えられている。単語の品詞は動詞、名詞、形容詞、副詞の 4 種類であり、それ以外の品詞を持つ単語については値を示していない。表 2 は単語感情極性対応表に書かれている、それぞれ最もポジティブ、最もネガティブな極性値をもつ 7 単語を表している。

表 1. Web ニュースデータの配信元一覧

配信元
ロイター
YOMIURI ONLINE
朝日新聞デジタル
毎日新聞
CNET Japan
Yahoo! ニュース
gooニュース
NHKニュース
産経ニュース
J-CASTニュース

表 2. 単語感情極性対応表

単語	たんご	品詞	極性値
優れる	すぐれる	動詞	1.000000
良い	よい	形容詞	0.999995
喜ぶ	よろこぶ	動詞	0.999979
褒める	ほめる	動詞	0.999979
めでたい	めでたい	形容詞	0.999645
賢い	かしこい	形容詞	0.999486
善い	いい	形容詞	0.999314
⋮	⋮	⋮	⋮
浸ける	つける	動詞	-0.999947
罵る	ののしる	動詞	-0.999961
ない	ない	助動詞	-0.999997
酷い	ひどい	形容詞	-0.999997
病気	びょうき	名詞	-0.999998
死ぬ	しぬ	動詞	-0.999999
悪い	わるい	形容詞	-1.000000

## 第4章 分析手順

本研究では、出現頻度分析を行い、既存の単語感情極性対応表との比較を行う。比較後、単語感情極性対応表にない頻度の高い単語から指数平滑法を用いて計算する。

### 4.1 前提条件

金[7]では文を文節ごとに区切り、ある文節と文節の距離の差をレンジと定義している。レンジの差の最小は3で検証を行っている。また、福島ら[8]では語の組み合わせ距離を **node** と呼ばれる中心語に対して、前後2単語が最少であることを示しており、同時に前後4単語、5単語の方法も可能であることを示している。本研究ではこれらの研究の条件を用い、特定の単語を含めた前後3単語が、特定の単語と強く関わりがあるとする。また、**Web** ニュースデータは時系列データのため、指数平滑法での計算を2008年から順に行う。最新データが2014年11月25日であるため、本研究で求める極性値はこの日時点での極性値とする。 $\alpha$  値を0.1ごとに区切ることによつての極性値の変化を確認するため、一つの特定の単語に対して9つの極性値を求めることを完成とみなす。

### 4.2 出現頻度分析

出現頻度分析とは、**Web** ニュースデータに形態素解析を行い、単語ごとに分割し、単語の出現頻度を表す分析方法である。本研究では、単語感情極性対応表と同様に句読点などの記号は除き、名詞、動詞、形容詞、副詞の4つの品詞に限定し出現頻度分析を行う。理由としては、句読点などの記号や他の品詞には、それ自体に意味が無いためである。出現頻度分析の結果と単語感情極性対応表を比較し、単語感情極性対応表にない単語を頻度の多い順のランキングを作成する。これは、**Web** ニュースデータで頻繁に見かける単語のため、値を算出する事で、その単語そのものの評価を得る優先順位が高いと考えるためである。

### 4.3 指数平滑法

**Web** ニュースデータにはそのニュースの配信された時間が記録されているため、時系列データとして扱われる。本研究では、2014年11月25日時点での極性値を算出するため、一期先の予測値を算出することができる指数平滑法を利

用する.

$$y_{t+1} = \alpha x_t + (1 - \alpha)y_t. \quad (1)$$

(1)式は指数平滑法を表す. (1)式の $y_{t+1}$ は $t + 1$ 期の予測値を,  $x_t$ は $t$ 期の実測値,  $y_t$ は $t$ 期の予測値を表している. 本研究では, 予測値をある特定の単語の感情極性の予測値, 実測値を Web ニュースデータに書かれている特定の単語を含んだ前後 3 単語の感情極性の実測値と定義する. また,  $\alpha$ は平滑化定数であり, 範囲は  $0 \leq \alpha \leq 1$  である[9].

つぎに, (1)式の $\alpha$ 値をそれぞれ変化させる.

➤  $\alpha=0$  の場合,

$$y_{t+1} = 0. \quad (2)$$

➤  $\alpha=0.5$  の場合,

$$y_{t+1} = 0.5x_t + 0.25x_{t-1} + 0.125x_{t-2} + \dots. \quad (3)$$

➤  $\alpha=1$  の場合,

$$y_{t+1} = x_t. \quad (4)$$

(2)式は $\alpha = 0$  の場合を表し, (3)は $\alpha = 0.5$ , (4)は $\alpha = 1$  の場合を表している. 指数平滑法の特徴として, (2), (3), (4)式を比較すると  $\alpha$ が 0 に近くなればなるほど, 過去の影響をより受けやすくなる. 反対に,  $\alpha$ が 1 に近くなればなるほど直前の影響を受けやすくなる. つまり,  $\alpha$ の変化により過去の影響度合いが変化するため, 算出値に差が生まれる事が分かる. 本研究では, 時系列データを用い, 2014 年 11 月 25 日時点での予測値を算出し, 比較を行うことを目的にしているため, 考察を行い易いという利点がある.

## 第5章 分析結果

本章では分析結果として、頻度分析、指数平滑法での計算結果を示す。

### 5.1 頻度分析結果

全 2,638,008 件の Web ニュースデータ本文を MySQL に格納し、形態素解析を行った。形態素解析により、単語ごとに分割し、11,828,438 語を得ることができた。表 3 は、得ることのできた単語の頻度を降順で表したものである。最も頻度の高い単語は、「する」で 12,563,830 回使われていることがわかる。日本語文章の終わり方として、「～する、～いる、～ます、～ある」が多いため、これらを表す動詞の中では、頻度の高い値が表されている。名詞では、時間を表す「日、年」が高い頻度を表していることがわかり、ニュースにとって時間を書くことは必要であることがわかる。また、人物を表す「人、者」も頻度が高くなっていることがわかる。一方表 3 を見ると、副詞や形容詞は動詞や名詞に比べ使用頻度が低いことがわかる。形容詞において、最も高い頻度を表す単語は「ない」で全単語の 31 番目。副詞においては「初めて」で 148 番目に位置づけしている。

表 3. ニュースデータの単語頻度

単語	品詞	品詞2	頻度
する	動詞	自立	12,563,830
いる	動詞	非自立	4,643,426
日	名詞	接尾	3,415,020
れる	動詞	接尾	2,890,121
こと	名詞	非自立	2,135,687
なる	動詞	自立	1,666,154
人	名詞	接尾	1,535,409
年	名詞	接尾	1,439,984
ある	動詞	自立	1,337,931
者	名詞	接尾	1,199,841
市	名詞	接尾	943,574
日本	名詞	固有名詞	896,317
の	名詞	非自立	827,871
さん	名詞	接尾	807,623
円	名詞	接尾	769,191
県	名詞	接尾	748,905
られる	動詞	接尾	689,244
的	名詞	接尾	686,396
⋮	⋮	⋮	⋮



表3と単語感情極性表を比較し、表3にあり、単語感情極性対応表にはない単語を、頻度順に表したものが表4である。これは、頻度の高い単語を極性値算出の優先順位が高いと定義するためである。本研究では、書かれている頻度が100,000回以上のすべて単語、「明らか」、「大臣」、「原発」、「分かる」、「使う」、「見る」、「強い」、「多い」、「目指す」、「計画」、「出す」、「社長」、「持つ」、「状況」、「県警」、「民主党」、「会長」、「線」、「選」の計19単語を特定単語とした。最も高い頻度を示す単語は「明らか」であり、199,665回本文に書かれていることがわかる。一方、特定単語の中で最も頻度の低い単語は、「選」であり、118,941回本文に書かれている。

表4. 新規極性値算出の単語類

単語	品詞	品詞2	頻度
明らか	名詞	形容動詞語幹	199,665
大臣	名詞	一般	196,750
原発	名詞	一般	183,118
分かる	動詞	自立	174,352
使う	動詞	自立	162,706
見る	動詞	自立	155,719
強い	形容詞	自立	152,824
多い	形容詞	自立	147,395
目指す	動詞	自立	146,913
計画	名詞	サ変接続	145,500
出す	動詞	自立	144,621
社長	名詞	一般	140,249
持つ	動詞	自立	139,104
状況	名詞	一般	138,561
県警	名詞	一般	137,162
民主党	名詞	固有名詞	133,208
会長	名詞	一般	124,238
線	名詞	接尾	122,734
選	名詞	接尾	118,941
⋮	⋮	⋮	⋮

## 5.2 指数平滑計算結果

表 4 で示した単語について、4.3 の方法で計算を行った結果が表 5 である。表 5 の結果について評価を行う。全体的に $\alpha$ 値に関係なくネガティブの値をとる単語が多いことがわかる。ここでそれらの単語がポジティブかネガティブかの評価を行う。最も 1 に近い値を示す単語は「目指す」であり、 $\alpha = 0.9$  のとき 0.855 というポジティブ極性値をとっている。一方、最も -1 に近い値を示す単語は「使う」であり、 $\alpha = 0.9$  のとき -0.871 というネガティブ極性値をとっている。 $\alpha$  を変化させることによって値が最も変化する単語は、「原発」であり、最も高い値と低い値の差は約 1.3 あることがわかる。いつの Web ニュースデータを重要とするかによってポジティブになることもあれば、ネガティブになることもある単語であることがわかる。 $\alpha$  が 0.1 のとき、ポジティブな極性をもつ単語は「明らか」、「選」、「目指す」の 3 単語のみとなった。また、これらの 3 単語は  $\alpha$  値が変化してもネガティブに傾く事がないため、ポジティブな意味を持つ単語と判断できる。一方、ネガティブな極性をもつ単語は「原発」、「会長」、「県警」、「使う」、「持つ」、「状況」、「大臣」、「分かる」、「民主党」、「強い」、「計画」、「見る」、「社長」、「出す」、「線」、「多い」の 16 単語となり、そのうち「会長」、「県警」、「使う」、「持つ」、「状況」、「大臣」、「分かる」、「民主党」、「強い」、「計画」、「見る」、「社長」、「線」、「多い」の 14 単語は  $\alpha$  値が変化してもポジティブに傾く事がないため、ネガティブな意味をもつ単語と判断できる。しかし、「原発」、「出す」の 2 単語については、 $\alpha$  値を変化させることで両極性に傾くため、ポジティブかネガティブのどちらかと判断する事は容易でない。つまり、時期や Web ニュースデータによって値が変化する単語と判断するべきだろう。

表 5. 指数平滑計算結果

$\alpha$ 単語	0.1	0.3	0.5	0.7	0.9
明らか	0.010751	0.137054	0.344425	0.585968	0.838905
大臣	-0.142776	-0.192750	-0.188706	-0.175903	-0.158253
原発	-0.382990	-0.185429	0.128065	0.461735	0.812110
分かる	-0.115514	-0.238843	-0.257139	-0.264886	-0.258011
使う	-0.579815	-0.521613	-0.638017	-0.754422	-0.870826
見る	-0.136793	-0.193037	-0.270743	-0.301784	-0.330748
強い	-0.354533	-0.332631	-0.365436	-0.510161	-0.766629
多い	-0.326266	-0.466363	-0.503191	-0.529426	-0.579691
目指す	0.065855	0.362095	0.507542	0.657594	0.854550
計画	-0.276461	-0.172817	-0.105232	-0.092384	-0.140742
出す	-0.311156	-0.112115	0.021241	0.118811	0.214693
社長	-0.225020	-0.242008	-0.267119	-0.250897	-0.190423
持つ	-0.404040	-0.408433	-0.399646	-0.390859	-0.382072
状況	-0.313652	-0.433367	-0.449455	-0.411005	-0.318471
県警	-0.260960	-0.222550	-0.252648	-0.274025	-0.279790
民主党	-0.164294	-0.194296	-0.251284	-0.287606	-0.276743
会長	-0.258643	-0.251398	-0.224798	-0.198447	-0.181104
線	-0.201750	-0.223829	-0.264152	-0.289306	-0.274751
選	0.165141	0.212040	0.291439	0.364531	0.391479

## 第6章 おわりに

本研究では、Web 上のニュースデータに書かれている単語がポジティブ、ネガティブのどちらかの極性をもつのか検証を行った。約7年分のWeb ニュースデータの本文を用い時系列解析を行った。これらのデータに形態素解析を行い、意味を持つ最小の単語ごとに分割した。さらに、分割された単語の頻度分析を行い、約7年間の中でより多く使われている19単語を極性値の求める対象とした。これらの対象となった単語を特定の単語とし、約7年間のWeb ニュースデータの中に書かれているすべての特定の単語を含む前後3単語を対象とする。特定の単語以外に単語感情極性値を与え、指数平滑法を用い計算を行い、 $\alpha$ の変化とともに特定の単語の極性値がどのように変化するかを検証を行った。

上記の手順を行い、得ることができた結果は第5章で述べた。出現頻度分析によって、Web ニュースデータで使われている単語の頻度の多い順のランキングを作成した。その中で、単語感情極性対応表にない単語を選出し、頻度の多い順でランキングを作成した。上位単語に対し、極性値を算出した。 $\alpha$ を変化させる事でポジティブの極性を得る単語、ネガティブの極性を得る単語、ポジティブ・ネガティブの両極性を得る単語の3パターンあることがわかった。また、どちらに強い極性をもつのかについて、値を算出する事で判断する事が容易となった。

本研究では、指数平滑法にて極性値を算出する方法を用い、 $\alpha$ ごとに値を求める事ができた。しかし、最終的には $\alpha$ を固定し値を求める必要がある。新規単語の極性値を求める方法は、第2章で述べたようにいくつかの手法がある。そのため、指数平滑法にて極性値を算出するためには、他の手法と比較を行い、 $\alpha$ を固定する必要がある。

## 参考文献

- [1] 総務省:情報通信白書 ICT サービスの利用動向 平成 27 年版,  
<http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h27/pdf/n7200000.pdf> (2016 年 1 月 22 日確認)
- [2] 日本新聞協会:調査データ 新聞の発行部数と世帯数の推移,  
<http://www.pressnet.or.jp/data/circulation/circulation01.php>  
(2016 年 1 月 22 日確認)
- [3] ICT 総研:モバイルニュースアプリ利用動向調査,  
<http://ictr.co.jp/report/20141104000070.html> (2016 年 1 月 22 日確認)
- [4] 菅原 久嗣, 石塚 満, 感情語辞書を用いた日本語テキストからの感情抽出.
- [5] 高村大也, 乾孝司, 奥村学, スピンモデルによる単語の感情極性抽出, 情報処理学会論文誌ジャーナル, Vol.47,No.2 pp.627-637, 2006.
- [6] 松本 裕治, 影山 太郎, 永田 昌明, 齋藤 洋典, 徳永 健伸, 単語と辞書, 岩波書店, 1997.
- [7] 金 明哲, 文節の係り受け距離の統計分析, 社会情報, Vol.5,No.2 pp.1-11, 1996.
- [8] 福島 一人, 定義されたコロケーションとその有用性:コロケーションの辞書執筆の見地から, 情報研究, Vol.36 pp.253-283, 2007.
- [9] 石村 貞夫, 金融・証券のための市場予測と回帰モデル・時系列モデル, 東京図書株式会社, 2000.

# 付録

表 6. 指数平滑計算結果一覧

単語 \ $\alpha$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
明らか	0.010751	0.062267	0.137054	0.233857	0.344425	0.463020	0.585968	0.711383	0.838905
大臣	-0.142776	-0.179394	-0.192750	-0.192979	-0.188706	-0.182829	-0.175903	-0.167814	-0.158253
原発	-0.382990	-0.314093	-0.185429	-0.033110	0.128065	0.293210	0.461735	0.634430	0.812110
分かる	-0.115514	-0.210917	-0.238843	-0.249439	-0.257139	-0.262786	-0.264886	-0.263036	-0.258011
使う	-0.579815	-0.463411	-0.521613	-0.579815	-0.638017	-0.696220	-0.754422	-0.812624	-0.870826
見る	-0.136793	-0.140973	-0.193037	-0.239836	-0.270743	-0.289162	-0.301784	-0.314521	-0.330748
強い	-0.354533	-0.349176	-0.332631	-0.335379	-0.365436	-0.423742	-0.510161	-0.624541	-0.766629
多い	-0.326266	-0.416156	-0.466363	-0.490495	-0.503191	-0.514383	-0.529426	-0.550874	-0.579691
目指す	0.065855	0.248594	0.362095	0.440369	0.507542	0.577617	0.657594	0.749980	0.854550
計画	-0.276461	-0.222556	-0.172817	-0.132893	-0.105232	-0.091480	-0.092384	-0.108526	-0.140742
出す	-0.311156	-0.206847	-0.112115	-0.038068	0.021241	0.071860	0.118811	0.165764	0.214693
社長	-0.225020	-0.224982	-0.242008	-0.258256	-0.267119	-0.265065	-0.250897	-0.225453	-0.190423
持つ	-0.404040	-0.412826	-0.408433	-0.404040	-0.399646	-0.395253	-0.390859	-0.386466	-0.382072
状況	-0.313652	-0.398249	-0.433367	-0.448530	-0.449455	-0.436899	-0.411005	-0.371665	-0.318471
県警	-0.260960	-0.222248	-0.222550	-0.236549	-0.252648	-0.265608	-0.274025	-0.278444	-0.279790
民主党	-0.164294	-0.176323	-0.194296	-0.222098	-0.251284	-0.274739	-0.287606	-0.287874	-0.276743
会長	-0.258643	-0.263929	-0.251398	-0.238153	-0.224798	-0.211288	-0.198447	-0.187747	-0.181104
線	-0.201750	-0.212729	-0.223829	-0.242612	-0.264152	-0.281295	-0.289306	-0.286593	-0.274751
選	0.165141	0.197881	0.212040	0.247092	0.291439	0.332975	0.364531	0.383579	0.391479