

PN 値に基づく名言検索手法の提案

The Search for Best Quotes based on the value of PN

高岡 幸一†
Kouichi Takaoka

灘本 明代‡
Akiyo Nadamoto

1 はじめに

現在の社会は、景気の低迷に伴い人件費が大幅に削減される等経済的にも精神的にも余裕がなくなりストレスをかかえる人が多くなっている。また現代人は、受験や就職活動、結婚活動など競争社会の中で生きており、ストレスをかかえる人が多くなっている。一方で、インターネットの普及により、一般の人々にとってもインターネットを利用した情報収集や SNS に代表されるような情報共有が盛んに行われるようになっている。しかしながら、現在これらインターネットを利用して人々を元気にするシステムは少ない。そこで、身近になったインターネットを利用し、ストレスをかかえている現代人を元気にするシステムを構築したいと考え、人々の気持ちをポジティブにする情報の検索手法の提案を行う。

具体的には、ユーザの「今の気分」と「なりたい気分」からセンチメント距離を求め、そのセンチメント距離に合った情報を検索して提示するシステムの提案を行う。本論文では、この提示する情報として「名言」を用いる。我々が名言を用いた理由として、名言は偉人達が残した言葉で人々を元気づけたり、勇気を与えたりする言葉であるからである。センチメント距離に合った名言をユーザに提示するために、名言のポジティブ・ネガティブ度を示す名言 PN 値を求める。我々の提案する名言 PN 値は、まず単語の出現頻度を用いて名言単語 PN 値を求め、その名言単語 PN 値から名言 PN 値を求める。実際には、Web 上にある名言を使いユーザ実験により、ポジティブ名言 350 個とネガティブ名言 350 個を取得する。ポジティブ名言とネガティブ名言各々から単語の出現頻度を求め、それを用いて名言単語 PN 値を作成する。そして、この名言単語 PN 値を用いて名言の値である名言 PN 値を求める。さらに、センチメント距離と名言 PN 値を照らし合わせて、センチメント距離に近い名言を提示する。本研究ではユーザをポジティブな気分にするを目的とし、PN 値に基づく名言の検索手法を提案する。

以下、2 章では関連研究を、3 章ではセンチメント距

離の求め方を、4 章では名言 PN 値の求め方を述べる。5 章ではまとめと今後の課題について述べる。

2 関連研究

高村ら [1] は単語の感情極性（肯定か否定）を判定する方法を提案している。提案手法では、感情極性を電子のスピンとみなし、語釈文を用いて語彙ネットワークをスピン系でモデル化している。平均場近似を利用してスピン系の状態を近似的に求めることにより、単語の感情極性を判定している。本研究ではこの感情極性を参考にしている。しかしながら、この感情極性は否定的な言葉が多いため、本研究の目的であるユーザの気分を元気にすることは難しいと考え、我々は新しいスコアである名言単語 PN 値を作成することを考案する。

河合ら [2] は、複数のニュースサイトから収集した記事を興味語と平均印象ベクトルのペア集合からなるユーザプロフィールに基づいて分類し、ユーザの興味と好みに合わせて記事を優先的に提示可能な MPV Plus を提案している。MPV Plus は、ユーザの閲覧履歴からユーザの興味のある語である興味語とユーザの好む印象である平均印象ベクトルを決定し、収集した記事のうち、この分類基準に合致する記事を優先的に提示している。河合ら [2] は、記事の印象を 4 つの尺度（「明るい」「暗い」「承認」「拒否」「緩和」「緊張」「怒り」「恐れ」）を用いて情報を提示している。それに対して、我々の研究はユーザの気分をポジティブにすることが目的であるので、PN 値（「ポジティブ」「ネガティブ」）を用いて研究を進めていく。

藤村ら [3] は、評判の肯定・否定分類に取り組み Web 全体からの評判抽出を行うシステムを構築している。肯定的（否定的）な評判には、肯定的（否定的）な概念を持った語が多く含まれているという仮定を基に、肯定的な評判と否定的な評判の差を取るスコアリング手法を提案している。我々は、藤村ら [3] が提案したスコアリング手法を基に名言スコアリング手法を提案する。本論文では、名言スコアリング手法の有用性についても検証する。

一方、情報発信者（話し手、書き手）から情報受信

† 甲南大学大学院 自然科学研究科、Konan University

‡ 甲南大学 知能情報学部、Konan University

者（聞き手，読み手）に伝達される情報のうち「もの，行為，状態」といった命題や「信念，意図」といった情報発信者の命題に対する態度に関しては，古くから研究されている．しかし，情報発信者が発信したコンテンツを見たり聞いたりすることによって情報受信者が感じる印象を抽出し，利用しようという研究は，まだ少ない．そこで熊本ら [4] は新聞記事を対象に，情報発信者が発信したテキストから情報受信者が感じる印象を抽出する印象マイニング手法を提案している．これに対し，本研究では情報受信者の感情を対象にして検索を行う点が異なる．

3 センチメント距離の求め方

一般に気分とは「楽しい」「悲しい」といったように形容詞を用いる場合が多いが，本研究では気分にあった名言を提示するため，気分を示す単語を数値化する．そして「今の気分」と「なりたい気分」の値を求め，ふたつの気分からセンチメント距離を求めることを提案する．

3.1 単語感情極性対応表

本研究では，センチメント距離を測る方法として，高村ら [1] の単語感情極性対応表の感情極性値を用いる．感情極性とは，その語が一般的に良い印象（positive）を持つか悪い印象（negative）を持つかを表した二値属性である．例えば「良い」「美しい」などは positive な極性「悪い」「汚い」などは negative な極性を持つ．感情極性値は，語彙ネットワークを利用して自動的に計算されたものである．感情極性値は，もともと二値属性であるが，-1 から +1 の実数値を割り当てられている．-1 に近いほど negative，+1 に近いほど positive とする．単語感情極性対応表の単語は「岩波国語辞書（岩波書店）」を使用し作成したものであり，単語の品詞は，名詞，動詞，形容詞，副詞，助動詞が含まれている．

3.2 センチメント距離

名言を求める際にセンチメント距離を用いる．我々が提案するセンチメント距離とは「なりたい気分」の値から「今の気分」の値を引いて，その絶対値を 2 で割った値である．距離を測る方法も，高村ら [1] の単語感情極性対応表の感情極性値を用いる．ここで単語感情極性対応表は 55,125 語あり非常に単語数が多い．そのため我々はその中から，60 語の単語（楽しい，悲しいなど）を使用する．そして中村明の感情表現辞書 [5] を用いて，その単語を 10 種の感情カテゴリに分類する．10 種の感情カテゴリとそれに含まれる単語の例

を表 1 に示す．

「今の気分」はこの 10 種の感情カテゴリの中にある 60 単語の中から 1 つ選択する．一方「なりたい気分」は 3 種の感情カテゴリ「喜」「安」「好」の中に含まれる単語を 1 つ選択する．「なりたい気分」の感情カテゴリを 3 つに制限したのは，ユーザがネガティブな言葉を選んでしまうとネガティブな名言が出てしまい，ユーザの気分を悪くしてしまう可能性があるためである．そこで，ポジティブなカテゴリ 3 種を対象とする．この 3 種のカテゴリは，経済産業省・産業技術総合研究所の「生活習慣の継続的モニタリングと行動変容に応じた健康改善サービスの検証」 [6] で用いられた値を参考とする．

この検証によると，10 種のカテゴリの中で「喜」，「好」，「昂」の 3 つのカテゴリがポジティブなカテゴリであることを示している．しかしながら，我々が作成した 10 種の感情カテゴリの「昂」カテゴリは，ネガティブな単語しか入っていないため，ポジティブな言葉が入っている「安」カテゴリを「昂」カテゴリの代わりに「なりたい気分」のカテゴリとする．センチメント距離は，0 から +1 の実数値を割り当てる．この値が +1 に近いほど距離が遠く，よりポジティブな名言を提示する．センチメント距離を測る式 (1) とセンチメント距離 (SD) の例 1 と例 2 を以下に示す．

$$SD = \frac{|(\text{なりたい気分の値}) - (\text{今の気分の値})|}{2} \quad (1)$$

例 1: なりたい気分 「嬉しい」 (値は 0.998871)
今の気分 「寂しい」 (値は -0.9990)

$$SD = |(0.998871) - (-0.9990)| \div 2 = 0.9989505$$

この場合は，センチメント距離が 1 に近いのでよりポジティブな名言を提示する．これによりユーザの気分をポジティブにすることができる．

例 2: なりたい気分 「うっとり」 (値は 0.36674)
今の気分 「むしゃくしゃ」 (値は -0.0523)

$$SD = |(0.36674) - (-0.0523)| \div 2 = 0.41904$$

この場合は，センチメント距離が 0 に近いので名言距離が 0 に近い名言を提示する．この場合は，距離が短いので，少し気分をポジティブにすることができる．

表 1: 10 種の感情カテゴリ

感情カテゴリ	単語
喜	楽しい, 面白い, 良い
怒	狂おしい, 腹立たしい
哀	哀しい, 痛い, 寂しい
怖	怖い, ときどき, 不安
恥	恥しい, もじもじ
好	恋しい, 愛しい, うきうき
厭	憎い, 暗い, 辛い
昂	もどかしい, 歯痒い, 苛立つ
安	のんびり, すっきり, さっぱり
驚	ぼんやり, 驚喜, 仰天

4 名言 PN 値

センチメント距離に合った名言をユーザに提示するために、各名言も数値化する。ここで、名言の値を求めるために、高村ら [1] の単語感情極性対応表の感情極性値を使用し、名言に含まれる単語に値を与え名言ごとの値を求めることを考えた。しかしながら、この単語感情極性対応表の感情極性値はマイナスな単語が多くあり、ネガティブな名言ばかりになった。そこで本研究では、ポジティブ名言とネガティブ名言各々の単語の出現頻度を用いて名言単語 PN 値を求め、その名言単語 PN 値から名言 PN 値を求めることを提案する。

4.1 名言実験

名言 PN 値を求めるために、Web サイトからポジティブ名言とネガティブ名言を取得する。はじめに、Web サイトから人手によりポジティブ名言の候補 500 個、ネガティブ名言の候補 500 個を取得する。

次に、被験者 8 名により、先程取得したポジティブ名言の候補 500 個、ネガティブ名言の候補 500 個を被験者の主観でポジティブであるかネガティブであるかどうかでもないかを判断する。そして、被験者 8 名の内 5 名以上からポジティブであると判断された名言 377 個をポジティブ名言、ネガティブであると判断された名言 434 個をネガティブ名言とした。次に、ポジティブ名言 377 個の中からより多くの被験者からポジティブと判断された 350 名言とネガティブ名言 434 個の中からより多くの被験者からネガティブと判断された 350 名言を用いて名言単語 PN 値を求める。表 2 に得られたポジティブ名言、ネガティブ名言の一例を示す。

表 2: 取得した名言の一例

ポジティブ名言
人間は負けたら終わりなのではない。辞めたら終わりなのだから前を向いて期待していれば「できる」可能性が見えてくる
失敗は成功への道しるべ
勤勉と熟達があれば不可能なことはほとんどない
いつも晴れ晴れとした気持ちでいる人は長生きする
ネガティブ名言
失敗を極度に恐れることは、失敗するよりも悪い。
愚か者は、金を持って死んでいくために、貧乏で暮らす人々が自分に調和してくれるように望むのは非常に愚かだ
怒りは無謀をもって始まり、後悔をもって終わる
死はありとあらゆる悲哀の週末なり

4.2 名言単語 PN 値

名言単語 PN 値とは、名言に含まれる単語のポジティブ、ネガティブの度合いを示す値である。名言実験で得られたポジティブ名言、ネガティブ名言の名詞、形容詞、副詞、動詞の単語の出現頻度 tf 値を求める。名詞と動詞と形容詞には、意味を持たない単語がいくつかあるので、その単語をストップワードとし、そのストップワードを除く。そして品詞が副詞と形容詞と動詞である場合は、単語の原型を用いて tf 値を求める。その tf 値を式 (2) に示すようにポジティブ名言に含まれる単語の名言 PN 値 y は $0 \sim +1$ 、ネガティブ名言に含まれる単語の名言 PN 値 y は $-1 \sim 0$ の範囲で正規化する。各品詞で一番頻度の多い単語を基準値 (ポジティブなら 1、ネガティブなら -1) とする。名言単語 PN 値を求める式 (2) を下に示す。

$$y = \frac{X_P}{X_{Pmax}} + \frac{X_N}{X_{Nmax}} \quad (2)$$

X_P は、ポジティブ単語の出現頻度を表し、 X_{Pmax} は、 X_P を含む品詞の中で出現頻度が最大の単語の出現頻度を表す。 X_N は、ネガティブ単語の出現頻度を表し、 X_{Nmax} は、 X_N を含む品詞の中で出現頻度が最大の単語の出現頻度を表す。ポジティブ名言にもネガティブ名言にも単語が含まれている場合は、二つの数値を足して単語のスコアを求める。名言単語 PN 値はあらかじめ求めておき、単語データベースに保存する。求めた名言単語 PN 値の単語総数は、1933 語である。

4.3 名言 PN 値

1つの名言のポジティブ、ネガティブの度合いを名言 PN 値と定義する．名言 PN 値を求めるためにさきほど求めた名言単語 PN 値を用いる．また，藤村ら [3] のスコアリング手法を基に名言スコアリング手法の式 (3) を提案する．

$$Score(m_i) = \frac{P_{num} + N_{num}}{P_{num} - N_{num}} \quad (3)$$

m_i は名言 i を表し， P_{num} は m_i にポジティブな言葉が出現する回数であり，同様に N_{num} は m_i にネガティブな言葉が出現する回数である．

以前までの我々の研究 [7] では，名言単語 PN 値と名言スコアリング手法を用いて名言 PN 値を求めてきた．しかしながら，名言スコアリング手法を使うと単語ごとの重みを考慮することができないという問題が出てきた．そこで，本研究では，名言単語 PN 値のみを使い名言 PN 値を求める場合と名言 PN 値と名言スコアリング手法を使い名言 PN 値を求める場合の比較を行う．

4.4 名言 PN 値の比較実験

名言 PN 値の有用性を測る実験として，4つの手法を用いて再現率と適合率による比較実験を行った．実験方法は，まず Web 上から 200 個の名言を人手により取得する．その 200 個の名言を人手でポジティブ名言，ネガティブ名言，どちらでもない名言の 3 種類に分類する．一方システムでは，4つの手法を用いて名言 PN 値を求める．名言 PN 値が 0 より大きい名言をポジティブ名言，名言 PN 値が 0 より小さい名言をネガティブ名言，名言 PN 値が 0 である名言はどちらでもない名言とする．比較実験では，人手で判断した名言とシステムで判定した名言を照らし合わせ，ポジティブ名言，ネガティブ名言を対象とした再現率と適合率をそれぞれ求めた．比較実験に用いた 4つの手法は以下の通りである．

- 手法 A: 名言単語 PN 値のみを用いた手法 (式 (2) のみ使用)
- 手法 B: 名言単語 PN 値と名言スコアリング手法を用いた手法 (式 (2) と (3) を使用)
- 手法 C: 名言単語 PN 値から名詞を抜いた名言単語 PN 値を使う．
- 手法 D: 名言単語 PN 値から名詞を抜いた名言単語 PN 値と名言スコアリング手法を使う．

手法 A と手法 B，手法 C と手法 D を比較し，名言スコアリング手法を使うべきなのかを実験した．手法 A と

表 3: 比較実験の結果 (ポジティブ名言)

	再現率	適合率
手法 A	72 %	54 %
手法 B	68 %	55 %
手法 C	62 %	47 %
手法 D	58 %	50 %

表 4: 比較実験の結果 (ネガティブ名言)

	再現率	適合率
手法 A	62 %	52 %
手法 B	50 %	53 %
手法 C	40 %	46 %
手法 D	34 %	46 %

手法 C，手法 B と手法 D を比較し，名言単語 PN 値の名詞を使うべきなのかを実験した．ポジティブ名言を対象とした再現率と適合率を表 3，ネガティブ名言を対象とした再現率と適合率を表 4 に示す．

4.5 名言 PN 値の比較実験結果

実験結果より名詞を抜いた名言単語 PN 値を使った手法 C と手法 D は，名詞を入れた名言単語 PN 値を使った手法 A と手法 B に比べてそれぞれ悪い結果となったことから，名言単語 PN 値には名詞を入れた方が良かった．また，名言スコアリング手法の有用性を比較した手法 A と手法 B では，再現率は手法 A の方が良いが適合率は手法 B の方が良い結果となった．適合率は，ポジティブ名言，ネガティブ名言の場合でも差は小さいが，再現率は，ポジティブ名言の場合は差が小さいがネガティブ名言の場合は，名言スコアリング手法を使わない方が大きく良い結果になった．また名言スコアリング手法を使うと単語の重みを考慮できない問題がある．これらのことから名言スコアリング手法を使わず名言単語 PN 値のみを使い名言 PN 値を求めるのが良かった．

5 まとめと今後の課題

本研究では，名言単語 PN 値を作成し，名言 PN 値とセンチメント距離に合った名言をユーザに提示する手法の提案した．本研究の目的は「ユーザの気分をポジティブにすること」である．そのため「なりたい気分」と「今の気分」から求めるセンチメント距離を提案し，

これを用いてユーザの気分を数値化した。また、名言単語 PN 値や名言スコアリング手法を提案し、ユーザの気分をポジティブにする名言を提示する手法を提案した。その他の今後の課題は以下の通りである。

- 名言 PN 値の有用性

本研究では、ポジティブな名言とネガティブな名言をはっきりと区別するために実験を行い、名言をポジティブ名言、ネガティブ名言に分類して名言単語 PN 値を作成した。そして、この名言単語 PN 値を用いて名言 PN 値を求めた。つまりポジティブな名言とは、ポジティブな単語が多く含まれているものであり、ネガティブな名言とはネガティブな単語が多く含まれていることを前提として本研究を進めてきた。しかし、ポジティブな言葉しか使われていない名言が本当にポジティブであるかは断定できない。良い名言とは、ポジティブな言葉とネガティブな言葉が良い具合に含まれているため、心に響く名言になるのかもしれない。心に響く名言を調べるために、さらなる実験を行い、名言 PN 値の有用性を測る必要がある。

- PN 以外のベクトルの考慮

本研究では、PN のベクトルを用いて、ユーザの気分にあった情報を提示してする手法の提案を行った。しかし、PN だけではユーザの気分にあった情報を提示するのは難しいのではないかという問題が出てきた。今後は、PN 以外のベクトルを用いて、更にユーザの気分にあった情報を提示していきたい。

- 名言以外の情報提示

本研究では、提示する情報として名言に注目した。今後は、名言だけではなくニュースや音楽や動画なども提示していきたい。次回の研究では、音楽に注目していきたいと考える。

謝辞

本研究の一部は、平成 22 年度科研費特定領域研究域「コミュニティ型コンテンツのコンテンツホール検索に関する研究」(課題番号: 21013044, 代表: 灘本明代) によるものである。ここに記して謝意を表します

参考文献

[1] 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. 02, pp. 627-637, 2006.

[2] 河合由起子, 熊本忠彦, 田中克己. 印象と興味に基づくユーザ選好のモデル化とニュースポータルサイトへの応用. 情報処理学会 情報科学技術レターズ (Information Technology Letters), Vol.4, No.LD-001, pp.65-68(2005).

[3] 藤村滋, 豊田正史, 喜連川優. 電子掲示板からの評判表現および評判情報の抽出. 第 18 回人工知能学会全国大会, CD-ROM, 4 pages, 2004.

[4] 熊本忠彦, 河合由起子, 田中克己. テキスト印象マイニング手法の開発と評価. Web とデータベースに関するフォーラム (WebDB Forum 2009), CD-ROM, 8 pages, 2009.

[5] 中村明. 感情表現辞書. 東京堂出版, 1993.

[6] 経済産業省・産業技術総合研究所. 生活習慣の継続的モニタリングと行動変容に応じた健康改善サービスの検証. 「平成 20 年度サービス研究センター基盤整備事業」に係る適用実証委託事業, 2008.

[7] 高岡幸一, 灘本明代. 気分距離に基づく名言検索手法の提案. 情報処理学会全国大会, A9-5, 2010 年 3 月.