

文のポジティブ・ネガティブ値を考慮した 名言検索手法の提案

高岡 幸一^{†1} 灘本 明代^{†2}

インターネットの普及により、誰でも容易に情報を取得できるようになった。しかしながら、ユーザの気分にあった情報を検索し提示するシステムは少ない。そこで、本研究では、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測り、そのセンチメント距離に合った情報を提示し、ユーザの気分をポジティブにすることを目的としたシステムの提案を行う。本論文では、提示対象の情報に名言を用いる。具体的には、名言の文構造からポジティブ/ネガティブ値(P/N 値)を求める手法を提案する。そして、実験により人が感じるPN判定とシステムが求めるPN判定をより近づけ、センチメント距離に合った名言を提示するシステムを構築する。

The Quotes Search System based on Positive/Negative Value of Sentences

KOUICHI TAKAOKA^{†1} and AKIYO NADAMOTO^{†2}

With the rapid advance of the Internet, everybody can get the information from it easily. There are, however, few system which extracts and presents the information suitable for user's sentiment. In this paper, we present the system that makes users positive to extract and present positive quotes based on calculating user's sentiment. In particular, we propose the method of calculating Positive-Negative(P/N) value for quotes based on structure of it. Then we have user experiment to measure the availability of our proposed method.

^{†1} 甲南大学大学院 自然科学研究科
Konan University

^{†2} 甲南大学 知能情報学部
Konan University

1. はじめに

近年景気低迷に伴い、不況、リストラ、就職難、賃金カットなどネガティブなニュースが多く取り上げられるようになってきている。そのため、現在の社会はストレス社会とも言われ、大きな社会問題の一つとなっている。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コンピュータは身近なものになっている。これに伴い、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きている。また、ユーザの過去の履歴から情報を推薦するシステムも多数ある。しかしながら、ユーザの気分を考慮したシステムが少ない。そこで、身近になったコンピュータシステムを利用し、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。本論文では、名言を分析し、複文から構成される名言に対応した名言PN値の式を提案し、よりユーザの気分にあった名言を提示する。

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章では単語PN値の求め方を、5章では名言PN値の求め方について述べる。6章では否定語への対応について述べ、7章では名言PN値の比較実験について述べ、8章ではプロトタイプシステムについて述べ、9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する。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単語PN値を作成することを提案する。

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。

一方、情報発信者(話し手、書き手)から情報受信者(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

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、本研究ではユーザの気分に合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、ユーザの「今の気分」と「なりたい気分」の値を求め、この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求める方法として、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

我々が提案するセンチメント距離とは、ユーザの入力する「なりたい気分」と「今の気分」から求める。この時に各気分の値として、高村ら¹⁾の単語感情極性対応表の感情極性値を用いる。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。そしてユーザが気分をより容易に選択するために、中村明の感情表現辞書⁵⁾を用いて、その 60 語の単語を 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、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。センチメント距離を求める式を以下に示す。ここでは、「なりたい気分」と「今の気分」に感情極性値を用いるため、センチメント距離は、0 から +1 の数値となる。この値が +1 に近いほど距離が遠いのでなりたい気分の差が大きいと考え、よりポジティブな名言を検索し提示することとする。

$$SD = \frac{|S_2 - S_1|}{2} \tag{1}$$

S_1 は今の気分の数値を表し、 S_2 はなりたい気分の数値を表す。例えば、なりたい気分が「嬉しい」場合、感情極性値のは 0.998871 であり、今の気分が「寂しい」場合その値は -0.9990 である。これらの値を式(1)に代入すると、センチメント距離は 0.9989505 となる。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

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, どきどき, 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

によりユーザの気分をポジティブにすることができる。

また、なりたい気分が「うっとり」の場合、感情極性値は 0.36674 であり、今の気分が「むしゃくしゃ」の場合その値は -0.0523 であり。これらの値も同様に式 (1) に代入すると、センチメント距離は 0.41904 となる。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。

4. 単語 PN 値の求め方

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、各名言のポジティブ、ネガティブの値から PN 値を求める。ここで、名言の値である名言 PN 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単語 PN 値を求め、その単語 PN 値から名言 PN 値を求めることを提案する。

4.1 予備実験

単語 PN 値を求めるために、ポジティブ名言とネガティブ名言を決定する実験を以下の手順にて行った。

- (1) 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。

表 2 取得した名言の一例

ポジティブ名言
希望さえあればどんな所にも たどりつくと決心している
生命のあるかぎり、希望はあるものだ
世の中には勝利よりも勝ち誇るに値する敗北がある
われらの最大の栄光は、一度も失敗しないことではなく、倒れるごとに起きることにある
すべての不幸は未来への踏み台にすぎない
ネガティブ名言
喜びは忘れられていく。しかし、悲しみは忘れることはない。
どんなにもチャンスが巡ってきて、僕がゼロなら意味がない
過ぎ去った不幸を嘆くのは、すぐにまた新しい不幸を招くもど
夢見る力のない者は、生きる力もない
怠惰は弱い頭の逃避場、愚者の休日にはすぎない

- (2) 被験者 8 名により、(1) にて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ちらでもないかを判断する。ここでは、被験者 8 名のうち 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。
- (3) 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 名言を各々ポジティブ名言、ネガティブ名言の教師データとして名言データベースに格納する。

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。

4.2 単語 PN 値

名言 PN 値を求めるために、名言に含まれている単語の PN 値である単語 PN 値を求める。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。以下の手順で単語 PN 値を求める。

- (1) 予備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。
- (2) 名詞と動詞と形容詞には、PN 判定において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。ストップワードの例を表 3 に示す。

表 3 ストップワードの例
 Table 3 Example of stop-words

ストップワードの例	
名詞	人間, 人生, 自分, 人
形容詞	ない
動詞	する, 持つ, いる, 言う, 考える

表 4 比較実験の結果
 Table 4 Result of comparison experiment

手法	ポジティブ名言対象		ネガティブ名言対象	
	再現率	適合率	再現率	適合率
手法 A	72 %	54 %	62 %	52 %
手法 B	68 %	55 %	50 %	53 %

- (3) 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.
 (4) (3) にて求めた tf 値を用いて, 名言に含まれるすべての単語の単語 PN 値を求める.
 ここで, 単語 i に対する単語 PN 値 WPN_i を式 (2) に示す.

$$WPN_i = \frac{X_{Pi}}{X_{Pmax}} - \frac{X_{Ni}}{X_{Nmax}} \quad (2)$$

X_{Pi} は, 4.1 節章で求めた名言データベース内のポジティブ名言群での単語 i の出現頻度を表し, X_{Pmax} は, 単語 i と同じ品詞の単語で且つポジティブ名言群の中で出現頻度が最大の単語の出現頻度を表す. X_{Ni} は, ネガティブ名言群での単語 i の出現頻度を表し, X_{Nmax} は, 単語 i と同じ品詞の単語で且つネガティブ名言群の中で出現頻度が最大の単語の出現頻度を表す. 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. 我々が求めた単語 PN 値の単語総数は, 1933 語である.

5. 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. 名言には, 単文で構成される名言と, 複文から構成される名言がある. そこで我々は単文からなる名言と複文からなる名言で名言 PN 値の求め方を各々提案する.

5.1 単文からなる名言の名言 PN 値

単文からなる名言の名言 PN 値を求めるためにさきほど求めた単語 PN 値のみを用いる方法と, 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を用いる方法を提案する.

単語 PN 値のみを用いる手法

単語 PN 値のみを用いる手法は以下の式により求める

$$MPN_j = \frac{\sum_{k=1}^n WPN_k}{n} \quad (3)$$

ここで, MPN_j は名言 j における名言 PN 値であり, n は名言 j に含まれる単語 PN 値を持つ単語数を示し, WPN_k はその単語の単語 PN 値を示す.

名言スコアリング手法

名言スコアリング手法は以下の式により求める.

$$MPN_j = \frac{P_{num} - N_{num}}{P_{num} + N_{num}} \quad (4)$$

P_{num} は名言 j にポジティブな単語が出現する回数であり, 同様に N_{num} は名言 j にネガティブな単語が出現する回数である.

これまで我々の研究⁷⁾ では, 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, 単語 PN 値のみを使い名言 PN 値を求める場合 (手法 A) と単語 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 (手法 B) の適合率と再現率の比較実験を行ったところ, わずかではあるが単語 PN 値のみを用いて名言 PN 値を求める方が良い結果になることがわかった. 表 4 に結果を示す. この結果より本論文では, 単語 PN 値のみを用いて単文からなる名言の名言 PN 値を求めることにする.

5.2 複文からなる名言の名言 PN 値

名言には, 単文で構成される名言だけでなく, 複文から構成される名言もある. 単文からなる名言であれば, 1 文がポジティブであればその名言がポジティブであり, 1 文がネガティブであればその名言もネガティブであるので, 前節の手法でも対応できる. しかし, "努力した者が全て報われるとは限らん. しかし成功した者は皆すべからく努力しておく." このような複文から構成される名言の場合, 我々の主観の判断では, 1 つ目の文はネガティブと判断し, 2 つ目の文はポジティブと判断し, 全体としてはポジティブと判断する. 一方システムでは, 文ごとの構成を考慮しない場合は, 全体としてネガティブと判定してしまうかもしれない. それは, 文の構成を考慮しないで単語の値だけでみているので, 値の大きい単語に引っ張られて全体が判断されてしまっているからである. これでは, 複文にうまく対応できずに名言 PN 値の適合率や再現率が向上しない問題が生じる. そこで, 本論文では名言を分析し, 複文から構成される名言に対応した名言 PN 値の式を提案する.

表 5 名言分析の実験結果
Table 5 Result of the quotes analysis

	一致している	一致していない
最後の文	88 %	12 %
後ろから 2 番目	63 %	37 %
後ろから 3 番目	63 %	37 %

5.2.1 名言分析の実験

複文で構成される名言のどの文が、全体に大きく影響を与えているかを分析するために、名言分析の実験を行う。

実験手法

はじめに、名言データベースにある名言 700 個を無作為に抽出する。次に、被験者 6 名により、先程取得した名言 700 個を文ごとに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を判断する。また、各名言がポジティブであるか、ネガティブであるか、どちらでもないかも被験者の主観で判断する。

名言分析の実験結果

実験で使用した名言 700 個の内分けは、短文で構成される名言が 252 個、2 文で構成される名言が 292 個、3 文以上から構成される名言が 156 個であった。このうち、単文で構成される名言については、文の PN 値がその名言の PN 値と同じと考えるので、実験では用いないこととし、複文のみを用いる。まず、名言に含まれる最後の文の値を調べ、その値と名言 PN 値が一致しているかどうかを調べた。ここでは、2 文以上から構成される名言 448 名言を用いる。次に、名言に含まれる後ろから 2 番目の文の値を調べ、その値と名言 PN 値が一致しているかどうかを調べた。ここでは、2 文以上から構成される名言 448 名言を用いる。最後に、名言に含まれる後ろから 3 番目の文の値を調べ、その値と名言 PN 値が一致しているかどうかを調べた。ここでは、3 文以上から構成される名言 156 名言を用いる。この 3 つを分析した結果を表 5 に示す。

考察

実験結果より、最後の文の PN 判定と名言の PN 判定が一致している確率は、88 % という高い確率であることがわかった。また、他の文に比べて、最後の文が名言全体に与える影響が大きいこともわかった。この分析結果を基に複文に対応した新しい名言 PN 値を提案する。

5.2.2 複文における名言 PN 値の式

名言分析の実験結果より、複文における名言 PN 値を以下の式で求める。

$$MPN_{fj} = \frac{\sum_{k=1}^{n-1} MPN_k + \alpha * MPN_{jn}}{n} \quad (5)$$

ここで、 MPN_{fj} は複文からなる名言 fj における名言 PN 値であり、 n は名言 fj に含まれる文の数を示し、 MPN_k は名言 fj に含まれる文の名言 PN 値を示す。ここでの MPN_k は、式 (3) で用いた式を用いて、文の名言 PN 値を求める。 MPN_{jn} は n 番目の文の名言 PN 値を示す。ここでは、名言 fj の最後の 1 文を表す。 α は最後の文の重みを示す。名言分析の結果より、最後の文が名言全体に影響する確率は 88 % で、他の文が名言全体に影響する確率は 63 % であることがわかった。つまり最後の文は、他の文よりも約 1.4 倍の影響力を持つということである。この名言分析の結果から、先程用いた最後の文の重みを表す α の値を、1.4 と設定する。この値を式 (5) に代入した式を用いて、複文からなる名言の名言 PN 値を求めていく。

6. 否定語への対応

文ごとの PN 値を判断する上で、人が明らかにネガティブと判断できる文であるのに、システムではポジティブと判定してしまう問題が生じる場合がある。例えば、「彼は逆上がりができない」という文があるとする。人の判断では、「彼は逆上がりができる」ということを否定しているのでネガティブな文と判断する。しかし現状のシステムでは否定語に対応しておらず、「彼は逆上がりができる」と判断し、ポジティブな文と判断してしまう可能性がある。そこで、否定語に対応することにより、人の判定とシステムの判定の誤差を減らし、名言 PN 値の再現率の向上を測る。いくつかある否定語の中でも本論文ははじめの一步として、「ない」という否定語に絞って否定語に対応する。PN 判定時に、否定語が出てきた場合、否定語に係る単語の単語 PN 値を反転させる。具体的には、先程の文の場合、「ない」は「できる」に係っているので、「できる」の単語 PN 値 (例. 0.8) を反転させことで、ポジティブ (0.8) であった値をネガティブ (-0.8) に反転させる。

6.1 ないの種類とその例

名言に含まれる「ない」を分析したところ、助動詞と形容詞の 2 種類があることがわかった。助動詞の「ない」は、「名詞+し+ない」、「名詞 (動詞)+じゃ+ない」などのようにいくつかの種類があるが、本論文では、「できない」、「作れない」などのように、動詞+助動詞の「ない」という種類に絞って対応する。形容詞の「ない」は、「名詞 (動詞)+こと+なんて+ない」、「名詞+の+ない」などのようにいくつかの種類があるが、本論文では、「不幸もない」、「死ぬしかない」などのように、名詞 (動詞)+助詞+形容詞の「ない」という種類に

表 6 名言 PN 値の比較実験の結果
Table 6 Result of comparison experiment of the P/N value

手法	ポジティブ名言対象		ネガティブ名言対象	
	再現率	適合率	再現率	適合率
手法 A	71 %	54 %	62 %	52 %
手法 B	70 %	53 %	62 %	52 %
手法 C	74 %	55 %	60 %	51 %
手法 D	71 %	54 %	60 %	50 %

絞って対応をする。これらの否定語への対応を用いて名言 PN 値を求めていく。他の種類やその他の否定語への対応については今後の課題とする。

6.2 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、まずインターネット上から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: 単語 PN 値のみを用いた手法 (式 (2) のみ使用)
- 手法 B: 単語 PN 値と否定語への対応を行った手法 (式 (2) のみ使用)
- 手法 C: 単語 PN 値と複文への対応の式を用いた手法。 (式 (2) と式 (5) を使用)
- 手法 D: 単語 PN 値と複文への対応、否定語への対応を行った手法。 (式 (2) と式 (5) を使用)

6.3 名言 PN 値の比較実験の結果

名言 PN 値の比較実験の結果を表 6 に示す。実験結果より、どの手法も類似した結果となった。その中でも手法 C は、ポジティブ名言を対象とした時の再現率が他の手法と比較して良い結果となった。これは、名言分析を行い複文に対応した式を提案したことで、複文からなる名言に対応できた結果と言える。しかし、否定語への対応は、良い結果が出なかった。これは、否定語への対応が未完成であり、これから改良していく余地があることを表している。どの手法も適合率が悪い結果を残してしまったのは、200個の名言を人手で判定した時、ポジティブでもネガティブでもないと判定された名言が多かったからである。今後

は、ポジティブ、ネガティブを判定し易い名言を用いて実験を行う必要がある。また、複文の式で用いた α の値を変化させ、名言 PN 値の再現率向上を行う必要がある。

7. プロトタイプシステム

Ruby on Rails を用いて、提案手法のプロトタイプシステムを開発した。システムのフローを図 1 に示す。図 1 より、まずユーザは興味のあるカテゴリを 10 種の中から 1 つ選択する。これはユーザにとって興味のない情報を提示されるより、ユーザにとって興味のある情報を提示した方が気分をより良くできると考え興味のあるカテゴリを追加した。興味のあるカテゴリは、「スポーツ」、「漫画」、「映画」、「ドラマ」、「恋愛」、「友情」、「努力」、「青春」、「人生」、「政治」である。名言データベースはあらかじめ作成しており、各々の名言は Web ページの名言サイトから取得した。「スポーツ」カテゴリの名言は、みんなの名言集*1 から取得し、「漫画」カテゴリは、名言ベスト.com *2 から、「政治」カテゴリは政治の名言・格言*3 から、「ドラマ」カテゴリと「映画」カテゴリは心に残る名言集*4 から取得し、「人生」、「恋愛」、「友情」、「青春」、「努力」カテゴリは、名言集.com*5 から取得した。

次に「今の気分」カテゴリを 10 種の感情カテゴリの中から 1 つ選択する。これは、中村明の感情表現辞書⁵⁾を用いた 10 種の感情カテゴリを使用する。3 番目に、「今の気分」をカテゴリの中に入っている言葉から選択する。4 番目に「なりたい気分」カテゴリを 3 種 (喜, 好, 安) の中から 1 つ選択する。最後に、「なりたい気分」をカテゴリの中に含まれている単語を 1 つ選択する。システムは、「なりたい気分」と「今の気分」から気分距離を求め、興味のあるカテゴリに含まれる名言の中で、気分距離が一番近い値の名言を提示する。

8. プロトタイプシステムの評価実験

作成したプロトタイプシステムを用い、被験者 5 名により提案手法の評価実験を行った。被験者には以下の 3 パターンのシチュエーションにおいて以下 1 と 2 の評価を行った。プロトタイプシステムの画面を図 2 に示す。

シチュエーション

*1 みんなの名言集 <http://quote.qooin.com/>
 *2 名言ベスト.com <http://www.meigen-best.com/>
 *3 政治の名言・格言 <http://www.oyobi.com/maxim01/13.html>
 *4 心に残る名言集 <http://www.cooking-korea.com/>
 *5 名言集.com <http://www.meigensyu.com/>

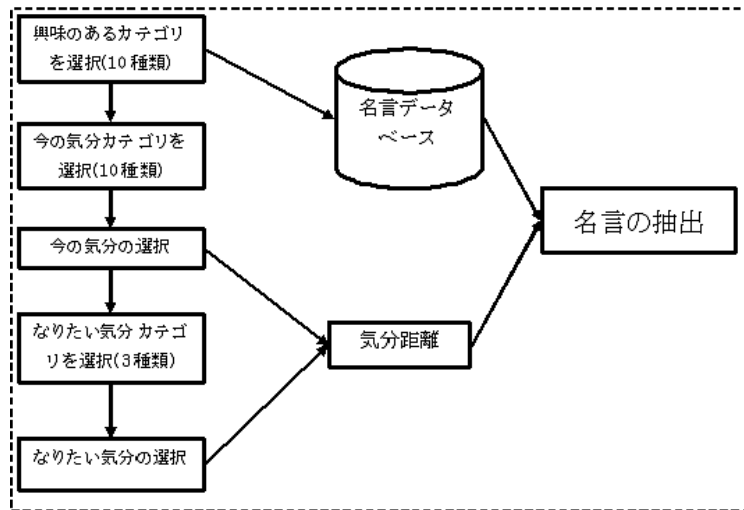


図1 システムのフロー
 Fig.1 System Flow

- パターン1: 普通の気分の時
- パターン2: ポジティブな気分の時
- パターン3: ネガティブな気分の時

実験は、パターン1のシチュエーションで各被験者にシステムを利用してもらい、全員が終了した後、気分を変える会話を行い、次のシチュエーションのパターンの気分に変えて実験をおこなった。各パターンの実験終了後に各々アンケートを取り被験者の気分の変化やシステムの評価をした。アンケートの質問事項を以下に示す。

- 質問項目 Q1: システム使用後被験者の気分がどのように変わるか?
- Q2: 気分に合った名言が提示されたか?

8.1 評価実験の結果

評価方法1に対する評価実験結果を図3に示す。図3では、縦軸が評価した人数を表し、横軸が「とてもネガティブになった」から「とてもポジティブになった」の5段階評価を表す。質問項目2に対する評価実験結果を図4に示す。図4では、縦軸が評価した人数を表

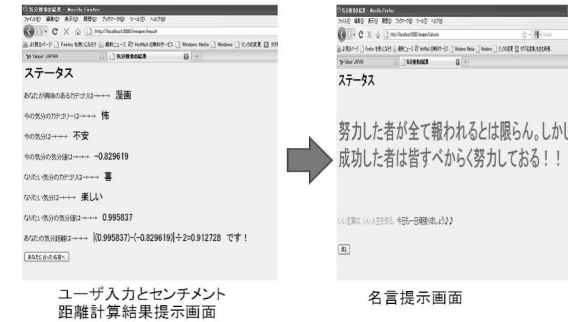


図2 プロトタイプシステムの画面
 Fig.2 Display of Prototype System

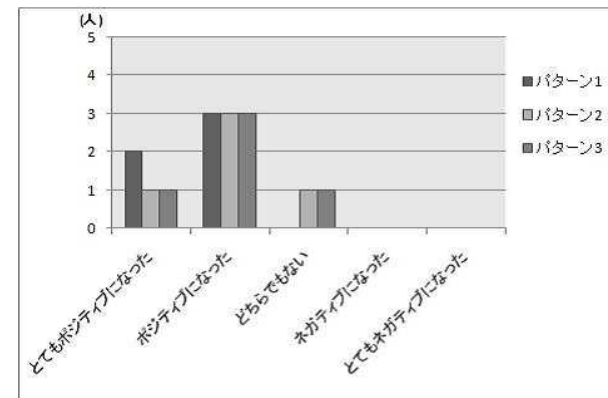


図3 質問項目1の結果
 Fig.3 Result of Q1

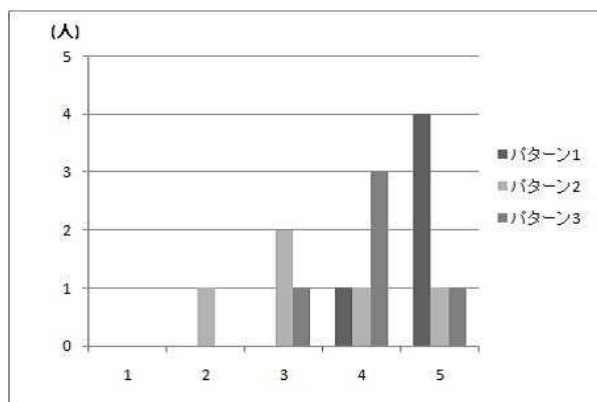


図4 質問項目2の結果
Fig.4 Result of Q2

し、横軸が1(とても気分が合っていなかった)から5(とても気分が合っていた)の5段階評価を表す。

図3の評価実験の結果1より、どのパターンでも被験者の気分を良くできた。実験前は、ネガティブな気分の時にこのシステムを使用した時が一番被験者の気分をポジティブにできると考えていたが、結果より普通の気分の時にこのシステムを使用した時が一番ポジティブになったので、普通の気分の時にこのシステムを使用するとポジティブな気分ができるという新たな発見があった。

図4の評価実験の結果2より、パターン1とパターン3では気分が合った名言を提示できた。しかし、パターン2では良い結果を得ることができず、気分が合った名言を提示できなかったと言える。この結果からパターン2であるポジティブな気分の時に合った名言を提示できなかった問題が残った。しかし、パターン1とパターン3で良い結果を得られたことから、本システムは、ネガティブな気分の時や普通の気分の時に使用することで、気分が合った名言を提示できることが分かった。

9. まとめと今後の課題

本研究では、単語PN値を提案し、名言PN値を用いてユーザの入力した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を提案した。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、単語PN値や複文からなる名言に対応した式を提案や否定語に対応する提案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値の再現率の向上
本研究では、名言分析実験を行い、複文からなる名言の式を提案した。また、否定語「ない」への対応を提案した。今後はよりユーザにあった名言を提示するために、文構造の考慮、他の否定語への対応などといった新たな提案をし、名言PN値の再現率向上に取り組んでいきたい。
- 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、ユーザの気分が合った情報を提示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が合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、更にユーザの気分が合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。

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, Vol.47, No.2, 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, 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, p.4 (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).
- 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 (2010).