

多次元感情ベクトルを考慮した 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、ユーザの気分に合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、ユーザの気분을考慮した検索システムの提案を行う。本論文では、はじめの一步として「名言」を対象とする。具体的には、実験により名言用の多次元感情ベクトルを提案し、その多次元感情ベクトルから名言に含まれる単語の感情値を定量化する。その単語の感情値を用いて名言の感情値を求める手法の提案を行う。そして、ユーザの入力した感情値と類似する名言を検索し提示するシステムを提案する。

Words-of-Wisdom Search System based on Multi-dimensional Sentiment vector

KOICHI 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. We propose the system that searches for the information based on user's sentiment. In this paper, we propose Words-of-Wisdom Search System as a first step of the research. Specifically, we first propose the multi-dimensional sentiment vector based on experiment, words sentiment value of words-of-wisdom, and words-of wisdom sentiment value.

^{†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

インターネットの普及により、誰でも簡単に情報収集をすることが容易になった。現在の検索エンジンは、Yahoo!JAPAN^{*1}や Google^{*2}に代表されるようにキーワード検索が主流となっている。しかし、現在のキーワード検索では、ユーザの気분을考慮した検索を行うことが出来ない。例えば、キーワードを「楽しい気分になる本」や「のんびりとした気分になる本」として検索した時、Q&A サイトが検索結果の上位に表示される。Q&A サイトの結果なら、目的の本を見つけることも可能だが、回答者の主観による回答になっており、ユーザにとって有益な情報とは言い難い。

一方、人間の感情は複雑だと言われている。また、その複雑な感情に合わせて行動を選択することがある。例えば、音楽を聞く際に、毎日同じ音楽を聴くのではなく、その日の気分ややりたい気分に合わせて音楽を選択することで、気分を良い方向に導こうとする。このように、ユーザの今の気分ややりたい気분을考慮した情報を提示することで、よりユーザが欲しい情報を提示することが可能であると考えられる。そこで、本研究では、ユーザの気분을考慮した検索システムの提案を行う。その初めの一步として検索対象を名言とし、感情を考慮した名言検索手法を提案する。

現在名言は、名言集の本や名言に関するテレビ番組、Web 上での名言集サイトなど様々なメディアで名言が扱われており、人々の心を良い方向へ導いている。名言集の本では、白取春彦の「超訳ニーチェの言葉¹⁾」という単行本が 2010 年度オリコンランキング^{*3}で総合 5 位、自己啓発部門で 1 位に輝くほど、名言はブームになっている。また、Web 上の名言集サイトはたくさんあり、誰でも簡単に名言を読むことができるようになっている。名言集サイトの名言ナビ^{*4}では、約 3 万もの大量の名言があり、その中からユーザに合った名言を見つけ出すのは困難である。そこで、本研究ではユーザの複雑な感情を定量化し、その数値に合った名言を提示する研究を行う。

以前の我々の研究²⁾では、感情のポジティブ・ネガティブに注目し、P/N 値を用いて名言検索を行った。しかしながら、人間の感情は複雑であり、P/N 値だけでは欲しい名言が

*1 Yahoo!JAPAN <http://www.yahoo.co.jp/>

*2 Google <http://www.google.co.jp/>

*3 オリコン 2010 年度本ランキング

<http://www.oricon.co.jp/entertainment/ranking/2010/bookrank1201/index02.html>

*4 名言ナビ <http://www.meigennavi.net/>

上手く取得できなかった。そこで、本研究では、多次元の感情ベクトルに基づいた名言検索手法の提案を行う。本研究の感情ベクトルは、中村明の感情表現辞典³⁾の10次元感情ベクトルに基づいた実験を行い、名言にとって最適な感情ベクトルを提案する。そして、この感情ベクトルを用いて、ユーザの気分と名言の定量化を行う。ユーザの気分と名言を定量化することで、大量の名言の中からユーザのなりたい気分にあった名言を抽出することが可能になる。ユーザのなりたい気分にあった名言を抽出することで、情報検索の支援とユーザの気分を少しでも良い方向へ導くことを目的とする。

以下、第2章では関連研究についてを、第3章では10次元の名言感情ベクトルについてを、第4章では感情ベクトルの結合についてを、第5章では名言単語値の求め方についてを述べる。第6章では名言値の求め方についてを、第7章ではシステムの評価実験についてを、第8章ではプロトタイプシステムについてを、第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

近年、感情表現抽出に関する研究は盛んに行われている。その中で、感情表現を表す感情モデルはさまざまなモデルが提案されている。さまざまな感情モデルの中で、多次元の感情ベクトルが多く提案されている。

代表的な感情モデルとして Plutchik^{4) 5)} のモデルがある。Plutchik のモデルでは、「受容」「嫌悪」「予期」「驚き」「喜び」「悲しみ」「怒り」「恐れ」の八つの感情を基本感情として分類され、4次元の感情ベクトルを表している。また、徳久⁶⁾ は、ユーザの発話内容から感情を推定する手法を提案している。「遠くの店に行ったら定休日だったので残念だ」という文から「残念」という感情が生起する要因として「遠くの店に行ったら定休日だった」という感情生起要因コーパスを用いて、Web テキストから感情生起要因と感情を抽出する手法を提案している。感情モデルには、(嬉しい、楽しい、安心、恐れ、悲しい、残念、嫌、寂しい、心配、腹立たしい) と、(neutral) の11種類の感情を用いている。

河合⁷⁾ 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。MPVPlus 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。

上記のように、多次元のベクトルを用いて感情表現を抽出する研究が多く行われている。

Dave ら⁸⁾ の研究では、Web 上のレビューを肯定、否定に分類する研究を行っている。肯定、否定の評価の差分をとることで評判を特徴づけるような表現を獲得してきている。また、小林⁹⁾ は P/N 判定に必要な辞書知識の集合「P/N 辞書」を国語辞典の語釈文からブートストラップ的に獲得し、単語の感情極性を抽出する手法を提案している。この研究では、国語辞典は、語釈文が見出し語の意味の説明になっているので、見出し語の P/N 評価は一致していると仮定して、研究を進めている。

藤村¹⁰⁾ 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。

また、高村¹¹⁾ 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。

上記のように、感情表現抽出に関する研究は盛んに行われている。しかし、コンテンツを見たり聞いたりすることでユーザが感じる印象や感情を抽出する研究はまだ少ない。

熊本¹²⁾ は、様々な感情のうち喜怒哀楽に焦点を当て、Web ニュース記事に含まれている喜怒哀楽の程度を決定するためのシステムを提案している。喜怒哀楽の程度は、2つの感情尺度「悲しいーうれしい」「怒りー喜ぶ」に対する評価値(0.1の実数値)という形で表されている。「感情語 e を含む記事はその感情語が感情を伝える」という仮定の下、新聞記事に現れる各単語が感情尺度を構成する2つの感情語のどちらと、より高い確率で共起するかという観点でシステムの設計を行っている。熊本^{12) 13)} は、コンテンツを見たり聞いたりした時に人々が感じる印象をコンテンツそのものから抽出する手法を提案している。本研究でも、情報発信者の感情ではなく、情報受信者が感じる気分を抽出する手法を提案する。

3. 10次元の名言感情ベクトル

はじめに、既存の感情ベクトルを、ユーザ実験により、名言に合った感情ベクトルを決定する。

表 1 10次元の感情ベクトルと単語の例
Table 1 Ten type of sentiment vector.

感情カテゴリ	単語	感情カテゴリ	単語
喜	楽しい, 面白い, 嬉しい	怒	怒る, 腹立たしい, 憤怒
哀	哀(悲)しい, 痛い, 孤独	怖	怖い, 不安, 震える
恥	恥しい, もじもじ, 照れる	好	恋しい, 愛しい, 好き
厭	憎い, 暗い, 辛い	昂	苛立つ, 感動, 動揺
安	のんびり, すっきり, 安らぎ	驚	ぼんやり, 驚喜, 慌てる

3.1 10次元感情ベクトル

我々の以前の研究²⁾では, PN 値(肯定「ポジティブ」 否定「ネガティブ」)を用いて名言に使われている感情を定量化することを行ってきた。しかしながら, 実験により PN 値だけでは感情を正確に測ることが困難であることがわかった。そこで本研究では, 多次元の感情ベクトルを用いてユーザの気分合った名言を提示する手法の提案を行う。多次元ベクトルのモデルとして, 中村明の感情表現辞典³⁾を用いる。感情表現辞典では, 感情を表現する言葉を 10次元の感情ベクトルに分類している。10次元の感情ベクトルとそれに含まれる単語の例を表 1 に示す。そして, この感情表現辞典を用いて, 名言を 10次元のベクトルに分類する。感情表現辞典では, この 10次元の感情ベクトルの要素名をカテゴリと呼んでおり, 本論文ではこの要素名を「感情カテゴリ」と呼ぶ。つまりは, 表 1 にあるように, 10次元のベクトル名の 1 つである「喜」は「喜カテゴリ」と呼ぶ。名言の中に感情表現辞典の単語が含まれていれば, 感情表現辞典においてその単語の示す感情カテゴリを名言の印象値とする。

3.2 10次元による名言からの感情の抽出

10次元の感情ベクトルを用いて, 名言から感情の自動抽出を行う。例えば「幸せとは, 楽しい考え方が心を占めている状態を言う」という名言の場合「楽しい」という単語が感情表現辞典の「喜」カテゴリに含まれているので「喜」と判定する。また, この名言に含まれる「幸せ」という単語は, 感情表現辞典の「喜」に含まれているので, この名言は「喜」カテゴリの感情を持つ名言として抽出する。この様に, 名言に含まれる単語と感情表現辞典の 10種の感情カテゴリに含まれている単語を照らし合わせ, 名言からの感情を抽出する。また「うきうきする」という単語の場合, 感情表現辞典の「喜」「好」カテゴリ両方に含まれているため「うきうきする」を含む名言は「喜」「好」両方の感情を持つ名言として抽出する。ここで抽出した感情をシステムの判定とする。

表 2 実験結果の例

Table 2 Example of experimental results.

名言	システムの判定	被験者の判定 (上位 3 件)	適合率
人を愛するよりも人に愛されるよりも まず自分を愛せたら幸せの始まりだと思う。 笑顔は万言に勝る	喜, 好	好, 喜, 安	1.0
インターナショナル・サインである。 女的美貌は高慢の種となっても, けっして愛情のこもった快楽を 約束するものではないのです。	喜	喜, 好, 安	1.0
恋の喜びは一瞬しか続かない。 恋の悲しみは一生続く。	喜, 好, 安	哀, 怖, 驚	0.0
	喜, 哀, 好	哀, 厭, 怖	0.33

3.3 予備実験

本研究では, 感情表現辞典の 10次元の感情ベクトルから 3.2 に示すように, 名言に適した感情ベクトルを生成する。その時, この 10次元の感情ベクトルが名言に適しているかを判断するための実験を行った。

予備実験手順

本実験では, インターネット上の 700 個の名言から無作為に 100 個を取得し, その名言を用いて, 被験者 15 人により各名言が 10次元の感情ベクトルのどの感情を持っているのかを実験した。被験者は各々の主観で判断し, 感情が判断できない場合は「どの感情にも当てはまらない」を選んでもらい, 感情は 10次元の内最大 2 つまで選択することを許可した。また, 実験環境としては, 感情に左右されやすい実験なので, 平常心の時に行ってもらった。

予備実験結果
実験で得られた被験者が感じた感情と 3.2 節で提案したシステムが判定した感情を比較し適合率を求めた。適合率は, 15 人の被験者が選んだ感情のうち, 感情の合計数の上位三位までを対象とした。表 2 に結果の一例を示す。上位三位までの適合率は 55 % となった。結果より, システムの判定と被験者の判定に大きな違いがあり, 10次元では次元の数が多いことがわかった。そこで, 次元を結合し次元数を減らすことを提案する。

4. 感情ベクトルの結合

3.3 節の実験の結果から, 感情ベクトルの次元を結合し 10次元の感情ベクトルの次元数を削減する。

4.1 感情カテゴリの共起確率

3.3 節の実験にて被験者が選んだ感情カテゴリが 1 つの名言において複数存在している場

合がある。例えば、表 2 より「人を愛するよりも人に愛されるよりもまず自分を愛せたら幸せの始まりだ」という名言の場合、被験者の判定は「喜」「好」「安」である。このような場合我々は一つの名言において感情が共起しているといい、この共起している感情を共起感情と呼ぶ。例えば先の例において「喜」と「好」と「安」は共起感情である。次元を結合するに当たり、共起感情は類似している感情である可能性が高いと考え、共起感情に基づいて次元結合を行う。次元結合を行うに当たり、まずは 3.3 節の実験で用いた 100 名言における共起感情の個数を求める。この個数を共起頻度数 CW_{ij} と呼ぶ。そして、この共起頻度数を用いて、各感情カテゴリが名言 100 件の中に出現する確率 CA_{ij} を以下の式で求める。

$$CA_{ij} = \frac{CW_{ij}}{AW_j} \quad (1)$$

i は比較元の感情カテゴリを表し、 j は i と共起している感情カテゴリを表し、 AW_j は 100 名言中に j が出現する名言数を表している。例として、「喜」カテゴリをベースにした結果を表 3 に示す。

4.2 次元結合

10 次元から次元を削減するために、上記共起確率から感情カテゴリをノードとする有向グラフを生成し、そのグラフから次元を結合することにより次元削除を行う手法を提案する。本論文では 2 つの手法を提案する。

(1) 共起確率上位 3 件による次元結合

被験者実験の結果から共起確率上位 3 件の結果を対象としてカテゴリ間のグラフを作成する。このグラフを感情グラフと呼ぶ。感情グラフの各ノードは共起確率上位 3 件とする。その為、各ノードの出次数は必ず 3 となる。つまり、あるカテゴリ A との共起確率上位 3 件のカテゴリを B, C, D とする場合、A-B, A-C, A-D となる。ここで、あるカテゴリともう一つのカテゴリを比較し、各々が双方向に連結し、且つ他の 2 つの双方向に連結しているノード(カテゴリ)が同じ場合、これらのカテゴリを結合する。例えば、図 1 の場合、 n_1 と n_2 のノードを比較する。 n_1 と n_2 は双方向にリンクしており、且つ各々のノードが双方向に連結している他の 2 つのノードは n_3 と n_4 は同じであるため、 n_1 と n_2 は結合できる。次に n_1 と n_3 を比較する。 n_1 と n_3 は双方向に連結している。しかし、各々のノードが双方向に連結している他の 2 つのノードは、 n_1 は n_2 と n_4 であるのに対し、 n_3 は n_2 と n_5 であり異なる。そのため、 n_1 と n_3 は結合しない。図 2 に次元の結合の結果を示す。ここでは、「喜」と「好」が結合され一つのベクトルとなり、同様に「哀」「怒」が結

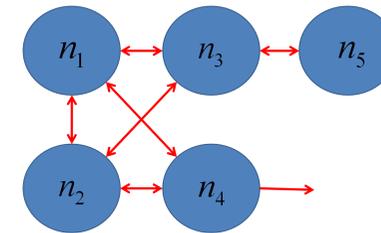


図 1 次元結合の例 1
Fig. 1 Example 1 of dimensions binding.

表 3 喜カテゴリをベースにした共起頻度数と共起確率
Table 3 Co-occur frequency and Co-occur probability based of ki category.

	喜	怒	哀	怖	恥	好	厭	昂	安	驚
喜	0	0	5	1	4	32	2	30	21	6
AW_j 全体	47	6	47	24	24	36	41	39	28	19
共起確率	0	0	0.10	0.04	0.16	0.90	0.04	0.78	0.76	0.32

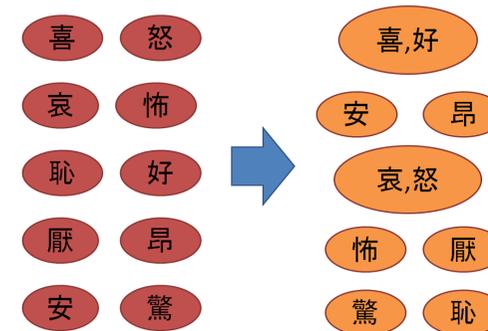


図 2 次元の結合の例 2
Fig. 2 Example 2 of dimensions binding.

合され、その結果 8 次元の感情ベクトルとなる。

(2) 共起確率閾値による次元結合

次に共起確率の閾値 以上を対象として感情グラフを取得する手法を提案する。まず始めに、感情カテゴリをノードとし、共起確率が 以上のカテゴリを連結して感情グラフを生成

する。例えば、閾値を0.35とした場合、表3の「喜」カテゴリを対象とした場合、閾値を超えて「好」「昂」「安」カテゴリそれぞれと共起関係になる。この共起関係を基に次元結合を行う。共起確率を使用する際の次元結合の定義を以下に示す。

- 定義1: 双方ともに共起関係である
例えば、「喜」カテゴリを対象とした場合、「好」カテゴリとの共起確率が閾値（ここでは $=0.35$ とする）を超えているので、共起関係（「喜」「好」）として抽出する。次に、「好」カテゴリを対象とした場合、「喜」カテゴリとの共起確率が0.96となり、閾値を超えているため、共起関係として抽出する。「喜」カテゴリと「好」カテゴリは双方ともに共起関係であるので、「喜」カテゴリと「好」カテゴリは双方ともに共起関係（「喜」「好」）であるとする。
- 定義2: 感情グラフで閉路である
図1より、 $n1$ と $n2$ は双方ともに共起関係であり、 $n2$ と $n3$ は双方ともに共起関係であり、 $n3$ と $n1$ は双方ともに共起関係である。図1より $n1, n2, n3$ を対象としたグラフは閉路になっているので、 $n1, n2, n3$ は共起関係となる。
- 定義3: 閉路が複数ある場合は共起確率が高いグラフを結合させる。
図1より、 $n1, n2, n3, n4$ を対象とした場合、 $n3$ と $n4$ が共起関係でないことから、 $n1, n2, n3, n4$ は共起関係ではなくなる。この場合、 $n1, n2, n3$ の閉路と $n1, n2, n4$ の閉路の2つの閉路ができる。 $n1, n2$ の様に、1つのカテゴリが複数の閉路を持つ場合は共起確率が高い方を共起関係として抽出する。

すべての定義を満たした共起関係のカテゴリを結合し、1つの次元として抽出する。

4.3 適合率の比較

10次元ベクトル、共起確率上位3件、閾値 $=0.2, 0.3, 0.35, 0.4, 0.5$ を対象として、それぞれ適合率を求め、比較する。ここで用いた、閾値 $=0.3, 0.35, 0.4$ の次元結合の結果を図3に示す。各実験で得た被験者が判断した感情とシステムが判定した感情との適合率を求めた。適合率の比較実験の結果を表4に示す。表4の結果より、閾値 $=0.3$ が一番良い結果となった。しかし、次元数を減らすと適合率は向上するので、次元の閾値を定義する。ドイツの心理学者である Wilhelm Max Wundt¹⁴⁾ は、感情分類、感情の構造に関する古典的研究をしており、感情を { 快, 不快, 興奮, 沈静, 緊張, 弛緩 } 6種類に分類している。また、デカルト¹⁵⁾ は基本情念は6つだけと提言している。基本的な情念は { 驚き, 愛, 憎しみ, 欲望, 喜び, 悲しみ } の6つに分類されており、他のすべての情念は、6つの情念のいくつかの複合、あるいは種であるとされている。上記のように基本的な感情は6種類だ

表4 適合率の比較

Table 4 Result of comparison experiment of precision.

	次元数	適合率
10次元	10	55%
上位3件	8	59%
閾値 $=0.2$	4	68%
閾値 $=0.3$	4	68%
閾値 $=0.35$	6	63%
閾値 $=0.4$	7	60%
閾値 $=0.5$	8	59%

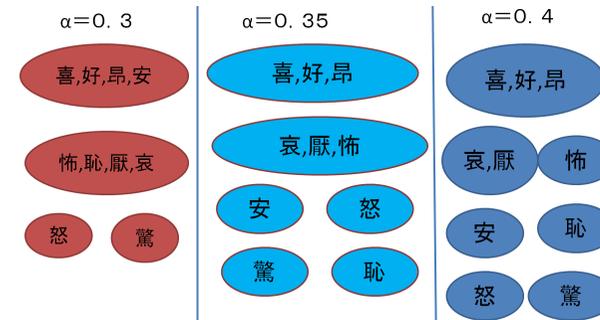


図3 閾値を用いた次元の結合

Fig.3 Dimensions binding using threshold value.

と提案されている。そこで、本研究での感情の次元数の閾値を6として研究を進める。表4より、次元数の閾値を超えておろかつ、適合率が良い結果である共起確率の閾値 $=0.35$ を用いて研究を進める。

5. 名言単語値

共起確率の閾値 $=0.35$ から得られた6次元を用いて名言の値を求めていく。名言の値は名言に含まれる単語の感情の値から求める。我々はこの名言に含まれる単語の感情の値を名言単語値と呼ぶ。本研究では、名言に含まれる単語を6次元に定量化した値を名言単語値と定義する。以下の手順で名言単語値を求める。

- (1) 名言ナビから無作為に抽出した17446名言を6次元に分類していく。分類方法は「意欲的な目標が人生を楽しくする。」という名言の場合「楽しい」という単語が感情表

表 5 Juman の結果変換ルールの一例
Table 5 Rules for transforming Juman's output.

1	サ変名詞/カタカナ/アルファベット/副詞/形容詞 (基本運用系/ダ列基本運用形)+ 動詞「する/できる」のとき、この 2 語を動詞 1 語に変換する
2	動詞+助動詞「ぬ」/形容詞性接尾辞「ない」のとき、この 2 語を動詞 1 語に変換する 名詞 (形式名詞と副詞的名詞を除く)/未定義語/動詞 (基本運用形)/副詞+ 判定詞のとき、この 2 語を形容詞 1 語に変換する
3	形容詞/動詞/判定詞+形容詞性述語接尾辞 (「ない」を除く) のとき、この 2 語を形容詞 1 語に変換する
4	名詞/未定義語/動詞/形容詞+形容詞性名詞接尾辞のとき、この 2 語を形容詞 1 語に変換する
5	形容詞+形容詞性述語接尾辞「ない」のとき、この 2 語を形容詞 1 語に変換する
6	

現辞典の「喜」カテゴリに分類されているため、この名言は「喜,好,安」カテゴリの名言と判断する。また、感情表現辞典の単語を含まない名言はどの次元にも含まれない名言とする。

- (2) 形態素解析器 Juman¹⁶⁾ を用いて、各次元で名言に含まれる単語を取得する。この際、Juman の出力をそのままの形では用いずに、熊本ら¹²⁾ の研究で用いられているルールを適用する。例えば、「運命だ」のような語は、普通名詞「運命」と判定詞「だ」に分けられるが、ルールを適用することにより、「運命だ」という形容詞 1 語として扱うことになる。また、「削除しない」のような語は、Juman によりサ変名詞「削除」、動詞「する」、形容詞性述語接尾辞「ない」の 3 語に分けられるが、ルールを適用することで「削除しない」という動詞 1 語として扱うことができる。熊本らの研究で用いられているルールは全部で 20 個あるがその一例を表 5 に示す。
- (3) 取得した単語から名詞、形容詞、動詞の出現頻度 tf 値を求める。名詞、動詞、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。ストップワードの例を表 6 に示す。
- (4) 各次元で求めた単語の値を 6 次元に結合し、名言単語辞書に登録する。

我々が求めた名言単語辞書にある総数は、4346 語である。

5.1 形容詞類語処理

名言単語値の中には形容詞が含まれているが、形容詞には類語が多く含まれているため、名言に適用した場合、その類語も考慮しなければならない。そこで類語に、Weblio 類語辞典^{*1}を用いて形容詞の類語を考慮し、以下の手順で形容詞の名言単語値を求める。

*1 Weblio 類語辞典 <http://thesaurus.weblio.jp/>

表 6 ストップワードの例
Table 6 Example of stop-words.

ストップワードの例	
名詞	仕事, 生活, 自分, 人間, 言葉, あなた, 学問, 思考
形容詞	同じだ, 事だ, 色々な, ほうだ, 自分だ, ものだ
動詞	いう, あう, する, 思う, なる, いる, 持つ, 見る

- (1) 3.3 節の実験で用いた 100 名言に含まれる形容詞を取得する。ここでは、133 語の形容詞を取得した。
 - (2) (1) にて取得した形容詞 133 語の各々の類語を Weblio 類語辞典を用いて取得する。例えば、「幸福」という単語の類語は「幸せ」「満足」「ハッピー」「至福」などが得られる。また、「幸福だ」のような単語の場合、「幸福だ」でキーワード検索すると検索結果が得られないため、類語を取得する際のルールとして「単語+だ」の単語に関しては「だ」を除いてキーワード検索した結果を類語として取得する。
 - (3) 各次元に含まれる名言を形態素解析器 Juman を用いて、各次元で名言に含まれる単語を取得する。この際、高精度の辞書を作成するために熊本らの研究で用いられているルールを適用し、単語を取得する。
 - (4) 取得した単語が形容詞でかつ 133 語の形容詞に含まれる類語であれば、類語処理をした形容詞の出現頻度を 1 回増やす。例えば、形容詞「幸福」の類語が「幸せ」「豊か」である場合は、「幸せ」が出てきた場合は「幸福」の出現頻度を 1 回増やす。
 - (5) 形容詞 133 語の出現頻度 tf 値を求める。
 - (6) 各次元で求めた単語の値を 6 次元に結合し、名言単語辞書に登録する。
- 以上の処理を行い、類語処理を行う前と後の結果を表 7 に示す。表 7 のデータは、予備実験から求めた 6 次元のデータである。

6. 名言値

本研究では、名言を 6 次元に定量化した値を名言値と呼ぶ。名言値は先程求めた名言単語値を用いて求める。名言には、単文で構成される名言と、複数の文から構成される名言がある。そこで我々は単文で構成される名言と複数の文から構成される名言で名言値の求め方を各々提案する。

6.1 単文で構成される名言の名言値

単文で構成される名言の名言値を求めるために以下の式を提案する。

表 7 類語処理の結果
Table 7 Results of synonym process.

類語処理前						
単語	喜, 好, 昂	怒	怖, 哀, 厭	恥	安	驚
幸福	1.0	0	0.19	0	0.11	0
楽しい	0.28	0	0.125	0	1.0	0
類語処理後						
単語	喜, 好, 昂	怒	怖, 哀, 厭	恥	安	驚
幸福	0.98	0	0.27	0	0.94	0
楽しい	0.82	0	0.42	0.33	1.0	0

$$MPN_j = \frac{\sum_{k=1}^n WPN_k}{n} \quad (2)$$

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

6.2 複数の文から構成される名言の名言値

複数の文から構成される名言値を以下の式で求める.

$$MPN_j = \frac{\sum_{k=1}^{n-1} MPN_k + \alpha * MPN_{jn}}{n} \quad (3)$$

ここで, MPN_j は複数の文からなる名言 j における名言値であり, n は名言 j に含まれる文の数を示し, MPN_k は名言 j に含まれる文の名言値を示す. ここでの MPN_k は, 式 (2) で用いた式を用いて, 文の名言値を求める. MPN_{jn} は n 番目の文の名言値を示す. ここでは, 名言 j の最後の 1 文を表す. α は最後の文の重みを示す. これは, 以前の我々の研究²⁾ で行った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.

7. システム評価実験

以前の我々の研究²⁾ では, ユーザの今の気分となりたい気分から気分距離を求め, その気分距離に合った名言を提示していた. しかし, 本研究では多次元感情ベクトルを用いたため, 気分距離を考慮することができなかった. そこで, 本研究では, ユーザのなりたい気分

表 8 適合率の比較

Table 8 Result of comparison experiment of precision.

対象	適合率
100 名言	43 %
700 名言	53 %
名言ナビ	57 %

表 9 評価実験結果の例

Table 9 Result of evaluation experiment.

なりたい気分	提示名言
幸福だ	幸せじゃない瞬間も含めて, 幸せだと思います.
幸福だ	幸せとは楽しい考え方が心を占めている状態を言う.
幸せだ	幸せに成功したければ自分が大好きなことを仕事にしない.
悲しい	人間にとって「孤独」は前提なのです。「ひとりぼっち」は当たり前の人間の姿です.
哀れだ	死の持つ恐怖はただ一つ. それは明日がないということである.

に合った名言を提示する. なりたい気分には, 名言単語辞書に含まれる 133 語の形容詞から, 気分に適した単語を 10 語を使用する. なりたい気分は {幸福だ, 哀れだ, 恥ずかしい, 恐ろしい, 幸せだ, 苦しい, 孤独だ, 悲しい, 辛い, 明るい} の 10 語である.

評価実験内容

なりたい気分 10 語にマッチする上位 3 件の名言を抽出し, 各名言がなりたい気分にはマッチした名言かを判定した. 提示する対象名言は 3.3 節の実験で用いた 100 名言, インターネットから取得した 700 名言, 名言ナビから取得した 17446 名言の 3 つのパターンを対象とした. システム評価実験結果

システムの評価実験結果を表 8 に示す. また, 実験で得られた結果の例を表 9 に示す. 表 8 の実験結果より名言ナビを対象とした場合は, 57 % という適合率を残した. 今後は, 更なる適合率向上を図るため名言単語辞書に登録される値の再考を行っていきたい.

8. プロトタイプシステム

Ruby と CGI を用いて, 提案手法のプロトタイプシステムを開発した. プロトタイプシステムの画面の例を図 4 に示す. システムを起動すると図 4 の左図のなりたい気分の選択画面が出てくる. ユーザには, リストボックスにあるなりたい気分のリストから 1 つ気分を選択してもらい, 検索ボタンを押してもらくと, 図 4 の右図の名言提示画面のように, ユーザのなりたい気分合った名言が提示されるシステムとなっている. 現在は, 名言のみの提

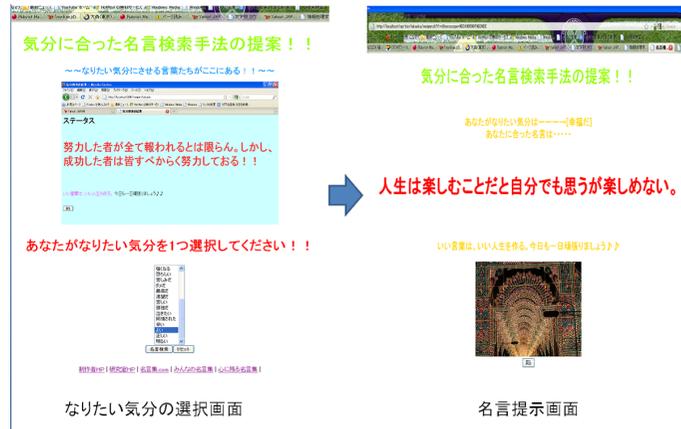


図4 プロトタイプシステムの画面
Fig.4 Display of Prototype System.

示となっているので、今後は名言の発言者の名前などの情報も付加していきたい。

9. まとめと今後の課題

本研究では、ユーザの気分にあった情報を提示する検索システムを構築するための第一歩として、名言を対象にしてなりたい気分にあった名言を提示する検索システムの構築を行った。また、10次元の感情ベクトルを実験を行い、6次元に結合した。本研究の目的は「ユーザの気分を少しでも良い方向へ導くこと」である。そのために、名言単語辞書の作成、複数の文から構成される名言に対応する式、Weblio辞書を用いて形容詞の値の再考などを行い、プロトタイプシステムを構築した。今後は、プロトタイプシステムの実験を行いたいと考えている。その他の今後の課題は以下の通りである。

- 気分距離への対応
本研究では、ユーザの「なりたい気分」にあった名言の提示を行ってきた。更にユーザの気分を良い方向へ導くために、今後は「今の気分」を取り入れ、「なりたい気分」との気分距離を測り、よりユーザの気分にあった情報を提示していきたい。
- 気分についての再考
本研究では「なりたい気分」に名言単語辞書にある形容詞の一部を自己判断で選択し用いた。しかし、選択した形容詞が「なりたい気分」に適した単語かを調査していない。

そこで、今後は気分に適した単語を調査し、ユーザに選択してもらい気分として用いていきたい。

● 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) 白取春彦. 超訳 ニーチェの言葉. ディスカヴァー・トゥエンティワン, 2010.
- 2) 高岡幸一, 瀧本明代. 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. 情報処理学会 第151回データベース研究会, No.14, 2010.
- 3) 中村明. 感情表現辞典. 東京堂出版, 1993.
- 4) R.Plutchik. The multi factor-analytic theory of emotion. *Psychology*, Vol.50, pp. 153-171, 1960.
- 5) R.Plutchik. The nature of emotions. *American Scientist*, Vol.89, pp. 344-350, 2001.
- 6) 徳久良子, 乾健太郎, 松本裕治. 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. 情報処理学会論文誌, Vol.50, No.4, pp. 1365-1374, 2009.
- 7) 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.
- 8) Kushal Dave, Steve Lawrence, and David M.Pennock. Mining the peanut gallery: opinion extraction and semantic classification of product reviews.
- 9) 小林のぞみ, 乾孝司, 乾健太郎. 語釈文を利用した「p/n 辞書」の作成. 人口知能学会 言語・音声理解と対話研究会, pp. 45-50, 2001.
- 10) 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.
- 11) 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.
- 12) 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).
- 13) 熊本忠彦, 河合由起子, 田中克己. 新聞記事を対象とするテキスト印象マイニング手法の設計と評価. 電子情報通信学会論文誌, No.3, pp. 540-548, 2011.
- 14) WilhelmMax Wundt. *Grundriss der Psychology*. 1896.
- 15) Rene Descartes. 情念論. 岩波文庫, 2008.
- 16) 黒橋禎夫, 河原大輔. 日本語形態素解析システム JUMAN VERSION 6.0.