

ユーザの気分を考慮した 名言検索支援システムの提案

高岡 幸一[†] 灘本 明代^{††}

[†] 甲南大学大学院 自然科学研究科 〒658-8501 兵庫県神戸市東灘区岡本 8-9-1

^{††} 甲南大学 知能情報学部 〒658-8501 兵庫県神戸市東灘区岡本 8-9-1

E-mail: [†]tmk.kt13@gmail.com, ^{††}nadamoto@konan-u.ac.jp

あらまし 現在の検索エンジンは、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、ユーザの気分を考慮した検索を行うことが出来ない。我々はユーザの感情を考慮した検索を行うことで、よりユーザに有益な情報を提示できると考えている。そこで本研究では、ユーザの気分を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」を対象とし、ユーザの気分にあった名言を検索するシステムを提案する。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。

キーワード 感情ベクトル, 気分検索, 名言

Words-of-Wisdom Search Support System based on User's Sentiment

Koichi TAKAOKA[†] and Akiyo NADAMOTO^{††}

[†] Graduate School of natural science graduate course, Konan University Graduate School Higashinada-ku Okamoto 8-9-1, Kobe-shi, Hyogo, 658-8501 Japan

^{††} Dept. of Intelligence and Informatics, konan University Higashinada-ku Okamoto 8-9-1, Kobe-shi, Hyogo, 658-8501 Japan

E-mail: [†]tmk.kt13@gmail.com, ^{††}nadamoto@konan-u.ac.jp

Abstract Nowadays, there are many keyword based search engines on the Internet. It is difficult, however, to search the information based on user's sentiment. We consider it is convenient for users if there is the search engine which searches for the user's sentiment like information. 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-dimensions sentiment vector based on our experiment. Next, we propose words sentiment value of words-of-wisdom and words-of-wisdom sentiment value. Finally, we show the results of our experiments using our prototype system that searches for the words-of-wisdom based on sentiment.

Key words dimensional vector, Sentiment search, Words-of-Wisdom

1. はじめに

現在の検索エンジンは、Yahoo!JAPAN や Google 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。そのため、検索する際に欲しい情報に関する単語をクエリとして入力し検索することで目的の情報を入手しようとする。例えば、映画を見たいユーザは検索クエリとして「作品名」や「映画 作品名」と入力し検索することで、目的である映画の作品に関する情報

を入手することが容易にできる。しかし、映画を見たいユーザが必ず作品名を入力するとは限らない。例えば、目的の作品はないが気分転換に映画を見たいと考えるユーザは、「映画 楽しい」や「映画 怖い」などのように気分や感情に関する単語をクエリとし検索することがあると考えられる。「映画 楽しい」とクエリに入力し検索した場合、ユーザは映画を見ると楽しい気分になる作品名を知りたいと考えられる。しかしこのように検索した場合、Q&A サイトや映画のタイトルに「楽しい」

が含まれている映画の公式サイトなどが検索結果の上位に表示される。Q&A サイトの結果ならば、目的的感情になる映画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いることや信頼できる情報でないことから、ユーザにとって有益な情報とは言い難い。また、タイトルに「楽しい」が含まれている映画が楽しい気分になるとは限らない。このように、現在のキーワード検索では、コンテンツを見たり聞いたりした時に感じる印象や感情をクエリとした検索を行うことが困難であるという問題が起きている。

一方、人間は日常生活をする上で、複雑な感情に沿って行動を選択することがある。例えば、音楽を聴く際、毎日同じ音楽を聴くのではなく、その日の気分やなりたい気分を考慮して、アーティスト名や曲名、プレイリストなどから曲を選択していくことがある。このように、人間は感情に沿って行動を選択することで、気分を良い方向へ導こうとする傾向がある。これらのことから、我々はユーザの今の気分やなりたい気分を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た。そこで我々は、ユーザの気分を考慮した検索システムの提案を行う。本研究では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。Web 上の名言集サイトの名言ナビ^(注1)では、約3万もの大量の名言がある。このようにインターネット上の大量の名言の中から、自分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、その感情に合った名言を提示する手法を提案する。以前の我々の研究[1]では、感情のポジティブ・ネガティブ(P/N)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 値だけではユーザの気分にあった名言が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を提案し、その感情ベクトルに基づいた名言検索手法の提案を行う。具体的には、中村明の感情表現辞典[2]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。本論文では、名言を読んで人が受けるであろう感情を「名言の感情」と呼ぶ。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感情の定量化を行う。ユーザの気分と名言の感情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。名言単語値を求める際、3つの正規化手法を提案し、名言単語値を求める際に適した式を検証する。また、興味のあるカテゴリを作成し、ユーザの興味となりたい気分にあった名言を提示する。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章では多次元感情ベクトルについてを、第4章では次元結合についてを、第5章では名言単語値

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を述べる。

2. 関連研究

近年、テキスト評価分析や文書の極性判定、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。テキスト評価分析や文書の極性判定に関する研究では、レビューなどのテキスト文書から評価を抽出し、分析している研究が多数あり、ポジティブ(肯定)・ネガティブ(否定)に注目した研究が多くある。Daveら[3]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取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、高村ら[4]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。

次に、多次元の感情ベクトルを提案した研究について説明する。Plutchik[5][6]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、これらの強度と組み合わせによって複雑な感情が作り出されるとしている。kampsら[7]は、WordNetを利用して、形容詞の評価極性を判定する手法を提案している。kampsらは、極性を表す組を「good-bad」、「weak-strong」、「passive-active」の3つを使用している。評価したい形容詞が各極性のどちらとより意味が類似しているかを判定し、形容詞の評価極性値を求めている。

上記のように「ポジティブ ネガティブ」のような体極関係を扱っている研究が多い。しかし、我々は感情は必ずしも対極関係で表せるとは限らないと考えている。感情の対極関係を用いていない研究も多数ある。徳久ら[8]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、恐い、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。遠藤ら[9]は、Web文書中の感情表現とその係り受け関係から書き手の感情が推定可能な表現を抽出する方法を提案している。文の感情を分類する際に中村明の感情表現辞典を使用している。本研究では、遠藤らと同様に感情表現辞典の10次元のベクトルを基本として名言に適した感情ベクトルを提案する。

上記のように、テキスト評価分析や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。熊本ら[10]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

(注1): 名言ナビ <http://www.meigennavi.net/>

表 1 10 次元の感情ベクトル

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, 驚き, 慌てる

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 うれしい」，「怒る 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．熊本ら [10] [11] 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感情を抽出する手法を提案する．

3. 多次元感情ベクトル

はじめに，ユーザ実験により既存の感情ベクトルが名言に適しているかどうかを判断する．

3.1 10 次元感情ベクトル

本研究では，名言に適した多次元の感情ベクトルを用いてユーザの気分にあった名言を提示する手法の提案を行う．様々な多次元ベクトルのモデルの中から，単語を分類して感情の多次元ベクトルを提案し，その単語の分類一覧を公開していることより，中村明の感情表現辞典を用いる．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．この辞典では，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「好」は「好カテゴリ」と呼ぶ．

3.2 10 次元による名言からの感情の自動抽出

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．ここでは，名言の中に 10 次元の感情ベクトルを示す単語が一つでも含まれていれば，その感情ベクトルがその名言の感情を示すとする．また，一つの名言に複数の感情ベクトルを示す単語が含まれている場合，その名言は複数の感情を示すとする．例えば「苦しくなったら，苦しみを味わえるだけ生きているんだと感謝した．嬉しいときはまた喜べるんだと，また感謝した」という名言の場合，「苦しみ」という単語が感情表現辞典の「哀カテゴリ」に含まれ，「嬉しい」という単語が「喜カテゴリ」に含まれているので，「哀」「喜」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．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．

3.3 予備実験

3.2 節で示した既存の 10 次元の感情ベクトルが名言に適しているかどうかを判断するために実験を行った．

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，被験者により各名言が 10 次元の感情ベクトルのどの感情ベクトルを持っているのかを判定した．幅広い年齢層に対応できる名言の感情ベクトルを生成するため，被験者は，16 歳から 22 歳，23 歳から 34 歳，35 歳から 44 歳，45 歳から 54 歳，55 歳から 64 歳，65 歳以上の 6 つの年代に分類し，各年代男女それぞれ 10 人ずつ，合計 120 名とした．被験者は各々の主観で名言の感情を判断し，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」を選び，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．尚，感情に左右されやすい実験なので，平常心の時に実験を行った．

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，120 人の被験者が選んだ感情のうち，感情の合計数の上位 2 位までを正解データとした．上位二位までの適合率は 37 %，再現率 35 %，F 値 36 % 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った．その原因の 1 つとして 10 次元感情ベクトルでは次元の数が多く，10 次元は名言に合った感情ベクトルではないということが分かった．また，被験者より 10 次元では次元数が多く，名言を判定するのは困難であったという意見が多数あった．そこで，感情ベクトルの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 節の予備実験結果から，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 節の予備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「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「喜カテゴリ」，「好カテゴリ」，「安カテゴリ」であった．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「喜カテゴリ」と「好カテゴリ」と「安カテゴリ」は共起感情である．次元を結合するにあたり，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．まずはじめに，3.3 節の実験で用いた 100 名言中にある感情カテゴリが他の感情カテゴリと共起している名言の数を求める．ここでの感情カテゴリは被験者 120 名が判定した感情カテゴリを用いる．この名言の数を共起頻度数 CW_{ij} と呼ぶ．例えば，表 2 の場合「好カテゴリ」と「喜カテゴリ」が共起している名言が 100 名言中 38 名言あった事を示す．よって「好カテゴリ」と「喜カテゴリ」の共起頻度数 CW_{ij} は 38 となる．このように，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

表 2 好カテゴリをベースにした共起頻度数と共起確率

Table 2 Co-occur frequency and Co-occur probability based of suki category.

	喜	怒	哀	怖	恥	好	厭	昂	安	驚
CW_{ij}	38	5	6	0	3	-	5	17	28	3
AW_j 全体	40	10	54	25	21	47	41	25	40	12
共起確率	0.95	0.5	0.11	0	0.14	-	0.12	0.68	0.7	0.25

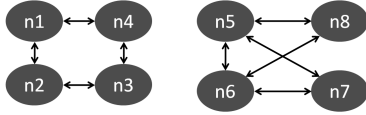


図 1 次元結合の例 1

Fig. 1 Example 1 of dimensions binding.

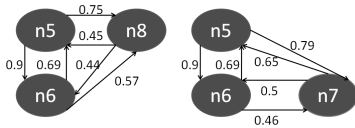


図 2 次元結合の例 2

Fig. 2 Example 2 of dimensions binding.

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，「好カテゴリ」をベースにした結果を表 2 に示す．

4.2 次元結合

10 次元から次元を削減するために，感情カテゴリをノードとし，共起確率が閾値 β 以上のカテゴリを接続して有向グラフを生成する．この有向グラフを感情グラフと呼ぶ．例えば，閾値 β を 0.35 とした場合，表 2 の「好カテゴリ」では，共起確率が閾値以上の「安カテゴリ」，「喜カテゴリ」，「昂カテゴリ」，「怒カテゴリ」それぞれと共起関係になる．この感情グラフを用いて次元結合を以下の条件で行う．まず，各ノードが双方向に接続している場合がある．例えば，閾値 β を 0.35 とした場合，「好カテゴリ」を対象とすると「喜カテゴリ」との共起確率が閾値以上なので，共起関係「好」→「喜」として接続する．逆に「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 0.81 となり，閾値以上のため，共起関係として接続する．よって「喜カテゴリ」と「好カテゴリ」は双方向に接続されている．このように作成した感情グラフにおいて，以下の 2 つの条件の場合，感情カテゴリの次元を結合する．

- 双方向に接続されており，且つ閉路の場合

図 1 の左図より， $n1$ と $n2$ ， $n1$ と $n4$ ， $n2$ と $n3$ ， $n3$ と $n4$ は互いに共起確率が閾値以上であり，感情グラフにおいて双方向に接続されている．また， $n1$ ， $n2$ ， $n3$ ， $n4$ は閉路になっているので， $n1$ ， $n2$ ， $n3$ ， $n4$ を結合する．図 1 の右図の $n5$ ， $n6$ ， $n7$ ， $n8$ の場合， $n7$ と $n8$ が双方に接続されていない．つまりは， $n5$ ， $n6$ ， $n7$ ， $n8$ は閉路になっていないため，結合しない．

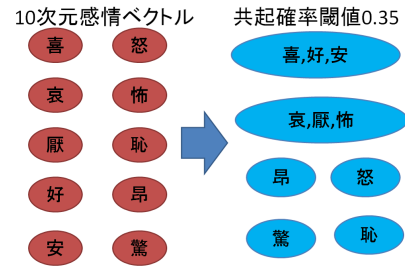


図 3 閾値を用いた次元の結合

Fig. 3 Dimensions binding using threshold value.

表 3 適合率の比較

Table 3 Result of comparison experiment of precision.

	次元数	適合率	再現率	F 値
10 次元	10	37 %	35 %	36 %
閾値 $\beta=0.35$	6	58 %	57 %	58 %
閾値 $\beta=0.45$	7	56 %	47 %	51 %
閾値 $\beta=0.6$	8	56 %	54 %	55 %

- 閉路が複数ある場合

図 1 の右図の場合，図 2 のように 2 つの閉路が存在する．このような場合，共起確率が低い値を持つ部分グラフは，そのノード間の接続が弱くつまりは，そのノードの感情カテゴリの関係が弱いと考える．そこで，共起確率が高い値を持つ閉路を構成しているノードを結合する．図 2 の左図と右図を比較した場合，右図の共起確率が左図に比べて高い値となっている．よって， $n5$ ， $n6$ ， $n7$ を結合する．

4.3 次元数の決定

次元を削減するために，10 次元ベクトル，共起確率閾値 $\beta=0.35, 0.45, 0.6$ を対象として，それぞれ適合率，再現率，F 値を求め，比較する．適合率の比較実験の結果を表 3 に示す．表 3 より， $\beta=0.35$ が最も高い適合率を示していることがわかる．よって，名言の為の感情ベクトルの次元を 6 次元とする．図 3 に名言の次元数として決定した $\beta=0.35$ の時の感情ベクトルを示す．

5. 名言単語値

名言の為の 6 次元感情ベクトルを用いて名言の値を求めていく．我々は名言の値を名言に含まれる単語の感情値から求める．本論文ではこの名言に含まれる単語の感情値を名言単語値と呼ぶ．感情表現辞典には感情カテゴリに含まれる単語は記載されているが，その単語は定量化されていない．そこで我々は名言に含まれる感情ベクトルを示す単語を定量化し名言単語値を決定する．そして，この名言単語値を用いて名言の感情を定量化することを行う．本論文では，3 つの正規化手法を提案し，それぞれの手法から名言単語値を求めていくことを提案する．

5.1 名言単語値の求め方

以下の手順で名言単語値を求めていく．

(1) まずインターネットから無作為に 17446 名言を抽出する．

(2) 次に名言それぞれを日本語構文解析システム KNP [12] を用いて，名言を基本句ごとに分類していく．分類した基本句

を構造解析処理を行うことで、基本句を6次元に分類していく。構文解析処理については、5.2節で説明を行う。

(3) 6次元に分類した基本句を形態素解析器 Juman [13] を用いて、単語ごとに分類する。この際、Juman の出力をそのままの形では用いずに、熊本らの研究で用いられているルールを適用する。例えば「前進できない」のような語は、Juman によりサ変名詞「前進」、動詞「できる」、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「前進できない」という動詞1語として扱うことができる。このルールを適用することで否定語にも対応できるようになる。本研究では、熊本らの研究で用いられている17個ルールを適用する。

(4) 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。 tf 値は各カテゴリ毎、品詞毎に求める。この時、名詞、動詞、形容詞には感情抽出に対して必要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削除する。

(5) 先程、求めた tf 値を3つの正規化手法を適用し、名言単語値を求める。3つの正規化手法については、5.3節で説明を行う。

5.2 構文解析処理

日本語構文解析システム KNP を使用し、名言を基本句ごとに分類し、基本句ごとに係り受け関係を取得する。基本句とは、京都大学テキストコーパスに含まれる格関係、照応・省略関係、共参照タグ付きコーパスにおいて各種関係を付与している単位のことを示している「今日はいい天気だ」を入力文にした場合、「今日は」「いい」「天気だ」の3つの基本句が生成される。この文の係り受け関係は「今日は」は「いい」に係っており、「いい」は「天気だ」に係っている。「今日は」「いい」の関係を1次の係り受け関係と呼ぶ。ここで作成した1次の係り受け関係を2つ組み合わせた関係を2次の係り受け関係と呼び、3つ組み合わせた係り受け関係を3次の係り受け関係と呼ぶ。すべての基本句に対して係り受け関係を作成する。ここで、最後の基本句のように係り受け先がない場合は空白として扱う。

5.2.1 基本句の分類

先程取得した基本句を6次元に分類する。基本句を感情表現辞典とマッチングさせ、基本句を6次元に分類していく。基本句と感情表現辞典をマッチングさせる際、前方手法、後方手法と前後手法の3つの手法を用いる。

5.2.2 前方手法

先程取得した3次の係り受け関係を使用し、基本句と感情表現辞典をマッチングさせることで基本句を6次元に分類する。この時、基本句を形態素解析を用いて単語ごとに分類し、基本句に含まれる単語と感情表現辞典に含まれる単語が一致していれば、マッチした感情表現辞典のカテゴリに基本句を分類する。前方手法では、3次の係り受け関係の前の基本句から感情表現辞典とマッチングさせる手法である。例えば「楽しい」「青春の」「夢を」「酒のような」という3次の係り受け関係の場合、まず「楽しい」と感情表現辞典をマッチングさせる。この場合、「楽しい」は感情表現辞典の「喜, 好, 安」カテゴリに含まれるので「楽しい」「青春の」「夢を」「酒のような」を「喜, 好, 安」カテゴリの基本句として取得する。次に、2番目の基本句

である「青春の」と感情表現辞典をマッチングする。もしマッチしていれば「青春の」「夢を」「酒のような」をマッチしたカテゴリの基本句として取得する。最後に、3番目の基本句である「夢を」と感情表現辞典をマッチさせる。マッチしていれば「夢を」「酒のような」をマッチしたカテゴリの基本句として取得する。

5.2.3 後方手法

前方手法と同様に、3次の係り受け関係を使用し、基本句と感情表現辞典をマッチングさせることで基本句を6次元に分類する。この時、基本句を形態素解析を用いて単語ごとに分類し、基本句に含まれる単語と感情表現辞典に含まれる単語が一致していれば、マッチした感情表現辞典のカテゴリに基本句を分類する。後方手法では、3次の係り受け関係の後ろの基本句から感情表現辞典とマッチングさせる手法のことである。例えば、「楽しい」「青春の」「夢を」「酒のような」のような3次の係り受け関係の場合、まず後ろの基本句である「酒のような」と感情表現辞典をマッチングさせる。もしマッチしていれば「楽しい」「青春の」「夢を」「酒のような」をマッチしたカテゴリに分類する。次に、その前の基本句である「夢を」と感情表現辞典をマッチングさせる。最後に「青春の」と感情表現辞典をマッチングさせる。

5.2.4 前後手法

前後手法では、2次の係り受け関係を利用し、基本句と感情表現辞典をマッチングさせることで基本句を6次元に分類する。この時、2次の係り受け関係の真ん中の基本句を形態素解析を用いて単語ごとに分類し、基本句に含まれる単語と感情表現辞典をマッチングさせ、もしマッチしていれば、マッチした感情表現辞典のカテゴリに基本句を分類する。例えば「後味の」「悪い」「事だ」の場合、真ん中の基本句である「悪い」と感情表現辞典をマッチングさせる。もしマッチしていれば「後味の」「悪い」「事だ」の3つの基本句をマッチしたカテゴリに分類する。このように、2次の係り受け関係の真ん中の基本句と感情表現辞典をマッチングさせる手法である。

5.3 名言単語値の計算手法

名言単語値を計算する際、3つの正規化手法を提案する。ここで提案した手法は7.2節で比較実験を行い名言単語値を求める際に適した計算手法を決定する。

(1) 縦方向の正規化

縦方向の正規化について説明をしていく。以前までの我々の研究 [14] では縦方向の正規化を使用し研究を進めてきた。本論文では、他の正規化手法との比較を行い、名言単語値を求める際に適した正規化の手法について議論する。

5.1節で取得した単語の中で各品詞ごとのカテゴリの中で最大の tf 値を1として各々の単語を正規化した値を名言単語値とする。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。

(2) 横方向の正規化

5.1節で取得した品詞ごとの単語の中で、単語の6次元のトータル tf 値を1として各々のカテゴリごとの tf 値を正規化した値を名言単語値とする。各次元で求めた単語の値を6次元に

表 4 名言数と重み

単語	喜, 好, 安	怒	怖, 哀, 厭	恥	昂	驚
名言数	2554	142	1546	69	166	46
重み	1	17.99	1.65	37.01	15.39	55.52

結合し, 名言単語辞書に登録する.

(3) 縦横複合の正規化

(2) の横方向の正規化を使用する場合, 6 次元の中で 1 つの次元にしか単語が出現しておらず, なおかつ 1 回しか出現していない場合でも, その単語の名言単語値が 1 になってしまうという問題がある. そこで, 縦方向の正規化と横方向の正規化を複合した正規化手法を提案し, 横方向の正規化の問題を解決していく. 縦横複合の正規化では, 6 次元の中で 3 つ以上の次元に値を持つ場合は, 横方向の正規化を適用し, 3 つ未満の場合は縦方向の正規化を適用する. 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.

5.4 重み

5.1 節で名言を分類した際に各次元に含まれる名言数に大きな偏りがあった. 名言数に偏りがあると横方向の正規化を使用する際に大きな問題が起こる. 例えば, ある次元に分類された名言数が 1000 個でその中に含まれる「楽しい」の出現頻度が 50 回だったとする. 一方, ある次元に分類された名言数が 100 個でその中に含まれる「楽しい」という単語の出現頻度が 50 回だったとする. この場合, 横方向の正規化を使用すると出現頻度が同じため, 同じ値の名言単語値を持つことになる. しかし, 同じ出現頻度でも 1000 名言の次元の「楽しい」より 100 名言の次元の「楽しい」の方が強い感情を持つ次元であることがわかる. そこで, この名言数の偏りを考慮した重みを提案し, 名言単語値の再考を行う. 各次元の名言数と名言数の偏りから求めた重みを表 4 に示す. 表 4 で示されている重みに出現頻度にかけることで名言による偏りをなくした. 重みをかけた後, 名言単語値の正規化を行う.

5.5 感情表現辞典拡張

5.2.2 節の後方手法で基本句の分類を行ったところ, 「怒」, 「恥」, 「昂」, 「驚」の 4 つのカテゴリにマッチした回数が大幅に少ないことが分かった. 基本句と感情表現辞典とのマッチ回数, マッチした単語数, 感情表現辞典に含まれる単語数を表 5 に示す. そこで「怒」, 「恥」, 「昂」, 「驚」の 4 つのカテゴリにマッチする回数を増やすために, 感情表現辞典の「怒」, 「恥」, 「昂」, 「驚」の 4 つのカテゴリの単語を増やすことを提案する. 4 つのカテゴリに含まれている単語を Yahoo!類語辞典^(注2)にかけて類語を取得し, 得られた類語を感情表現辞典に追加する. 拡張をする際, 感情を示さない類語に関しては追加しないようにした. 類語処理を行い, 感情表現辞典拡張を行い, 基本句分類を行い, 後方処理での基本句と感情表現辞典とのマッチ回数, マッチした単語数, 感情表現辞典に含まれる単語数を表 6 に示す.

6. 名言値の求め方

本研究では, 名言の感情を示す 6 次元に定量化した値を名言

表 5 感情表現辞典とのヒット数

Table 5 the number of hit with sentiment dictionary.

カテゴリ	ヒット回数 (回)	ヒットした単語数 (語)	感情表現辞典の単語数 (語)
喜, 好, 安	9529	117	579
怒	584	23	207
怖, 哀, 厭	4917	146	965
恥	123	7	70
昂	818	28	277
驚	119	8	144

表 6 感情表現辞典拡張とのヒット数

Table 6 the number of hit with extended sentiment dictionary.

カテゴリ	ヒット回数 (回)	ヒットした単語数 (語)	感情表現辞典の単語数 (語)
喜, 好, 安	9529	117	579
怒	584	35	285
怖, 哀, 厭	4917	146	965
恥	189	9	97
昂	2987	53	467
驚	126	10	184

値と呼ぶ. 名言値は 5 章にて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. 複数の文から構成される名言は文毎の感情が異なる場合がある為, 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.

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

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 は単語 k 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

以前の我々の研究 [1] では, 複数の文から構成される名言の各文がどれだけその名言の感情に影響しているかの分析を行った. その結果, 最後の文の感情の 88 % がその名言全体の感情と同じであるのに対し, その他の文は 63 % がその名言の感情と同じである結果となった. つまりは, 最後の文の感情がより名言全体の感情に影響することがわかった. よって, 複数の文から構成される名言値を以下の式で求める.

$$MPN_j = \frac{\sum_{s=1}^{n-1} MPN_s + \alpha \times MPN_{jn}}{n} \quad (3)$$

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_s は名言 j に含まれる文の名言値を示す. ここでの MPN_s は, 式 (2) を用いて, 各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, 我々の以前の研究より, α の値を 1.4 とする.

7. 実験

7.1 構文解析処理の比較実験

5.2 節で提案した, 前方手法, 後方手法, 前後手法, 前方手

(注2): Yahoo!類語辞典 <http://dic.yahoo.co.jp/>

法 (辞典拡張後), 後方手法 (辞典拡張後), 前後手法 (辞典拡張後), の 6 つの手法の比較のための実験を行った。まず, インターネット上の名言サイトである名言ナビから無作為に 250 個の名言を取得した。各々の名言を, システムが判定した感情 (1) と被験者が名言を読んだときに感じる感情 (2) を比較し, 適合率, 再現率, F 値を求める。

(1) システムの判定

5 章で求めた名言単語値を用いて, 250 名言の名言値を求めた。名言単語値の求め方は 6 つの手法を使用した。

- 前方手法
- 後方手法
- 前後手法
- 前方手法 (辞典拡張後)
- 後方手法 (辞典拡張後)
- 前後手法 (辞典拡張後)

各手法で取得した基本句から縦横複合の正規化を用いて名言単語値を求め, 6 章で説明を行った名言値を求める式を使用し, 名言値を求める。求めた名言値から 6 次元の中で値が高い上位 2 位までをシステムの判定とした。ここで, 6 次元の値がすべて 0 の場合は対象外とした。

(2) 被験者の判定

被験者 19 名に対し, 250 名言を使用した実験を行ってもらった。実験環境は, 感情に左右されやすい実験である為, 平常心の際に実験を行った。被験者は, 名言が 6 次元の名言感情ベクトルのどの感情を持っているのかを主観で判断した。名言感情ベクトルは「喜, 好, 安」「怒」「哀, 怖, 厭」「恥」「昂」「驚」の 6 次元と「どの名言感情ベクトルにも当てはまらない」の中から最大 2 つまで選択することを可能とした。被験者の判定は, 各々の名言で「どの名言感情ベクトルにも当てはまらない」を除いた 6 つの選択肢で回答数が多い上位 2 位までの感情ベクトルを対象とした。

比較実験の結果

実験結果を表 7 に示す。表 7 より基本句を取得する際, 後方処理を行い, 6 次元に分類する手法が良いことが分かった。また, 感情表現辞典を拡張することで更に F 値を向上させることを可能にした。表 8 にカテゴリ毎の実験結果を示す。表 8 より, 以前の我々の研究 [14] と比較した際, 全体の F 値向上だけでなく「昂」カテゴリが大幅な F 値向上を図ることが可能となった。しかし「怒」「恥」「驚」の 3 つのカテゴリの F 値は依然良い結果が得られていない。そこで, 今後は全体の F 値向上だけでなく「怒」「恥」「驚」の 3 つのカテゴリの F 値向上を目指したい。また人が名言を読んだときに「はっと」したり, 「なるほど」と驚くことがある。しかし, システムでは, 人間が感じる驚きを判定するのが困難であることがわかった。そこで我々は「驚」の次元を感情ベクトルから削除することも視野に入れている。

7.2 名言単語値の計算手法の比較実験

名言単語値の式の比較のための実験を行った。まず, インターネット上の名言サイトである名言ナビから無作為に 250 個の名言を取得した。各々の名言を, システムが判定した感情 (a) と

表 7 6 つの手法の比較実験結果

Table 7 Result of comparison experiment for six method.

	前方	後方	前後	前方 (拡張後)	後方 (拡張後)	前後 (拡張後)
適合率	53.8 %	58.9 %	56.5 %	54.4 %	59.4 %	56.0 %
再現率	44.4 %	48.1 %	46.8 %	45.9 %	50.2 %	48.2 %
F 値	48.7 %	52.9 %	51.2 %	49.8 %	54.4 %	51.8 %

表 8 カテゴリ毎の実験結果

Table 8 Result of experiment for each category.

	全体	喜, 好, 安	怒	怖, 哀, 厭	恥	昂	驚
適合率	59.4 %	67.0 %	14.2 %	50.5 %	0 %	63.2 %	0 %
再現率	50.2 %	87.4 %	0.08 %	58.5 %	0 %	36.8 %	0 %
F 値	54.4 %	75.9 %	10.5 %	58.5 %	0 %	46.5 %	0 %
システムの判定数	-	176 回	14 回	109 回	0 回	68 回	0 回
被験者の判定数	-	135 回	24 回	94 回	24 回	117 回	40 回

被験者が名言を読んだときに感じる感情 (b) を比較し, 適合率, 再現率, F 値を求める。

(a) システムの判定

5 章で求めた名言単語値を用いて, 250 名言の名言値を求めた。名言単語値の求め方は 5 つの手法を使用した。

- 手法 1: 縦方向の正規化
- 手法 2: 横方向の正規化
- 手法 3: 重みを考慮した横方向の正規化
- 手法 4: 縦横複合の正規化
- 手法 5: 重みを考慮した縦横複合の正規化

各手法で求めた名言単語値を使用し, 6.2 節で説明した名言値を求める式を使用し, 名言値を求める。求めた名言値から 6 次元の中で値が高い上位 2 位までをシステムの判定とした。ここで, 6 次元の値がすべて 0 の場合は対象外とした。

(b) 被験者の判定

7.1 節の実験の被験者の判定を本実験の被験者の判定とする。実験内容, 実験環境などは 7.1 節の実験と同様である。

比較実験の結果

実験結果を表 9 に示す。表 9 より縦横複合の正規化を使用して名言単語値を求め, その名言単語値を使用して名言値を求めた場合が一番良い結果を得ることができた。これは, 5.3 節で述べた横方向の正規化での問題点を解決していることを示している。実験結果から, KNP を用いて名言を基本句に分類し, 3 次の係り受け関係を作成する。先程作成した 3 次の係り受け関係を後方処理で感情表現辞典を拡張した辞典とのマッチングを行うことで, 基本句を 6 次元に分類する。分類した基本句を 5.1 節の (3), (4) を適用し, tf 値を求める。ここで求めた tf 値を縦横複合の正規化を用いて, 名言単語値を求める。ここで求めた名言単語値を名言単語辞書に格納する。ここで求めた名言単語値を 6 章の名言値の式を適用し, 名言値を求める。これらの手法を本研究の提案手法とする。

8. プロトタイプシステム

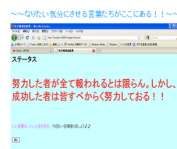
提案手法のプロトタイプシステムを開発した。システムを起動すると図 4 の左図のように, 興味のあるカテゴリとなりたい気分の選択画面が出てくる。ユーザには, リストボックスにあ

表 9 計算手法の比較実験実験の結果

Table 9 Result of comparative experiment for calculation method.

	手法 1	手法 2	手法 3	手法 4	手法 5
適合率	52 %	58 %	50 %	60 %	40 %
再現率	41 %	46 %	39 %	48 %	31 %
F 値	46 %	51 %	43 %	53 %	36 %

気分にあった名言検索手法の提案！！



興味のあるカテゴリを1つ選択してください！！



あなたがなりた気分を1つ選択してください！！



システム起動時画面

気分にあった名言検索手法の提案！！

あなたがなりた気分は———《幸福だ》
 あなたの興味のあるカテゴリは———《音楽》
 あなたに合った名言は———

幸福というのは、いつも持ち歩いているものです。
 みんなのポケットにも、たくさん入っているんですよ。

さだまさし



さだまさし(本名:佐田 善三(読みは同じ)、1952年4月10日-)は長崎県長崎市出身のシンガーソングライター、タレント、小説家、ファンとスタッフの間では「まっさん」の愛称で親しまれている。身長は165cm、血液型はA型。

いい言葉は、いい人生を作る。今日一日頑張らねようよ♪

引用

検索結果表示画面

図 4 プロトタイプシステムの画面

Fig.4 Display of Prototype System

興味のあるカテゴリとなりた気分のリストから1つ気分を選択してもらい、検索ボタンを押してもらうと、図4の右図のように、ユーザのなりた気分と興味のあるカテゴリに合った名言が提示されるシステムとなっている。提示名言は、Web上の名言サイトの名言ナビの名言の中からユーザの興味のあるカテゴリとなりた気分にあった名言、名言の発言者、発言者の情報と発言者の画像を提示している。

図4の結果は、興味のあるカテゴリを「音楽」、なりた気分を「幸福だ」にした場合の実行例になっている。なりた気分には、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。気分の値は、名言単語辞書の形容詞の値を使用している。興味のあるカテゴリは、名言ナビの発言者の情報から16個のカテゴリを作成した。カテゴリは{スポーツ、音楽関連、経営・経済関連、歴史、宗教・思想家、皇帝・王、評論家、映画関連、タレント、学者(文系)、学者(理系)、文芸、代表者、作家、政治、ジャンルなし}となっている。例えば、アインシュタインなら学者(理系)に分類されている。発言者の情報はWikipediaから引用し、発言者画像はYahoo APIの画像検索から画像を検索し提示している。今後は、プロトタイプシステムのユーザインタフェースを再考し、より使いやすいシステムにしていきたい。

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、名言を対象にしたりた

い気分と興味のあるカテゴリに合った名言を提示するための手法の提案を行った。本論文では、構文解析処理を行い、名言を基本句ごとに分類した。この基本句と感情表現辞典を拡張したものを後方処理を用いてマッチングさせ、名言に含まれる係り受け関係を6次元に分類した。また、名言単語値を求める際に縦横複合の正規化手法を提案した。今後は、開発したプロトタイプシステムを使用した実験も行っていきたい。その他の今後の課題を以下に示す。

● 気分についての再考

本研究では「なりた気分」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利用した。しかし、選択した形容詞が「なりた気分」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う気分として用いていきたい。また、ユーザの様々な感情表現に対応したシステムの提案も行いたい。

● 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。

文 献

- [1] 高岡幸一, 瀬本明代. 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.
- [2] 中村明. 感情表現辞典. 東京堂出版, 1993.
- [3] Kushal Dave, Steve Lawrence, and David M. Pennock. Mining the peanut gallery: opinion extraction and semantic classification of product reviews. In *Proceedings of the 12th International World Wide Web Conference (WWW)*, pp. 519–528, 2003.
- [4] 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.
- [5] R. Plutchik. The multi factor-analytic theory of emotion. *Psychology*, Vol. 50, pp. 153–171, 1960.
- [6] R. Plutchik. The nature of emotions. *American Scientist*, Vol. 89, pp. 344–350, 2001.
- [7] Jaap Kamps, Maarten Marx, Robert J. Mokken, and Maarten De Rijke. Using wordnet to measure semantic orientation of adjectives. In *proceedings of the 4th International Conference on Language Resources and Evaluation (LREC-2004)*, pp. 1115–1118, 2004.
- [8] 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.
- [9] 遠藤大介, 齋藤真実, 山本和英. 係り受け関係を利用した感情生起表現の抽出. 言語処理学会第12回年次大会, pp. 947–950, 2006.
- [10] 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).
- [11] 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.
- [12] 黒橋禎夫. 結構やるな knp. 情報処理学会, Vol. 41, , 2000.
- [13] 黒橋禎夫, 河原大輔. 日本語形態素解析システム JUMAN VERSION 6.0.
- [14] 高岡幸一, 瀬本明代. 名言のための多次元感情ベクトルの生成. Webとデータベースに関するフォーラム 2011 (WebDB2011), 2011.