

気分距離に基づく名言検索手法の提案

高岡 幸一[†] 灘本 明代[‡]

甲南大学理工学部情報システム工学科[†] 甲南大学知能情報学部[‡]

1. はじめに

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たが、インターネットを活用して人々を元気にするシステムは少ない。また現代人は、競争社会の中で生きており、心の病にかかる人が多くなってきている。例えば気分が優れなかったり落ち込んだりした時に、自分に合ったポジティブな言葉を提示されると落ち込んだ気分が少しは良い気分になれると思われる。そこで我々は、人々を元気にするシステムとして、ユーザのなりたい気分と合った名言を提示する気分検索を提案する。具体的には、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検索して提示するシステムの提案を行う。

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。これらの理由により本研究では、名言を用いてユーザを元気にしたいと考える。

2. 気分距離

我々の提案する気分距離とは、なりたい気分と今の気分との距離である。距離を測る方法として、高村らの単語感情極性対応表の感情極性値を用いる[1]。感情極性値は二値属性に分けられ、-1 から+1 の実数値を割り当てている。-1 に近いほど negative, +1 に近いほど positive な言葉である。ここで単語感情極性対応表は、55125 語あるため、非常に単語が多い。その為我々はその中から、60 語の単語(楽しい, 悲しいなど)を使用する。そして中村明の単語表現辞書[2]を用いて、その単語を 10 種の感情カテゴリに分類する。

我々の提案する気分距離は、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。気分距離は、0 から+1 の実数値を割り当てている。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。

例えば、なりたい気分が「嬉しい」(値は 0.998871)であり、今の気分が「寂しい」(値は-0.9990)である場合、この気分距離は 0.9989505 となる。この場合は、気分距離が 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。

3. 名言 PN 値

名言 PN 値とは、1 つの名言がポジティブ (P) かネガティブ (N) かを表す尺度であり、-1 に近いほどネガティブ

ブ、+1 に近いほどポジティブな名言とする。

名言 PN 値を求める方法としてまず、高村ら[1]の単語感情極性対応表の感情極性値を使用することを考えた。しかしながら、単語感情極性対応表はネガティブな言葉が多く、ネガティブな名言ばかりになった。

そこで我々は、単語の出現頻度から新しいスコアである名言単語 PN 値を提案する。具体的には、Web サイトの名言集のポジティブ名言、ネガティブ名言から、形容詞、名詞、副詞の単語の出現頻度 tf 値を求める。

その tf 値を式 1 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。

$$y = \frac{X}{X_{\max}} \quad (1)$$

X は、一単語の出現頻度数を表し、

X_{\max} は、出現頻度で最大の数を表す。

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2で割り単語スコアを求める。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値の総数は、1323語である。

(1) 式で求めた、名言単語PN値を使用し名言PN値を求める。名言PN値は、藤村らのスコアリング手法[2]を基に以下の名言スコアリング手法の式 (2) を提案する。

$$\text{Score}(m_i) = \frac{P_{num} - N_{num}}{P_{num} + N_{num}} \quad (2)$$

$$(-1 \leq \text{Score}(m_i) \leq 1)$$

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数。同様に N_{num} は m_i にネガティブな言葉が出現する数である。

4. システムのフロー

まずユーザは、興味のあるカテゴリ (10 種類) と、今の気分と、なりたい気分を選択する。今の気分は、中村明の単語表現辞書[3]を用いた 10 種のカテゴリを使用し、なりたい気分は、10 種のカテゴリの中の 3 種を使用する。これは、経済産業省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[4]を用いて、10 種のカテゴリの中、「喜」、「好」、「昂」の 3 種がポジティブなカテゴリであることによる。しかし、昂カテゴリにはネガティブな言葉が多いため、なりたい気分

The Search for Best Quotes based on Sentiment Distance

[†]Kouichi TAKAOKA, and [‡]Akiyo NADAMOTO

[†]Faculty of Science and Engineering information system engineering department

[‡]Dept. of Intelligene and Informatics

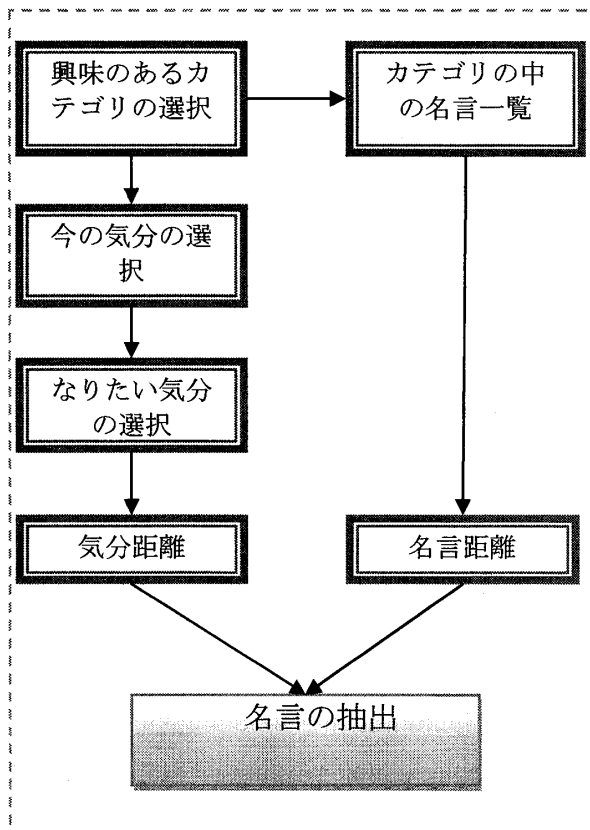


図 1: システムの流れ図

カテゴリから外す。また、安カテゴリは、ポジティブな言葉が多いため、なりたい気分カテゴリに入れる。これよりなりたい気分カテゴリは喜, 好, 安の 3 種を用いる。ここで、我々はユーザをポジティブな気分させたいので、なりたい気分をポジティブなカテゴリに制限する。

次にユーザが選択したなりたい気分と今の気分から気分距離を測り、興味あるカテゴリ一覧からそれぞれの名言距離を計算し、気分距離と名言距離が近い名言を抽出する。システムの流れ図を図 1 に示す。

5. 実験

我々は実験として、Web サイトの 100 個の名言を用いて、その 100 個の名言の名言 PN 値の平均値とその名言の中でポジティブな名言 (名言 PN 値が 0 より大きい) の数を調べた。

5.1. 実験手法

我々は実験手法として、提案手法の有用性を求めるために、以下の 3 つの手法を用いた。

手法 A: 単語感情極性対応表の感情極性法を用いる手法。

手法 B: ポジティブ, ネガティブの出現頻度から求めた名言単語 PN 値を使用した数値を用いた手法 (式 (1) のみ使用)。

手法 C: 名言単語 PN 値と名言スコアリング手法を用いた手法 (式 (1) と (2) を使用) の 3 種類を用いて比較実験を行った。

表 1: 実験結果

	平均値	ポジティブ名言の数
手法 A	-0.519550668	4 個
手法 B	-0.006967	62 個
手法 C	0.20454788	59 個

5.2. 実験結果

実験結果を表 1 に示す。実験結果より、我々の提案した手法 C は手法 A に比べて、良い結果となった。手法 C は、平均値がプラスになったことで、名言がポジティブであることがわかった。また手法 C の結果よりネガティブな名言 (名言 PN 値が 0 より小さいもの) が 13 個あったので、名言の中にもネガティブな名言がある発見があった。以下に良い例と悪い例を示す。

例:

良い例: 名言距離 (0.789473684210526)

別れることがなければ、めぐり逢うこともできない。

悪い例: 名言距離 (-1.0)

ドアを叩け、さすれば開かれん。

6. おわりに

本研究では、ユーザの興味のあるカテゴリを基に情報を探し、気分距離に合った名言をユーザに提示するシステムの提案をした。実験結果から高村らの単語感情極性対応表よりも単語数が少ないという問題が残った。

今回は、提案する情報として名言を用いたが、今後は情報を名言以外の情報を提示する研究もしていきたい。我々は、提案システムを用いてユーザの興味と気分距離に合った情報を提示して、人々を元気にしていきたいと考える。

謝辞

本研究の一部は、平成 21 年度科研費特定領域研究域「コミュニティ型コンテンツホール検索に関する研究」(課題番号: 21013044, 代表難本明代) 及び甲南大学平生太郎基金科学研究援助助成金による。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] 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, 3F1-03, 2004.6
- [3] 中村明, “単語表現辞書”, 東京堂出版
- [4] 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