E-007

感情語に基づくことわざ推薦システム
Proverb Recommendation based on Affective Words

佐藤 祥子† 高瀬 裕‡ 中野 有紀子‡
Shoko Sato Yutaka Takase Yukiko Nakano

1. はじめに

近年，インターネットの普及により，簡単に掲示板やブログなどに書き込みができるようになった．それらの内容としては日々の感想などが多いが，中には人には言えない愚痴や悩みなども見受けられる．そこで，このような文章を書く人に対して先人の知恵であることわざや故事成語を薦めることによって，ユーザの精神的負担を軽減できるのではないかと考えた．

前述を踏まえ，本研究ではユーザの日々の感想を書いた文書に対して，ことわざや故事成語を出力するシステムを構築することを目的とする．構築したシステムは，ユーザの文章とことわざに含まれる感情語から感情を推定し，ユーザの感情に合致したことわざや故事成語を出力する．

2. 感情語ベクトル空間の構築

本研究では，文章の感情を推定するために，感情語を収集し，これらに基づく感情語のベクトル空間を構築した．その作成方法を図 1 に示す．以下の節では，各ステップについて説明する．

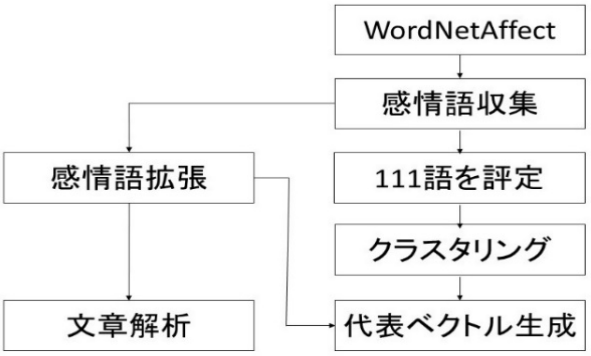

図 1　感情語ベクトル空間構築の手順

2.1 感情語の収集

ユーザの入力とことわざに対して感情を推定するために，感情を表現すると考えられる単語を収集する．まず英語の感情語をまとめた WordNetAffect[1]を参考に 111 語の感情語を選定した．

2.2 感情モデルに基づく感情語の評定

それぞれの感情語がどのような感情を表しているかを定量化するために，Plutchik による感情研究[2]を参考に，10 人の評定者に各感情語のアノテーションを行ってもらった．Plutchik は人間の感情は「喜び，信頼，心配，驚き，悲しみ，嫌悪感，怒り，予測」の 8 種類の基本感情から構成されていると考え，8 次元の感情モデルを提案している．評定者には，各感情語に対し，8 次元の感情について-5 から 5 までの値を付与してもらい，その際，強くその感情を感じる場合には正値，感じられない場合は 0，かけ離れていると感じた場合には負の値を選ぶように指示した．評定結果を表 1 に示す．縦に感情語，横に Plutchik の 8 次元の感情が記されている．アノテーションの結果を集計し，各次元の平均値をその感情語の評定結果とした．

表 1　評定結果

	喜び	信頼	心配	驚き	悲しみ	嫌悪感	怒り	予測
愛情	3	2.9	0.5	0.4	-0.2	-1.8	-1.3	-0.2
悲しみ	-2.7	-0.8	0.9	0.3	3.7	0.5	0.6	0.1
恐怖	-1.7	-1.6	2.7	2.0	2.0	1.7	0.5	0.7

2.3 感情語のクラスタリング

より細かい感情語分類を行うために，2.2 節の評定結果を k-menas 法によりクラスタリングした．繰り返しクラスタリングした結果を確認したところ，

・それぞれのクラスタで似た意味の単語がまとめられた
・クラスタ数を増やしても分類結果が大きく変わらない

という点から，クラスタ数は 17 が最適だと考えた．

2.4 代表ベクトルの生成

ユーザの入力文章の感情を推定するには，感情語の語彙は，WordNetAffect から取り出された 111 語では不十分だと考えた．そこで，単語を 200 次元でベクトル化することを可能にする Word2Vec[3]を使って，各感情語の拡張を行った．Word2Vec の学習データには語彙が豊富であるという理由から Wikipedia を使用した．それぞれの感情語に対してコサイン類似度上位 10 位までの単語を類義語とし，これらを同じクラスタの単語とすることにより感情語を 858 語まで拡張した．

感情語拡張後の各クラスタにおいて，クラスタ内の感情語を Word2Vec によってベクトル化し、その重心をクラスタの代表ベクトルとした．

3. ことわざ推薦システム

2 章で構築した感情語のベクトル空間を利用したことわざ推薦システムを提案する．本システムの処理の流れを図 2 に示す．

3.1 ユーザ文章の感情語ベクトルによる表現

ユーザの文章から感情語を取り出すために，ユーザの文章を形態素解析器 Mecab[4]により解析した．単語に区切り，感情語のいずれかと一致すると，その単語をユーザ文章の感情を表す単語として抽出した．検出された感情語を

†成蹊大学理工学研究科理工学専攻
‡成蹊大学理工学部

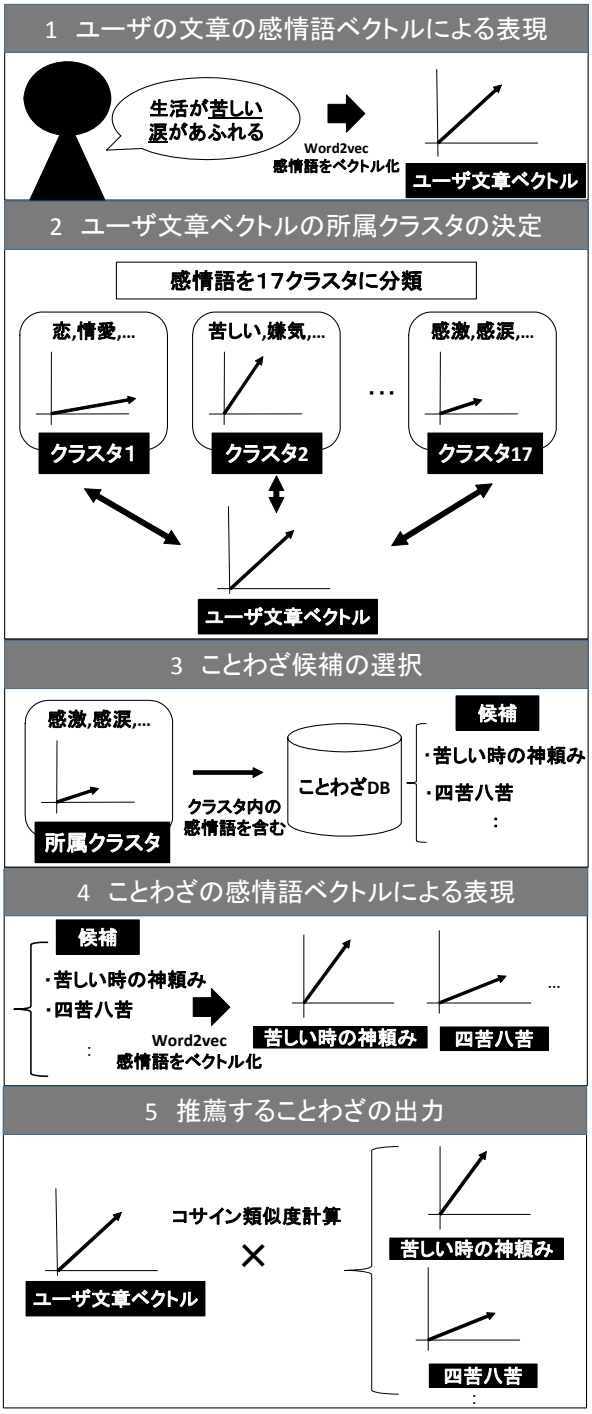

図 2　システム処理の流れ

Word2Vecにより200次元ベクトル化し，その重心をユーザ文章ベクトルとした．

3.2 ユーザ文章ベクトルの所属クラスタの決定

ユーザ文章の感情を特徴づけるために，2.3 節で作成した17クラスタを用いて，ユーザの文章を分類した．コサイン類似度を用い，17 クラスタの各代表ベクトルと 3.1 節で作成したユーザ文章ベクトルとの類似度を算出し，最も類似度の高い代表ベクトルのクラスタをユーザ文章ベクトルの所属クラスタとした．

3.3 ことわざの候補

形態素解析機 Mecab を用いて，三省堂 Web Dictionary[5]に掲載されていることわざとその説明の文章を解析し，858 種類の感情語のいずれかを含むものを選択した．その結果，733 種類のことわざ，慣用句，故事成語を得た．

次に，3.2 節で決定したユーザ文書ベクトルの所属クラスタに含まれる感情語を含むことわざを推薦することわざの候補とする．例えば，図 2-2 において，ユーザ文書ベクトルがクラスタ 2 に分類されている場合，図 2-3 では「苦しい」「嫌気」といった感情語を，ことわざかその説明のいずれかに含む「苦しい時の神頼み」，「四苦八苦」が推薦する候補のことわざとして選択される．

3.4 ことわざの感情語ベクトルによる表現

3.1 節でユーザ文章をベクトル化したのと同様の手法を用い，候補のことわざをベクトル化する．まず，ことわざとその説明を Mecab を使って形態素解析し，感情語を抽出した．これらの感情語を Word2Vec で 200 次元ベクトル化し，その重心をことわざのベクトルとした．

3.5 推薦することわざの出力

ユーザの入力文章に応じたことわざを選択するために，ユーザ文章ベクトルと 3.4 節で選定された候補とのコサイン類似度の計算を行う．その結果，類似度が高い値のことわざを推薦することわざとする．

本システムの実行例を示す．例えば，「仕事への後悔や怒りがわいてきます．大変だったけど楽しかった．なんだか涙が溢れます．過去や未来への不安しかありません．早くこの苦しみから解放されたい．」というユーザの文章に対して，表 2 のような出力結果が得られた．

表 2　出力結果

ことわざ：四苦八苦
説明：非常な苦しみ．あらゆる苦しみ．仏教で，人生の生・老・病・死の四苦に，愛別離苦・怨憎会苦・求不得苦・五陰盛苦の四苦を合わせたもの．

4. おわりに

本研究では，感情語から感情語ベクトル空間を構築することで，ユーザの文章とことわざの感情をベクトル空間で表現し，ユーザの感情と似た感情のことわざを出力するシステムを構築した．現在の 858 語の感情語では，ユーザの文章に対して感情語を検出することができない例もあった．そのため，語彙をさらに拡張することで，対応できる文章を増やす必要がある

参考文献

[1] WordNet-Affect: http://wordnet.princeton.edu/
[2] Plutchik, R.,"The Emotions",University Press of America(1991)
[3] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean,"Efficient Estimation of Word Representations in Vector Space" , Proceedings of Workshop at ICLR,(2013)
[4] Taku Kudo, Kaoru Yamamoto, Yuji Matsumoto,"Applying Conditional Random Fields to Japanese Morphological Analysis", Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing (EMNLP-2004), pp.230-237 (2004.)
[5] 三省堂 WebDictionary:http://www.sanseido.net/