HOKKAIDO UNIVERSITY

Title	重文および否定形を考慮した感情認識システムの改善
Author(s)	滝澤, 満; ジェプカ, ラファウ; 荒木, 健治
Citation	ファジィシステムシンポジウム講演論文集, 29, 774-777
Issue Date	2013-09-09
Doc URL	http://hdl.handle.net/2115/63645
Type	proceedings
Note	第29回ファジィシステムシンポジウム(29th Fuzzy System Symposium). 2013年9月9日～11日．大阪国際大学枚方キャンパス,枚方市.
File Information	FSS29-2013_774-777.pdf

Instructions for use

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

重文および否定形を考慮した感情認識システムの改善

Improvement of Emotion Recognition System Considering the Negative Form and Compound Sentence

○滝澤 満[1], ラファウ ジェプカ[2], 荒木 健治[3]
○Takizawa Mitsuru[1], Rafal Rzepka[2], Araki Kenji[3]
[1]北海道大学大学院情報科学研究科
[2]Graduate School of Information Science and Technology, Hokkaido University

Abstract: A system that is capable of recognizing an emotion of user's statement would be very effective to have a smooth conversation with users. Therefore, studies of the emotion recognition have been actively conducted, which quantify the emotion by analyzing the surface pattern in the sentence. However, recognizing the emotion from complicated sentences, such as compound sentences and sentences that include negative expressions remains difficult to solve. This paper describes our improvement of emotion recognizing system using parsing. The effectiveness of recognizing these complicated sentences was confirmed through the evaluation experiments.

1. はじめに

現在, PC やモバイル, タブレットなどの急速な普及に伴い, WWW (World Wide Web) の利用が非常に盛んになっている. その利用目的は電子メールをはじめ, チャット, ブログなどと多様な広がりを見せている. そうした中, WWW 上には膨大な言語資源が日々更新されている. しかし, こういった言語資源をコンピュータが人間のように理解を行うためには様々な課題が残されている. 実際に人間が理解を行うためには感情や一般常識, 文法など様々なことを考慮し, 理解を行なっている. このような扱いにくい情報に対し数値化を行うことは, コンピュータが人間と同等に言語資源を理解することに有益であると思われる.

近年, ユーザの感情を推定する研究が盛んに行われている. 感情推定のアプローチを大きく分けると二種類に分類される. 一つは表層的な表現に着目したもので, 構文構造を用いて感情の処理を行うものなどがある[1][2]. もう一つは, Web マイニングを用いたものである. テキストの情報を用い, Web 上のテキスト情報との因果関係を用いて感情を推定するものがある[3]. また, これらの手法を組み合わせ, 感情を推定する研究もなされている[4]. しかし, これらの手法だけでは感情を複数含む重文や, 二重否定のような否定形を考慮していないため, 適切な感情抽出を行うことは困難である. 本稿では, 重文や否定形を考慮した構文解析を導入する手法を提案し, 実験及びシステムの評価実験の結果について述べる.

2. 感情とは

2.1. 感情の定義

人間の感情は五感と価値基準などにより総合して捉えられる複雑なものであり, その感情を判断するには人間であっても困難を伴う. 例えば,「日本酒を飲んだ」という文では, お酒が苦手で嫌だと感じる人もいれば, その逆にお酒が好きで嬉しいなどと感じる人もいる. このように同じ発話でも人の価値観などによりその内容の捉え方は異なり, 感情も異なる場合がある.

2.2. コンピュータで感情を扱う

感情を人間が判断する際には先にも述べた通り, 様々なことを考慮している. コンピュータで扱いやすい情報は数値化されたものであるため, こういった様々な考慮されるべき要因を数値化し扱うことはコンピュータが人間の感情を理解するために有効であると考えられる. そのため, 人間の感情をカテゴリ化し, それぞれ数値化することによって感情をコンピュータで扱うことが可能になると考えられる.

2.3. 感情の分類

本研究では, 感情表現を網羅されている『感情表現辞典』[5]を利用して人間の感情のカテゴリ化を行った. ここで扱う感情分類は, この著者である中村が感情を{喜, 怒, 哀, 怖, 恥, 好, 厭, 昂, 安, 驚}の 10 種類に分類し収集を行ったものである. 感情表現辞典には日本小説などから収集した 2, 167 語が掲載されている. 本手法では, 中村により分類された 10 種類の感情を利用する.

3. システム概要

3.1. 処理過程

システムの概要を図 1 に示す. 提案手法では, まず, 入力文が重文であるかどうかを判定する. 次に, 二重否定を考慮した文の表層的表現を用いた感情推定を行う. 感情表現を抽出できなかった場合, ウェブ上の因果関係を用いた感情推定を行う. 重文, 二重否定が含まれていた場合は, 感情の重点や曖昧性を付加し, より深い感情解析を行う. 表層的表現では文自体に含まれる文末表現や感情表現により感情を推定する手法[1]を用いる. 統計的手法では, 文との因果関係を持つ文を Web 上から取得し感情認識を行う手法用いシステム構築の際に施らのシステムと Ptaszynski らの ML-Ask システム[1]を利用した. ML-Ask システムに二重否定を考慮するシステムを導入した.

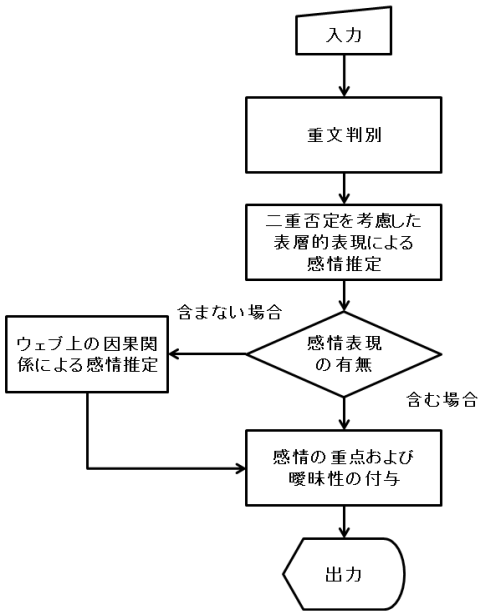

図 1.　システムフロー

3.2.　重文の判定

ここでは感情推定に入る前に入力文に対して前処理を行う．重文とは文中に節と節が並列的に含まれる文のことである．「サッカーは好きだけど，観るのはあんまり好きじゃない」などがこれにあたる．

重文となるためには 2 つパターンがある．「太郎は 6 時に起き，次郎は 8 時に起きる．」のように連用中止法を用いる場合と「太郎は 6 時に起きて，次郎は 8 時に起きた」のように「て」のような接続助詞を伴う場合である．本稿では接続助詞を用いた重文の判定を行う．接続助詞にはいくつかの種類があるが，重文に伴う接続助詞である「けど，けれど，でも，と，でも，が，し，て，のに」の 9 個の接続助詞を対象とした．構文解析にこれらの接続助詞を用いることにより「サッカーは好き」と「観るのはあんまり好きじゃない」のように文を分割しそれぞれ感情認識を行う．

3.3.　表層的表現による感情推定

3.3.1.　感情推定と否定形

表層的表現とは文中に含まれる表現のことである．「大学に受かったあああ！」などのように感情的な表現をする場合に文末には普段とはことなる表現が表出する．この場合「あああ」と「！」が文末に含まれる感情的な表現となる．この表現を感情要素として抽出を行う．また，「ラーメンが好きだ」などのように文自体に「好き」などの直接的な感情が含まれている場合にも，感情表現として抽出を行う．その際，中村の感情分類[4]を用いて分類を行う．つまり，表層的表現では文自体に含まれる感情表現と文末の感情要素を抽出し感情を含んでいるのかどうかを判断し，感情を推定する．また，ここで二重否定以外の否定形の処理を行う．

否定による感情分類は，Russellらによる活発的か非活発的，肯定的か否定的の2次元情報を保持した感情[6]に対し，Ptaszynskiらにより2次元空間にマッピングされた中村の感情分類を用い，感情分類を遷移させ感情抽出を行う[4]．2次元空間にマッピングされた中村辞書を図2に示す．　例えば，「ラーメンは好きじゃない」のように感情表現の後ろに否定形が現れた場合，「好き」という感情表現から2次元空間にマッピングされた反対である厭という感情を抽出することになる．

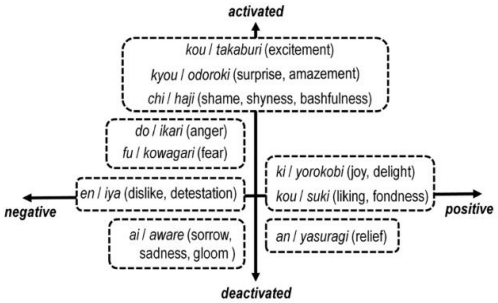

図 2.　Russell 空間における中村感情分類

3.3.2.　二重否定の考慮

表層的表現による感情推定においての二重否定の処理について本項で述べる．二重否定を伴う文は「落ち着かなくもない」のように否定を 2 度重ねることにより強調や曖昧性などの意味が伴う．二重否定の形式は多様であるが，ある程度類型化することができるため[6]，本研究では，第一著者により感情を含む場合に用いる二重否定を限定し類型化したものを対象に処理を行った．曖昧の意味を伴う「～ないとは（とも）かぎらない，ないとは（とも）言えない，～ないとは（とも）～ない，～ないでは（でも）～ない，～なくは（も）ない，～ないという（体言）ではない」と強調などの意味を伴う「～ない（体言）は（も）ない」の 7 種類に分類した．

3.4.　Web 上の因果関係による感情推定

3.4.1.　文の検索クエリの作成

文に含まれるフレーズを抽出し，文末に接続詞を付与する．それらを Web 上で検索する際のクエリとして検索を行う．ここで，フレーズとは文法上の単語の集合のことである．以下のような作成ルールを用いている．図 3 はフレーズの作成例である．

- 3-gram からフレーズの作成
- フレーズの先頭は助詞，接続詞ではない
- フレーズの最後は動詞又は形容詞
- 符号，顔文字は無視する

入力文：ちょっと、辛すぎた。
フレーズ：辛すぎた

図 3.　フレーズ作成例

これにより得られたフレーズに対し文末に接続詞を付与する．連続の役割を担う助詞や単語のうち，施らによって得られた接続詞「て，と，ので，から，たら」の 5 種類を付与した[3]．また，検索ヒット数の向上のためにフレーズ単体もクエリとして利用した．

3.4.2.　Web からの結果節の抽出

入力文から得られたフレーズをクエリとして Web 上から入力文との因果関係が存在する文の抽出を行う．クエリの直後に存在する文をクエリとの因果関係を保持する文とする．例えば，表4の「辛すぎた」という文の

場合，表4のように接続詞が付与され検索により得られたスニペット中の検索クエリである「辛すぎたら」の直後にある「泣くなり落ち込むなりするといいでしょう」を因果関係を保持する文とした．検索エンジンはYahoo を利用した．一つのクエリに対して 100 個のスニペットの取得をテキストデータとして保存した．

入力：辛すぎた。
接続詞の付与：辛すぎたと
取得スニペット：辛かったら泣くなり落ち込むなりするといいでしょう
保存する文：泣くなり落ち込むなりするといいでしょう

表 4 因果関係を保持する文の取得の例

3.4.3. 取得データの分析

取得したテキストデータから，フレーズが存在する文の抽出を行う．フレーズよりも後に存在する 1 文を感情推定に用いるためのデータとした．しかし，フレーズよりも後ろにある場合であっても，以下の場合はデータとして保存しない．

- 後ろの文に「しかし」などの逆接を表す接続詞が現れる場合
- 後ろの文に「ない」などの否定を表す単語が現れる場合

こうして得られたデータに対し形態素解析を行い，解析結果に対し中村の 10 種類の感情分類[4]それぞれをカテゴリとしたデータベースとの比較を行う．各感情のカテゴリの数をそのカテゴリの点数とし，各カテゴリの合計が上位から 60%を超えるものをその文の感情として分類する．複数の感情が含まれている場合も考慮するために上位からカテゴリの合計が 60%を対象とした[3]．ただし，ヒット件数が少ないものは入力文とは関係のない感情を出力している可能性が高いと第一筆者により判断したため 60%以内に含まれていても各感情カテゴリの点数が全体の 10%未満の場合は出力しない．例えば，「喜 50，哀 9%」の場合，喜と感情を出力する．

3.5. 感情の重点および曖昧性の付与

重文否定判定により 2 つの節に分けられた場合，重文の重点を付与する．例えば，「サッカーは好きだけど，観るのはあんまり好きじゃない」の場合，「サッカーが好き」という感情と「観るのはあんまり好きじゃない」の二つに感情分けることができる．元の文自体の感情はどちらの感情に重点が置かれているかの情報を，重点が置かれている感情に付与する．感情の重点の付与については，4.1 に後述するアンケート結果から二つの節に分けた際の後ろに感情を含む場合に付与する．また，入力文が曖昧性を含む二重否定を含んでいた場合はその感情に曖昧性を付与させて結果とし出力する．

4. 評価実験

4.1. 事前アンケート

今回の実験に用いるためのデータの作成を行った．ブログやツイッターなどから，重文になりうる 9 個の接続詞と否定形を含む文を第一筆者により接続詞または否定形をクエリとして 50 文，抽出を行った．その際，顔文字を含む文は対象としていないため，顔文字の含まれていないものを抽出した．文自体の感情のみを推測するため今回は顔文字を対象から除外した．男性 7 名女性 1 名の計 8 名に対してアンケート調査を行った．被験者は理系学生が 8 名である．それぞれの文に対して中村による 10 種類の感情分類[4]のどの感情を含んでいるかを判断してもらった．二つ以上感情を含んでいると判断した場合には二つまで感情を選んでもらい，感情を含んでいないと判断した場合には無感情を選択してもらうようにした．感情が曖昧性を保持しているかについても判断してもらった．また，文を節に分割して考えることができる場合は，それぞれの感情を判定してもらった．その際，前後の節のどちらに重点が置かれているかも判断してもらった．8 人中 2 人以上でかつ上位 2 位までの感情を正解データとして作成した．被験者が 2 つの節に分けられると判断した文の感情の重点は 35 文中 32 文が後ろに重点があると判断した．

4.2. 実験

事前アンケートで収集した 50 文に対し，提案手法と提案手法を含まない文の表層的表現とウェブ上の因果関係をベースラインとし比較を行った．重文による分割の影響と二重否定の考慮による影響を調査するため，ベースラインに二重否定判別のみを導入したシステムと重文による分割のみを用いたシステムとの比較も行った．施らのシステムでは入力文をクエリとしてウェブ上のスニペットから感情推定する．ML-Ask では文自体に含まれる感情と文末表現から感情の推定を行う．システムが推定する感情が入力文感情と一つでも一致した場合に正解とした．実験結果と成功例，失敗例を表 1 と表 2 に示す．重点と曖昧性における再現率と適合率については表 3 に示す．

表 1 提案手法とベースラインの比較

	成功	失敗	精度[%]
ベースライン	24 文	26 文	44.0
二重否定判定のみ	26 文	24 文	52．0
重文判定のみ	30 文	20 文	60．0
提案手法	39 文	11 文	80．0

表 2 提案手法成功例

	入力文感情	ベースライン	提案手法
リハビリを経て復帰を果たしたが、不安がないとは言えない！	好 厭	安	好 厭
海もきれいで最高だったし、ご飯も美味しかった！	安	無感情	喜 怒

表 3 提案手法失敗例

	入力文感情	ベースライン	提案手法
このこなれたオシャレさには夢中にならずに入られないハズ！	昂 喜	厭	好
超パーティー行きたいけど、一人はやっぱり辛い	昂	無感情	無感情

表 4 重点と曖昧性の再現率と適合率

	再現率[%]	適合率[%]
重点	62.9	100.0
曖昧性	75	64. 3

4.3. 実験の考察

評価実験において，二重否定のみを考慮したシステムの精度は 52.0%，重文のみを考慮したシステムの精度は 60.0%とベースラインの 44.0%よりもそれぞれ大きく精度の向上が確認でき，本手法の有効性の確認を行った．また，重文と二重否定の両方考慮した提案手法では 80. 0%とベースラインより 36. 0 ポイントと大幅な改善を確認することができた．重文を考慮することにより，節ごとでウェブ上の因果関係から感情の推測を行うため，ベースラインの入力文自体をクエリとするよりも取得できるスニペット数が異なるため精度が向上したと考えられる．そのような文は 5 文存在した．また，二重否定を考慮することによりベースラインでは取れなかった感情を取得できることを確認した．失敗した文の原因としては，表層的表現に用いた中村辞書に登録されていない単語が現れた場合が 2 文存在し，二重否定を含む文による失敗が 6 文存在した．二重否定を含む場合は感情表現をそのまま出力したため感情の推定に失敗したと考えられる．認識する際に検索クエリによって得られたスニペット中に文とは関係ない感情表現が含まれている場合が 2 文存在した．

感情の重点については再現率 62.9%，適合率 100.0%という結果が得られた． 被験者が 35 文のうち重点がある文だと判断した文のうち，11 文を提案手法で取得することができなかった．複文についての考慮をしないために失敗した文が 2 文存在した．重点が置かれている節の感情を抽出することができなかったために失敗した文が 9 文存在した．感情を含まない場合は重点を付与しなかったためであると考えられる．

二重否定による感情の曖昧性については再現率 75.0%適合率 64. 3%という結果が得られた．被験者が判断した 12 文のうち本システムにより 9 文を抽出し，3 文を抽出することができず，本来取得するべきではない 4 文を抽出していた．本研究では二重否定による曖昧性のみの抽出を行ったが，感情表現を含まない文に対しての曖昧性を考慮していなかったためであると考えられる．二重否定の後に「！」が現れていたことにより，曖昧性の抽出に失敗した文が 2 文存在した．感情の重点，曖昧性の抽出にはそれぞれ課題は残ったが抽出された感情に感情の重点，曖昧性を付与することにより，より深い感情認識を行える可能性があることが確認された．

5. まとめと今後

本稿では，文の表層的情報を用いた感情認識手法とウェブ上の因果関係を用いた感情認識手法のハイブリッド手法に対し，重文による文の分割と二重否定判別による解析を導入する手法の提案を行った．重文による分割と二重否定を導入することで，80. 0%と高い精度で感情認識を行えることができるという結果が得られた．また，重文における感情の重点と二重否定による感情の曖昧性を確認することができ，システムに導入することでより深い感情認識を行うことができることを確認した．今後の課題として，ウェブ上の因果関係を用いた手法に対しさらなる改善が必要であることが確認できた．例えば，スニペットにおける雑音の除去などがあげられる．また，本稿では重文と二重否定のみを対象としたが，複文や程度の副詞など感情認識の課題は様々に残っているので，これらの研究を進める予定である．

参考文献

[1]Michal Ptaszynski， Pawel Dybala， Wenhan Shi， Rafal Rzepka， Kenji Araki: A System for Affect Analysis of Utterances in Japanese Supportedwith Web Mining， 日本知能情報ファジイ学会誌，Vol. 21， No. 2， pp. 194-213， 2009.

[2] 遠藤 大介，斎藤 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”，言語処理学会年次大会発表論文集，pp. 947-950，2006.

[3] 施 文翰，ジェプカ ラファウ，荒木 健治：“ウェブ上の因果関係を用いたユーザ入力文からの感情情報の推測”，情報科学技術フォーラム講演論文集，pp. 267-268，2008.

[4] Michal Ptaszynski， Pawel Dybala，Wenhan Shi， Rafal Rzepka and Kenji Araki: Contextual Affect Analysis: A System for Verification of Emotion Appropriateness Supported with Contextual Valence Shifters， International Journal of Biometrics， Vol. 2， No. 2， pp. 134-154， 2010.

[5] 中村 明，『感情表現辞典』，東京堂出版 1993

[6] 林楽 常：“二重否定表現の一考察：形式と意味の相関性を中心に”，人間文化研究，Vol．3，pp. 27-39，2005.

[7] James A. Russell: A circumplex model of affect， Joural of Personality and Social Psychology，Vol. 39，No. 6， pp. 1161-1178， 1980.

連絡先

〒060-0814
札幌市北区北 14 条西 9 丁目
北海道大学大学院情報科学研究科
言語メディア学研究室
滝澤 満
TEL: 011-706-7389