

情緒推定における状況の対称性を考慮するための パターン辞書の拡張

野口和樹 徳久雅人 村田真樹

鳥取大学大学院 工学研究科 情報エレクトロニクス専攻

{s072036, tokuhisa, murata}@ike.tottori-u.ac.jp

1 はじめに

言語の意味理解の一つとして、言語表現から書き手や登場人物の情緒を推定する技術に期待が寄せられている。この技術はテキストマイニングへの応用に可能性がある。たとえば、ブログや掲示板などに蓄積されたテキストデータから情緒を推定することで、商品や社会事情に対する大衆の気持ちを知るといったことが挙げられる [1],[2],[3]。そのため、文章の用言や格要素などから、情緒を明示する表現や、情緒の生じる原因を表す表現などを検出し、解析する必要がある。

そこで用いられる手法の中でも、「情緒生起の原因に着目した推定手法」について、関連研究を以下に述べる。まず、目良らは深層格フレームと好感度計算式で、入力文から情緒を推定する手法を提案した [4]。問題点は、あらゆる語において好感度を用意する必要があること、および、人物や物事の省略や照応の解析が必須であることである。次に、徳久らは Web から感情生起要因となる文を獲得し、機械学習を用いて感情を推定する手法を提案した [5]。しかし、マイニングへの応用を考えたときの問題点としては、情緒主や情緒対象が直接得られないことが挙げられる。

一方、田中らと吾郷らは、日本語語彙大系 [6] に、「情緒属性」として、「判断条件、情緒原因、情緒名、情緒主、および、情緒対象」を付与し、情緒推定用結合価パターン辞書を作成することで、パターンベースの情緒推定の手法を示した [7],[8]。情緒推定の方法は、もし、入力文と結合価パターンが意味属性制約を充足しながらマッチし、かつ、判断条件が成立するならば、情緒生起の原因が成立するといえるので、対応する情緒属性を出力するというものである。ここで、判断条件が成立すると仮定することで、用言の語義に基づく情緒推定が可能であり、目良らの手法ほど格要素に依存することなく情緒推定が可能である。しかし、判断条件を常に成立すると仮定すると、過剰に情緒が推定される問題が発生した。そこで、滝川らは、辞書を改良して、判断条件においての二者の

関係の方向性である「接近」と「乖離」の関係を扱えるようにした [9]。問題点は、対称な情緒属性をもたないため、状況の対称性が考慮されないことがあり、十分には状況の対称性が考慮されていないことである。

そこで、本稿では、情緒原因ごとに対称な情報属性が必要であるかを調査する。調査の結果、必要性の高い情緒原因を持つ用言（結合価パターン）に対し、対称な情緒属性の追加を行う。そして、この拡張の妥当性を確認するために、判断条件が理想的に運用できる状況での情緒推定の性能を、実験により評価する。

2 情緒属性付き結合価パターン辞書

2.1 辞書の構成

本辞書は、日本語語彙大系の結合価パターン 14,819 件に、11,712 セットの情緒属性を付与することで作成された [7]。各パターンには、0 セット以上の情緒属性が付与されている。

判断条件に記載するものは、命題もしくは命題関数であり、引数に代入される値（主に格要素の情報が代入される）によって真偽が決まる（表 1） [8],[9]。

表 1: 判断条件に使われる従来の命題関数

命題関数	意味
生理・近 (x,y)	x と y が生理に「接近」の関係
生理・離 (x,y)	x と y が生理に「乖離」の関係
心理・近 (x,y)	x と y が心理に「接近」の関係
心理・離 (x,y)	x と y が心理に「乖離」の関係
目標実現・近 (x,y)	x と y が目標実現に「接近」の関係
目標実現・離 (x,y)	x と y が目標実現に「乖離」の関係
対人関係・近 (x,y)	x と y が対人関係に「接近」の関係
対人関係・離 (x,y)	x と y が対人関係に「乖離」の関係
対人関係・上下・近 (x,y)	x にとって y が対人関係において「接近・優劣」の関係
不要	常に真
その他	未定義

情緒原因は、〈信頼〉や〈嘘〉など 89 種類がある。たとえば、情緒原因〈嘘〉が付与されたパターンは 16 件存在し、その内、情緒属性は、26 セット存在する。

情緒名は、《喜び》、《悲しみ》、《好ましい》、《嫌だ》、《期待》、《恐れ》、《怒り》、《驚き》、および、《なし》の 9 種類である。

本辞書のレコード例を図1に示す。レコードには、まず「日本語文型パターン」と「一般名詞意味属性制約」があり、さらに2セットの情緒属性が存在する。

セット1を見ると、情緒原因は〈自由〉である。〈自由〉は、「条件や制約が減少し、実行可能なプランの数が増加した」という特徴のラベル名である。「訪れる」という語義だけでは、この特徴をカバーできないが、 $N1$ にとって $N2$ が目標実現に有利に働く場所であるならば、この特徴がカバーできるようになる。そこで、「目標実現・近($N1,N2$)」によって〈自由〉の特徴を補う。「目標実現・近(x,y)」は、 x と y が目標実現に「接近」の関係であることを表す命題関数である。あえて、粗く命題の意味を与えている。

一方、セット2を見ると、情緒原因は〈制限〉である。〈制限〉は、「条件や制約が増加し、実行可能なプランの数が増加した」という特徴のラベル名である。「訪れる」という語義だけでは、この特徴をカバーできないが、 $N1$ にとって $N2$ が目標実現に不利に働く場所であるならば、この特徴がカバーできるようになる。そこで、「目標実現・離($N1,N2$)」によって〈制限〉の特徴を補う。

「 $N1$ が $N2$ を訪れる」は、「接近」と「乖離」、どちらの判断条件も必要な用言（結合価パターン）である。判断条件「目標実現・近」と「目標実現・離」の対と、それに基づく情緒名《好ましい》と《嫌だ》の対が存在することにより、状況の対称性を考慮している。

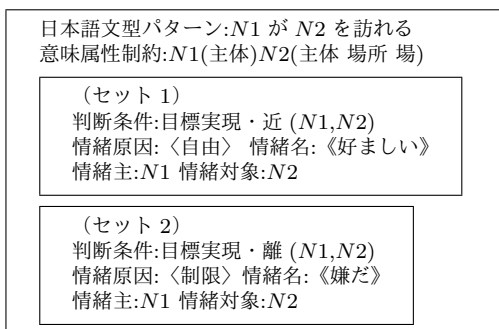


図1: 情緒属性付き結合価パターン辞書のレコード例

2.2 情緒推定方法

本辞書を利用した自動推定の手順を以下に示す。

- (1) 本辞書の全てのパターンと入力文を照合する。
- (2) マッチしたパターンに付与されている「判断条件」 C 、および「情緒属性（情緒原因・情緒名・情緒主・情緒対象）」 E を得る。これで、 $C \rightarrow E$ という命題を得たことになる。なお、 C と E は、元々は命題関数であり、そこに用いられる変数 $N1, N2$ 等には入力文の該当部分が代入されている。
- (3) 判断条件 C の真偽を判定する。真の場合は、情緒名等の E を出力する。偽の場合は、《なし》を出力する（情緒の出力を抑制する）。

なお、判断条件 C の真偽を問わない場合は E をそのまま出力する。動詞からデフォルトとしての情緒が推定されることになる。

3 状況の対称性を考慮しない問題

本辞書には、「接近」と「乖離」、どちらの判断条件も必要な用言（結合価パターン）が存在する。しかし、その内、一方の判断条件しか付与されていない用言（結合価パターン）がある。これらは、もう一方の判断条件を追加し、それに基づく情緒名を追加する必要があると予想される。追加する情緒名の決定は、動詞の種類などを考慮しなければならないため、用言（結合価パターン）ごとに人手で行うべきである。しかし、手作業で行うにはコストがかかる。そこで、機械的に行う方法を提案する。まず、判断条件において、「接近」と「乖離」の対を対称と定義する。次に、それぞれの情緒名に対して、対称な情緒名を定義する。対称な判断条件を追加した場合、元々の情緒名に対して、対称な情緒名を追加する。どちらの追加も機械的に行うことで、コストを削減し、本辞書の網羅性を高める。

たとえば、入力文「子供がピーマンを食べる」を情緒推定する場合を考えてみよう。一般的にピーマンは子供が嫌いな食べ物とされているため、正解情緒は《嫌だ》とする。この文には、図2のパターンがマッチし、情緒属性を得る。この状況設定では、「生理・近(子供, ピーマン)」は偽であり、《なし》を出力する（情緒の出力を抑制）。正解情緒である《嫌だ》と一致しないため、不正解となる。

一方、図3を追加した場合を考えてみよう。追加の情緒属性の「生理・離(子供, ピーマン)」は真であり、《嫌だ》を出力する。正解情緒である《嫌だ》と一致するため、正解となる。こうして対称な情緒属性を追加することで、状況の対称性を考慮した推定を行うことができる。

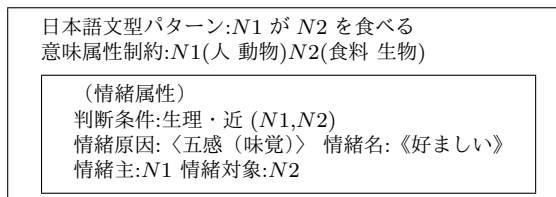


図2: 状況の対称性が考慮されていないレコード例

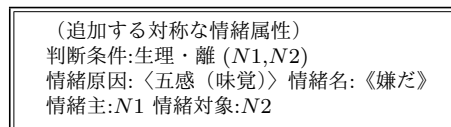


図3: 追加する対称な情緒属性

4 対称な情緒属性の調査・追加

4.1 調査方法

本辞書の11,712セットの情緒属性を、88種類の情緒原因（〈その他〉を除く）ごとに、振り分ける。次に、振

り分けられたそれぞれ的情緒属性に対して、付与されている情緒名と判断条件（「接近」もしくは「乖離」の意味を含む命題関数のみ）の組の件数を調査する。調査の結果、最も件数が多い組を含む情緒属性が付与された用言（結合価パターン）をランダムで10件（10件未満の場合、存在する件数）選出する。選出した用言（結合価パターン）それぞれに対して、対称な情緒属性が必要であるかを調査する。

判断条件は、「接近」と「乖離」の対を対称とする。情緒名は、《喜び》と《悲しみ》の対、《好ましい》と《嫌だ》の対、《期待》と《恐れ》の対を対称とする。これらは双方向に対称性があると仮定する。しかし、《怒り》と《驚き》は、対とはせず、一方向に対称性があると仮定し、それぞれに対して、《期待》が必要であるかを調査する。付与された情緒属性に対して、対称な情緒属性が必要である場合、○をカウントし、必要でない場合、×をカウントする。表2に示す通り、5つのランクに分類する。

表2: 対称な情緒属性の必要度に基づくランクの分類

ランク	10件の場合	10件未満の場合
<i>S</i>	○:10 ×:0 ○:9 ×:1 ○:8 ×:2	○>0 ×=0
<i>A</i>	○:7 ×:3 ○:6 ×:4	○>× ○>0, ×>0
<i>B</i>	○:5 ×:5	○=×
<i>C</i>	○:4 ×:6 ○:3 ×:7	○<× ○>0, ×>0
<i>D</i>	○:2 ×:8 ○:1 ×:9 ○:0 ×:10	×>0 ○=0

4.2 調査結果

調査結果を表3に示す。88種類的情緒原因の内、83種類は、「接近」もしくは「乖離」の意味を含む命題関数が付与された用言（結合価パターン）が存在したが、5種類は、存在しなかった。83種類的情緒原因のうち39%（32種類）は*S*・*A*に分類された。

表3: 調査結果の分類一覧

ランク	総数(種)	割合	該当する情緒原因の例
<i>S</i>	18	22%	〈同意〉 〈教訓〉 〈出会い〉
<i>A</i>	14	17%	〈完遂〉 〈判明〉 〈服従〉
<i>B</i>	8	9%	〈協力〉 〈喪失〉 〈内的な快〉
<i>C</i>	13	16%	〈別れ〉 〈保護〉 〈成行き〉
<i>D</i>	30	36%	〈叱責〉 〈非難〉 〈無計画〉
合計	83	100%	

4.3 追加方法

表3において、*S*・*A*に分類される情緒原因を持つ用言（結合価パターン）に対し、対称な情緒属性の追加を行う。元々付与されている判断条件・情緒名に対して、第4.1節で示した対称な判断条件および、情緒名を追加する。その際、情緒主・情緒対象は既存の情緒属性と同じものを追加する。

4.4 追加結果

対称な情緒属性の追加を行った結果、*S*が1,461セット、*A*が681セット追加された。本辞書全体を見る

と、情緒属性の総数11,712セットに対して、2,142セット、18%(=2,142/11,712)が追加された。日本語の用言約6,118語の内、941語、15%(=941/6,118)に、情緒推定用の情報を追加した。たとえば、*S*に分類される情緒原因〈五感(聴覚)〉を持つ用言「*N1*が*N2*を聞く」には、情緒属性（判断条件「生理・近」・情緒名《好ましい》）が存在する。これに対して、対称な情緒属性（判断条件「生理・離」・情緒名《嫌だ》）が追加された。追加により、*N1*にとって*N2*が生理に不快である状況も考慮できるようになった。

5 情緒推定実験

情緒主と関連事物との関係が明示された上で、1文単位で情緒を推定するという実験を行う。本実験により、辞書の拡張の妥当性を確認する。

5.1 実験方法

実験の入力文の作成者は、本研究室の学生20名である。3種類の事柄（生理・目標実現・対人関係）について、9種類的情緒ごとに作成してもらい、108文作成する。作成する文の条件は、判断条件の命題関数の引数が全て代入されること、もしくは、判断条件が「不要」であること、否定語を含まないこと、述語が動詞であること、時制が現在または過去であること、情緒主は「太郎」、情緒主以外で人名を使用する場合「花子」であることである。作成した108文それぞれに対して、作成者が、情緒主と関連事物との関係を「接近・乖離・どちらでもない」から1つを選択し、その選択に基づき説明文を作成する。

次に、正解データの作成者は、本研究室の学生5名である。各作成者には、1文単位で、入力文、および、情緒主と関連事物との関係の説明文を与える。これらから、作成者が情緒を1種類推定し、正解データとする。また、kappa値を用いて5名の作成者間の推定した情緒の一致率を評価したところ、 $\kappa=0.43$ であった。

一方、計算機の出力する情緒は、次の手順で求める。まず、入力文をパターン辞書を用いて解析する。その結果、判断条件、情緒名、情緒主などが得られる（判断条件不使用時には、ここで得られた情緒名、情緒主などをそのまま出力する）。次に、判断条件の命題関数の真偽を、情緒主と関連事物との関係から決定する。判断条件の真偽に基づき情緒を出力するか否かを定める。入力1文に対して、複数出力を行う場合もある。

正解データと、計算機により出力される情緒とを比較し、一致率で評価する。一致率は、正解データと出力された情緒の一致した割合である（一致件数/全出力数）。

次に、正解データの作成者は計算機の出力する情緒を読み、同意できるかどうかを答える。これは、同意率で評価する。同意率は、被験者が計算機の出力に同意した割合である（同意件数/全出力数）。

5.2 実験の様子

入力文「太郎が花子に出会う」の場合の、正解データの作成者に提示したものを図4に、計算機の入出力を図5に示す。正解データの作成者は、入力文と、説明文「乖離（太郎にとって花子は離れていたいものである）」から推定される情緒を答える（記入欄に書き込む）。入力文と本辞書がマッチした結果、情緒属性セット1（判断条件「対人関係・近（太郎, 花子）」、情緒名《好ましい》、情緒主「太郎」と、セット2（判断条件「対人関係・離（太郎, 花子）」、情緒名《嫌だ》、情緒主「太郎」）を得る。情緒主と関連事物との関係が「乖離」であるため、「対人関係・近（太郎, 花子）」は、偽であり、「対人関係・離（太郎, 花子）」は、真であるので、計算機は、《嫌だ》を出力する。正解データの作成者の推定した情緒と、計算機の出力した《嫌だ》が一致するかを調査する。

次に、正解データの作成者は、計算機の出力した《嫌だ》を読み、同意できるかどうかを答える（yesかnoの一択）。このようにして、一致件数、同意件数を調査する。

- *入力文:太郎が花子に出会う
- *情緒主と関連事物との関係:乖離
(太郎にとって花子は離れていたいものである)
- *正解データの作成者の推定する情緒名:(記入欄)
- *自動推定された情緒名:嫌だ
(正解データの作成者が情緒推定を終えてから表示する)
- *正解データの作成者の同意:yes/no

図4: 正解データの作成者に提示したもの

- (入力)
 - *入力文:太郎が花子に出会う
 - *判断条件(セット1):対人関係・近(太郎, 花子)
 - 判断条件(セット2):対人関係・離(太郎, 花子)
 - (情緒主と関連事物との関係からT/Fを決定)
- (出力)
 - *出力情緒:嫌だ

図5: 計算機の入出力

5.3 実験結果

一致率の実験結果を表4に示す。判断条件の有無、および、対称な情緒属性の追加の仕方に応じて5通りの結果を示した。5人ぶんの正解データがあるので、のべ540文での実験である。#1と#2の比較から、出力数が抑制され、情緒《なし》による一致が増加したことから一致率の向上が確認できる。#2と#3の比較から、対称性の考慮により《なし》以外の情緒名の出力で改善が見られる。#3と#5の比較によると、過剰な対称性の考慮が抑えられたといえる。参考として、#4によると、判断条件の不使用と大きすぎる拡張により、より過剰な推定になることが確認できる。以前の方式#2と今回の方式#5の比較から、改善が確認できた。

また、S・Aに分類される情緒原因を持つ用言（結合価パターン）に対し、対称な情緒属性を追加をした上での判断条件使用時の同意率（5人の合計）は、95%（=608/640）であった。以上から、判断条件が理想的に運用できる状

況では、同意できる情緒推定が可能であること、そして、辞書の拡張の妥当性を確認した。

表4: 一致率の評価

#	推定方法		一致率	(一致件数/全出力数)
	判断条件	対称追加		
1	不使用	なし	42%	(285/675)
2	使用	なし	50%	(320/640)
3	使用	全ランク追加	52%	(338/645)
4	不使用	S・A追加	35%	(303/870)
5	使用	S・A追加	53%	(338/640)

6 おわりに

本稿は、日本語語彙大系における結合価パターンに情緒推定用の情報を付与する課題を進めた。先行研究[9]にて、付与が不十分となっていた対称な情緒属性が必要であるかの調査を、情緒原因ごとに行った。調査の結果、83種類の情緒原因のうち39%は必要性の高い分類であるが、61%は低い分類であった。日本語の用言約6,118語の内、941語(15%)に、情緒推定用の情報を追加した。これらの結果から、全ての用言（結合価パターン）に対して一律に対称な情緒属性を追加するべきではなく、用言（結合価パターン）ごとに動詞の種類などを考慮して対称な情緒属性を追加するべきであることを確認した。

情緒推定実験の結果、理想的な情報の下では、対称な情緒属性を追加した上での判断条件使用時の情緒推定の一致率は、53%となり、以前の方式よりも改善した。また、推定された情緒の95%は正解データの作成者から同意が得られた。この結果から、辞書の拡張の妥当性が確認できた。

謝辞 本研究は、科研費・若手研究(B)(課題番号:22700100)の助成の下で行いました。

参考文献

- [1] 福原知宏, 中川裕志, 西田豊明: “感情表現と用語のクラスタリングを用いた時系列テキスト集合からの話題検出”, 人工知能学会大会, 2E1-02, 2006.
- [2] 福原知宏, 宇津呂武仁, 中川裕志, 武田英明: “複数の言語で記述されたブログ記事を対象とした言語横断型関心解析システム”, 人工知能学会全国大会予稿集, 2F4-3, 2007.
- [3] 徳久雅人: “ブログにおける情緒対象の分析システムの試作”, 電子情報通信学会技術研究報告, 思考と言語, TL-2009-25, pp.1-6, 2009.
- [4] 目良和也, 市村匠, 相沢輝昭, 山下利之: “語の好感度に基づく自然言語発話からの情緒生起手法”, 人工知能学会論文誌, Vol.17, No.3, pp.186-195, 2002.
- [5] 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.
- [6] 池原悟, 宮崎正弘, 白井諭, 横尾昭男, 中岩浩巳, 小倉健太郎, 大山芳史, 林良彦: “日本語語彙大系”, 岩波書店, 1997.
- [7] 田中努, 徳久雅人, 村上仁一, 池原悟: “結合価パターンへの情緒生起情報の付与”, 言語処理学会第10回年次大会発表論文集, pp.345-348, 2004.
- [8] 吾郷裕昭, 徳久雅人, 村上仁一, 池原悟: “情緒生起原因を表す結合価パターン辞書への判断条件の補強”, 電子情報通信学会2008年総合大会講演論文集, 基礎・境界, A-13-1, p.232, 2008.
- [9] 滝川晃司, 徳久雅人, 村上仁一, 池原悟: “情緒推定用パターン辞書における荒いレベルの情緒原因判断条件”, 電子情報通信学会技術研究報告, 言語理解とコミュニケーション, NLC2009-40, pp.43-48, 2009.