

情緒属性付き結合価パターン辞書により 格要素の評価極性を考慮した情緒推定

野口 和樹[†] 滝川 晃司[†] 徳久 雅人[†]

[†] 鳥取大学 大学院工学研究科 情報エレクトロニクス専攻

〒680-8552 鳥取県鳥取市湖山町南 4-101

E-mail: †{s072036,s052032,tokuhisa}@ike.tottori-u.ac.jp

あらまし 本稿は、文単位でのパターンに基づく情緒推定を実現する。情緒推定は、文とパターンの照合により情緒属性から情緒推定の規則を生成し、別途行う判断条件の真偽の推定により、情緒を結論付けるという方式である。本稿では次の3点を行う。(1) 日本語語彙大系における結合価パターンに情緒属性を付与する課題を進めてきたが、そこで、付与が保留となっていた1,600パターンの補修を行う。(2) 理想的な条件下で情緒推定を実験する。(3) 判断条件の真偽推定のための知識ベースの構築を行う。これらの結果、本辞書の95%が補修済みとなった。情緒推定の性能は、被験者による推定との一致率が45%であり、推定された情緒の85%は被験者から同意が得られた。情緒原因文のコーパスから知識ベースを構築したところ、名詞の評価極性と類似する傾向が観測された。以上により、パターンに基づく情緒推定の基本構成が確立できた。

キーワード 情緒, 感情, 推定, パターン, 原因

Emotion reasoning considerate to the polarity of words by affective valency pattern dictionary

Kazuki NOGUCHI[†], Koji TAKIGAWA[†], and Masato TOKUHISA[†]

[†] Department of Information and Electronics, Graduate School of Engineering, Tottori University

4-101, Koyama-Minami, Tottori, 680-8552, Japan

E-mail: †{s072036,s052032,tokuhisa}@ike.tottori-u.ac.jp

Abstract In this paper, we realize a pattern based emotion reasoning from sentences. The method is deductive reasoning. For instance, a conditional statement is made by the pattern matching, a precondition is estimated, and then emotions are deduced from them. We address the following three points: (1) we repair 1,600 invalid patterns remaining on the pattern set developed in our previous work. (2) we make an experiment to infer the emotions under the desirable conditions in order to confirm the basic performance. (3) we try to make the knowledge base for estimating the precondition. As the results, the 95% of the pattern sets became available. The 45% of the estimated emotions exactly corresponded to the correct ones and the 85% of the emotions were acceptable to the subjects of the experiment. We observed the polarity of nouns in the knowledge base. Thus, we realized the fundamental components for pattern based emotion reasoning.

Key words emotion, affect, reasoning, pattern, causality

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、情緒対象を追加し、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、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してしまうと、過剰に情緒が推定されてしまう問題が発生した。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、情緒属性が不明確なパターンが 1,600 件残っていること、および、判断条件を導入した効果が不明であること、自動的な判断条件の推定動作の確認ができていないことである。

本稿では、まずこの情緒属性が不明確なパターン 1,600 件を再分析し、補修を行う。そして、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評価する。さらに、判断条件の推定動作を、コーパスから得た情報で行うことで確認する。

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追加することで作成された。情緒属性には、判断条件、情緒原因、情緒名、情緒主、および、情緒対象が設けられている。情緒名は、《喜び》、《悲しみ》、《好ましい》、《嫌だ》、《驚き》、《期待》、《恐れ》、《怒り》、および、《なし》の 9 種類を使用する。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）によって真偽が決まる。具体的には表 1 の通り設計されている [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」の関係である
生理・離 (x, y)	x と y が生理に「乖離」の関係である
心理・近 (x, y)	x と y が心理に「接近」の関係である
心理・離 (x, y)	x と y が心理に「乖離」の関係である
目標実現・近 (x, y)	x と y が目標実現に「接近」の関係である
目標実現・離 (x, y)	x と y が目標実現に「乖離」の関係である
対人関係・近 (x, y)	x と y が対人関係に「接近」の関係である
対人関係・離 (x, y)	x と y が対人関係に「乖離」の関係である
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」の関係である
その他	未定義
不要	常に真である

本辞書のレコード例を図 1 に示す。レコードには、まず「日

本語の結合価パターン」と「一般名詞意味属性制約」があり、さらに 2 つの情緒属性セットが存在している。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。

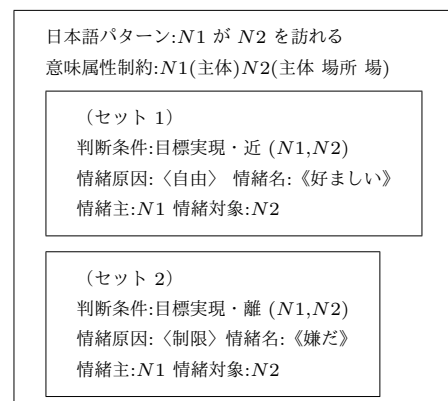


図 1 情緒属性付き結合価パターン辞書の一部

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 、情緒原因・情緒名 E 、情緒主・情緒対象を得る。これで、 $C \rightarrow E$ という命題を得たことになる。なお、 C と E の中の変数 $N1, N2$ 等には入力文の該当部分が代入される。
- (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。

たとえば、「お父さんと子供が遊園地を訪れる」という入力文を情緒推定する場合を考えてみよう。遊園地は、遊ぶところがたくさんあり、自由度が大きいため、子供は嬉しいと思っている。しかし、もし「お父さんが仕事の締め切りに追われている」という状況であれば、「子供と遊んでいる場合ではないのに、遊園地は嫌だなあ」と思うかもしれない。この状況設定では、「目標実現・近 (子供, 遊園地)」は真であり、「目標実現・離 (お父さ

ん、遊園地)」は真である。こうして情緒推定の過程に妥当な説明がつく。

2.3 判断条件の推定方法

情緒推定の過程は、パターンマッチにより「 $C \rightarrow E$ 」を前提の1つとし、判断条件の推定により「 C 」をもう1つの前提とすることで、情緒の推定値 E を結論として得るとなっている。この過程を成立させるためには、判断条件の推定を実現しなければならない。

判断条件の推定には、情緒生起の原因文のコーパスを利用する。コーパスには原因文と対応する情緒の組が存在するので、原因文とのパターンマッチで $C \rightarrow E_{\text{estimated}}$ が得られ、対応する情緒より E_{correct} が得られる。 $E_{\text{est}} = E_{\text{cor}}$ が比較的多いならば、 C は真である確率が高く、 $E_{\text{est}} \neq E_{\text{cor}}$ が比較的多いならば、 C は偽である確率が高いといえる。

たとえば、「主婦が温泉を訪れて喜んでいる」という文が得られたとする。図1と照合し、セット1を用いると、情緒名が一致するので、目標実現・近(主婦,温泉)が真である確率が高まることになる。一方、セット2を用いると、情緒名が不一致なので、目標実現・離(主婦,温泉)が偽である確率が高まることになる。なお、「訪れる」以外の動詞からでも、ある1つの命題関数の真偽の確率が支持されるため、比較的収集しやすいと予想している。こうして得られた命題関数の真偽の確率表を「判断情報知識ベース」と呼ぶことにする。

ただし、文脈や修飾節を考慮して情緒推定を行う場合、文脈や修飾節の情報から判断条件の真偽が得られることがある。その際は、判断条件の推定方法は上記の限りではない。

3. 判断条件の補修

第1章で述べた通り、本稿の1つめの課題は、情緒属性が不明確なパターン1,600件が残っていることである。本章では、「保留」が付与されている1,600件のレコードに対して手作業で補修を行う。

3.1 補修手順

各レコードの判断条件のスロットに、「命題関数」もしくは「作業ラベル」を記入する。

ここで、「作業ラベル」は、補修作業用のものであり、命題関数としての役割はない。「作業ラベル」の種類と意味を表2に示す。

表2 作業ラベルの種類と意味

作業ラベル	意味
パターン不使用	他の結合価パターンがその結合価パターンを包含しているためパターンそのものが不要場合に使用
情緒原因再考 保留(文脈依存)	結合価パターンの情緒原因の見直しが必要な場合に使用 文脈からの情報に依存しすぎるため、結合価パターンの情報のみでは、判断条件の付与が不可能な場合に使用
保留(格要素不足)	情緒推定において必須の格要素が不足しているため判断条件の付与が不可能な場合に使用
保留	その他の理由で判断条件の付与が不可能な場合に使用

補修手順を以下に示す。各手順において、「命題関数」もしくは「作業ラベル」の付与が可能ならば、そのレコードへの作業は終了するが、付与が不可能ならば、次の手順へと進むもの

とする。全手順を終えて、付与が不可能ならば、「保留」を付与する。

手順1 他の結合価パターンがその結合価パターンを包含しているかどうかを調べる。包含する場合、作業ラベル「パターン不使用」を付与する。

手順2 情緒原因の見直しが必要なため判断条件の付与が不可能な場合、作業ラベル「情緒原因再考」を付与する。

手順3 文脈からの情報に依存しているため判断条件の付与が不可能な場合、作業ラベル「保留(文脈依存)」を付与する。

手順4 情緒推定において必須の格要素が不足しているため判断条件の付与が不可能な場合、作業ラベル「保留(格要素不足)」を付与する。

手順5 情緒対象が空欄のものは格要素で表せるものに限定し、追加する。

手順6 従来の命題関数で付与できる場合、その命題関数を付与する。

手順7 新たな命題関数を設計し、付与する。

3.2 補修例

a) 「パターン不使用」への補修例

図2は「パターン不使用」と補修されたレコードの例である。情緒原因〈獲得〉は、「目標実現に必要な物事を努力して手にいれた」という特徴のラベル名である。結合価パターン「 $N1$ が $N2$ を集金する」が結合価パターン「 $N1$ が集金する」を包含していると考え、「パターン不使用」とした。

*結合価パターン: $N1$ が集金する
*意味属性制約: $N1$ (主体)
*判断条件:パターン不使用 *情緒原因:〈獲得〉
*情緒主: $N1$ *情緒対象: *情緒名:《好ましい》

図2 パターン不使用へ補修したレコード例

b) 「情緒原因再考」への補修例

図3は「情緒原因再考」と補修されたレコードの例である。情緒原因〈教訓〉は、「目標実現のために役に立つ情報だった」という特徴のラベル名である。「ぼち」には、「人間が犯した悪事に対して神仏が与えるこらしめ、天罰」という意味がある。判断条件の付与は、本研究室の学生が引き継ぎながら手作業で進めている。そのため、先に作業した者の誤りがありうる。この例では、先の作業者が情緒名を《好ましい》とした理由が分からないため、情緒原因の見直しが必要と考え、「情緒原因再考」とした。

*結合価パターン: $N1$ にぼちが当たる
*意味属性制約: $N1$ (主体)
*判断条件:情緒原因再考 *情緒原因:〈教訓〉
*情緒主: $N1$ *情緒対象: *情緒名:《好ましい》

図3 情緒原因再考へ補修したレコード例

c) 「保留（文脈依存）」への補修例

図4は「保留（文脈依存）」と補修されたレコードの例である。情緒原因〈有効〉は、「目標実現への行動結果は、予想以上だった」という特徴のラベル名である。「再生」には、「廃物を加工して、再び同種のものにつくり出すこと」という意味がある。この語義からは、〈有効〉の特徴である「行動すること」を満たす。しかし、目標実現に向かうことができたかどうかを、動詞の語義はカバーしていない。そこで、再生した結果得られるものを y とすると、目標実現・近 ($N1, y$) という命題関数を判断条件に置くことが考えられる。ここで、「(古紙から) トイレトトペーパーを再生する」と「古紙を (トイレトトペーパーに) 再生する」という例文から分かる通り、「 $N2$ を再生する」というとき $N2$ が再生した結果得られるもの y かどうか不明確であるため、「目標実現 ($N1, N2$)」とすべきとは一概にいえない。したがって、「保留（文脈依存）」とした。

*結合価パターン: $N1$ が $N2$ を再生する
 *意味属性制約: $N1$ (主体) $N2$ (人工物)
 *判断条件:保留(文脈依存) *情緒原因:〈有効〉
 *情緒主: $N1$ *情緒対象: *情緒名:〈喜び〉

図4 保留（文脈依存）へ補修したレコード例

d) 「保留（格要素不足）」への補修例

図5は「保留（格要素不足）」と補修されたレコードの例である。情緒原因〈叱責〉は、「自分の行動に不祥事があり、おとがめがあるかもしれない」という特徴のラベル名である。「買う」の語義は「おとがめがあること」をカバーしている。ここで、対人関係を保ちたい者 y に関しておとがめがあると情緒が生じると考えられるので、「対人関係・近 ($N1, y$)」という命題関数を設けることが考えられる。たとえば、「太郎と花子は仲良しである。しかし、太郎は花子から反感を買った」と「太郎と次郎は犬猿の仲である。太郎は次郎から反感を買った」とを比べて分かる通り、仲良しの者から反感を買う場合のほうが《恐れ》や《悲しみ》が大きいと思われる。しかし、 y に該当する格要素は、このパターンに存在しない。したがって、「保留（格要素不足）」とした。

*結合価パターン: $N1$ が $N2$ を買う
 *意味属性制約: $N1$ (人) $N2$ (対人感情)
 *判断条件:保留(格要素不足) *情緒原因:〈叱責〉
 *情緒主: $N1$ *情緒対象: *情緒名:〈恐れ〉

図5 保留（格要素不足）へ補修したレコード例

e) 「対人関係・上」への補修例

図6は「対人関係・上 (x, y, z)」と補修されたレコードの例である。これは新しい命題関数であり、「 x が y よりも z の観点で上位の人間関係である」という際に真となる。情緒原因〈優越〉は、「自分の能力を周囲の者より高く自己評価した」という特徴のラベル名である。「間隔が空く」の語義からは、「他人との比較により差が生じる」ことは分かるが、どちらが上位であるかは分からない。そこで、判断条件に「対人関係・上 ($N1, N2, \phi$)」を置くことで、 $N1$ が $N2$ よりも上位である際に、 $N1$ が〈優

越〉による《好ましい》の生じることが判断できる。なお、 ϕ は、任意であることを示す。

*結合価パターン: $N1$ が $N2$ と間隔が空く
 *意味属性制約: $N1$ (主体) $N2$ (主体)
 *判断条件:対人関係・上 ($N1, N2, \phi$) *情緒原因:〈優越〉
 *情緒主: $N1$ *情緒対象: $N2$ *情緒名:《好ましい》

図6 対人関係・上へ補修したレコード例

f) 「適度」への補修例

図7は「適度 (x, y)」と補修されたレコードの例である。《好ましい》における情緒原因〈五感（触覚）〉は、「痛みや温度、固さ、手触り、衝撃は体の各部位に馴染む」という特徴のラベル名である。「暖房」には、「建物や部屋の内部を暖めること」という意味がある。この語義からは、〈五感（触覚）〉の特徴である「体の各部位に馴染む」を満たしていない。「体の各部位に馴染む」は、 $N1$ の生理状態に依存する。従来の判断条件では、生理状態のみを表すことはできたが、複数の状態を1つの判断条件で表すことはできないと考え、新たな判断条件「適度 (x, y)」を設計した。「適度 (x, y)」は、 x にとって y が適度であることを関係づける場合に使用する。「適度 ($N1, 暖房$)」と補修することで、「体の各部位に馴染む」を満たす。引数を暖房としたのは、パターン上の字面だからである。

*結合価パターン: $N1$ が $N2$ を暖房する
 *意味属性制約: $N1$ (人) $N2$ (施設 建造物 場)
 *判断条件:適度 ($N1, 暖房$) *情緒原因:〈五感（触覚）〉
 *情緒主: $N1$ *情緒対象: *情緒名:《好ましい》

図7 適度へ補修したレコード例

g) 「可能性・高」への補修例

図8は「可能性・高 (e)」と補修されたレコードの例である。情緒原因〈失敗〉は、「可能性が高い行為にもかかわらず失敗した」という特徴のラベル名である。「誤る」の語義からは、〈失敗〉の特徴である「失敗した」を満たす。しかし、「可能性が高い行為にもかかわらず」を満たしていない。従来の判断条件では、可能性が高い行為を表すことができないと考え、新たな判断条件「可能性・高 (e)」を設計した。「可能性・高 (e)」は、事象 e の発生する可能性が高いという意味がある。「可能性・高 ($N1$ が $N2$ を $N3$ と思う)」と補修することで、〈失敗〉の特徴である「可能性が高い行為にもかかわらず」を満たす。「 $N1$ が $N2$ を $N3$ と思う」という事象の可能性が高いということである。

*結合価パターン: $N1$ が $N2$ を $N3$ と誤る
 意味属性制約: $N1$ (主体) $N2$ () $N3$ (*)
 *判断条件:可能性・高 ($N1$ が $N2$ を $N3$ と思う) *情緒原因:〈失敗〉
 *情緒主: $N1$ *情緒対象: $N2$ *情緒名:《驚き》

図8 可能性・高へ補修したレコード例

3.3 判断条件の補修結果

補修結果を表3に示す。1,600件のうち65%は「命題関数」が付与され、35%は「作業ラベル」が付与された。本辞書全体を見ると、「作業ラベル」が残る判断条件は、レコードの総数11,712件に対して5%以下であった。なお、命題関数（新）とは表4の通りである。

表 3 判断条件の補修結果

判断条件	総数 (件)	割合
命題関数		
(従来)	975	61%
(新)	68	4%
作業ラベル		
(パターン不使用)	8	1%
(情緒原因再考)	23	1%
(保留 (文脈依存))	453	28%
(保留 (格要素不足))	3	0%
(保留)	70	4%
合計	1,600	100%

表 4 判断条件のために新たに追加した命題関数

命題関数	意味
対人関係・上 (x, y, z)	x が y よりも z の観点で上位の人間関係である
対人関係・下 (x, y, z)	x が y よりも z の観点で下位の人間関係である
適度 (x, y)	x にとって y が適度である
過度 (x, y)	x にとって y が過度である
可能性・高 (e)	事象 e の発生する可能性が高い

4. 情緒推定実験

第 1 章で述べた通り、本稿の残りの課題は、判断条件の導入で情緒推定がより正確になることを確認することと、自動的な判断条件の推定動作を確認することである。そこで、2 通りの実験を行う。1 つめは、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評価する。2 つめは、判断条件の推定動作をコーパスから得た情報で行うことで確認する。

4.1 実験 1: 判断条件が与えられた場合の情緒推定

4.1.1 実験目的

情緒主が状況を肯定的にとらえられているか否かが明示された上で、ある 1 文から情緒を推定するという実験を行う。本実験により、本情緒推定の手法の可能性を検討する。

4.1.2 実験方法

複数の被験者による情緒推定と、計算機により出力される情緒とを比較する。被験者には、情緒推定の元となる 1 文、および、情緒主の状況説明文を与える。被験者は、まず、これらから推定される情緒名を答える。次に、計算機の出力する情緒 (後述する) を読み、同意できるかどうかを答える。これらは、一致率と同意率で評価する。一致率は、被験者の推定した情緒と計算機の出力した情緒の一致した割合である (一致件数/提示件数)。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/提示件数)。

一方、計算機の出力する情緒は、次の手順で求める。まず、情緒推定の元となる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。次に、判断条件の命題関数の種類と真偽を状況説明文から定める。これは人手で行う。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。

以上の方法に対して、次のように行った。まず、被験者は、本研究室の学生 6 名である。次に、情緒推定の元となる文は、単文コーパス [10] から選出した。選出の基準は、判断条件の命題関数の引数が全て代入されること、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である。この基準を満たす文をランダムで 30 文得た。最後に、状況説明文は、1 文から複数通りを作成した。その結果 56 件が実験で提示される。

4.1.3 実験の様子

入力文「娘は今年中学校に上がった」の場合の、被験者に提示したものを図 9 に、計算機の入出力を図 10 に示す。被験者には、図 9 の自動推定された情緒名以外のものを提示する。被験者は、入力文「娘は今年中学校に上がった」と、状況説明文「娘は中学校に行きたい」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、判断条件「目標実現・近 (娘, 中学校)」, 情緒名《喜び》, 情緒主「娘」を得る。状況説明文「娘は中学校に行きたい」から、「目標実現・近 (娘, 中学校)」は、真であるので、計算機は、《喜び》を出力する。被験者の推定した情緒と、計算機の出力した《喜び》が一致するかを調査する。次に、被験者は、計算機の出力した《喜び》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:娘は今年中学校に上がった
- *状況説明文:娘は中学校に行きたい
- *被験者の推定する情緒名:(記入欄)
- *自動推定された情緒名:喜び (被験者が情緒推定を終えてから表示する)
- *被験者の同意:yes/no

図 9 被験者に提示したもの

(入力)
*入力文:娘は今年中学校に上がった
*判断条件:目標実現・近 (娘, 中学校)
(状況説明文から手動で T/F を与える)
(出力)
*出力情緒:喜び (記入欄)

図 10 計算機の入出力

4.1.4 実験結果

実験結果を表 5 に示す。6 人の被験者の合計の一致率は 48%、同意率は 85%であった。以上から、判断条件の真偽が正しく得られれば、ある程度、同意できる情緒推定が可能であることを確認した。

表 5 一致と同意の結果

被験者	一致率	(一致件数/提示件数)	同意率	(同意件数/提示件数)
A	54%	(30/56)	86%	(48/56)
B	45%	(25/56)	91%	(51/56)
C	48%	(27/56)	86%	(48/56)
D	59%	(33/56)	88%	(49/56)
E	41%	(23/56)	77%	(43/56)
F	43%	(24/56)	82%	(46/56)
合計	48%	(162/336)	85%	(285/336)

4.2 実験 2: 判断条件の推定

4.2.1 目的と方法

2.3 節で示した方法で、判断情報知識ベースを構築することを試みる。

本知識ベースの構築には、大量の原因文が必要である。そこで、Kawahara らの 5 億文 Web コーパス [11] から、「ので」を含むこと、「ので」の後に感情表現語が存在すること、かつ、否定語を含まないこと、という条件で、原因文と正解の情緒の組を収集する。感情表現語は小林らの評価表現辞書 [12] から人手で抽出した 414 語を用いる。正解の情緒は、Positive, Negative, Surprise の 3 通りとする。本辞書で推定される情緒は 9 分類であるので、《喜び》, 《好ましい》, 《期待》は Positive とみなし、《悲しみ》, 《恐れ》, 《嫌だ》, 《怒り》は Negative とみなして、推定の情緒と正解の情緒の一致を判定する。

4.2.2 構築の結果

まず、原因文の収集の結果、13,175 文を得た。次に、本辞書を原因文に照合したところ、変数に具体値の代入された命題関数を 10,323 件得た。その一部を表 6 に示す。

表 6 判断情報知識ベースの一部

命題関数	引数	T	F
対人関係・近	(ϕ , 友達)	86%	14%
対人関係・近	(ϕ , 子供)	86%	14%
目標実現・近	(ϕ , ゲーム)	81%	19%
対人関係(上下)・近	(ϕ , ϕ , 勉強)	78%	32%
目標実現・近	(ϕ , 火)	50%	50%
対人関係・近	(ϕ , 医者)	50%	50%
対人関係・離	(ϕ , 上司)	50%	50%
目標実現・近	(ϕ , 汗)	26%	74%
目標実現・近	(ϕ , テレビ)	25%	75%
対人関係・離	(ϕ , 母)	20%	80%

表 6 を眺めると、名詞に対する一般的な評価極性と類似していることが分かる。たとえば、「対人関係・近 (ϕ , 友達)=T(86%)」より、一般に「友達」は対人関係において接近の方向が望まれていると読み取れる。通常、「友達」は肯定的な評価の方が高いと予想されることと一致する。

また、「対人関係・離 (ϕ , 母)=F(80%)」より、一般に母は乖離の関係にあることが望ましくないことが読み取れる。これは、「対人関係・近 (ϕ , 母)=T」と読み替えが可能かもしれないが、読み替えの可能性は今後の調査課題である。

4.2.3 考察

判断情報知識ベースを用いて情緒推定を行う際、Positive, Negative, Surprise の 3 通りの情緒推定でのクローズドテストという実験条件だとしても、その性能は 100%にはならないことが、表 6 より読み取れる。たとえば、「友達」を含む文において、14%は誤判定となる。

動詞の語義および格要素の評価極性という 2 つの情報で情緒推定する際に限界のあることが予想される。したがって、今後は、まずは、情緒推定実験を行うことで、限界を具体的に求める。次に、誤り原因の分析に基づく対応はもちろんだが、格要素への修飾表現、動詞への修飾表現(副詞や連用節)などを考慮に入れた情緒推定を目指す。

5. 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] にて付与が保留となっていた 1,600 件を再分析し、補修を行った。未補修の箇所が、辞書全体の 5%以下になった。

一方、本辞書による情緒推定は、パターンと文を照合することで、 $C \rightarrow E$ 、すなわち、判断条件 C のもとで情緒 E が推定されるという仮定が得られ、判断条件 C の真偽の推定により E の推定が結論付けられるという方式である。理想的な情報の下では、情緒推定の一致率は 48%、同意率は 85%となった。

判断条件 C の推定については、コーパスから得た情報で行うことを示した。また、その推定は名詞の評価極性に類似していることを述べた。しかし、動詞の語義および格要素の評価極性のみで情緒推定を行うには、正解率には限界があることが見えてきた。今後は情緒推定に用いる情報を、修飾語句など 1 段階増やすことが課題である。

謝辞 本研究は、科学研究費補助金(若手研究(B):22700100)のもとで行いました。本研究を進める際、河原氏らおよび小林氏らの言語資料を使用させて頂きましたので、深く感謝致します。

文 献

- [1] 福原宏知, 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”, 第 20 回人工知能学会大会, 2E1-02, 2006.
- [2] 福原宏知, 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”, 第 21 回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, 中岩浩己, 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.
- [10] 村上仁一: “日英対訳データベースの作成のための 1 考察”, 言語処理学会第 17 回年次大会発表論文集, D4-5, pp.979-982, 2011.
- [11] Daisuke Kawahara and Sadao Kurohashi: “Case Frame Complication from the Web using High-Performance Computing”, Proceedings of the 5th International Conference on Language Resources and Evaluation, pp.1344-1347, 2006.
- [12] 小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一: “意見抽出のための評価表現の収集”, 自然言語処理, Vol.12, No.2, pp.203-222, 2005.