

概要

言語の意味理解の一つとして、言語表現から登場人物や話者の情緒を推定する技術に期待が寄せられている。なぜなら、この技術はテキストマイニングとして応用できる可能性があるからである。例えば、ブログなどに蓄積されたテキストデータを情緒推定することにより、観光地の情報や、旅行者の気持ちや行動を分析するといったことが挙げられる。

田中らは、情緒推定へのアプローチの一つとして、「情緒生起原因に着目した手法」を提案した。この手法は、結合価パターン辞書を構築することで、情緒生起原因を明記した特徴（情緒原因）を用言の語義から解析し、情緒を推定する手法である [3]。さらに、吾郷らは、不足する情緒原因の特徴を補うために、本辞書に「判断条件」を追加した [4]。それに加えて、滝川らは、判断条件において情緒主と情緒対象の関係の方向性である「接近」と「乖離」の関係に注目し、辞書を改良した [5]。野口らは、判断条件「保留」と付与された、判断条件が不明確なパターン 1,600 件に対し、再分析と補修を行った [8]。本辞書を用いた情緒推定方法は、もし、入力文と結合価パターンがマッチし、意味属性制約を充足し、かつ、判断条件が成立するならば、対応する情緒属性として「情緒主」、「情緒対象」、「情緒名」を出力するというものである。しかし、この手法は、判断条件における格要素同士の関係を判定する際に、格要素に係る修飾語句を読み捨てている。例えば、「私は美味しいご飯を食べる。」という文も「私は不味いご飯を食べる。」という文も、修飾語句である「美味しい」、「不味い」が捨てられてしまい、同じ情緒が推定されてしまうという問題がある。そこで竹本らは、修飾語句の評価極性を「好評極性」、「不評極性」および「極性なし」の3分類で捉え、それらを利用することで、修飾語句の意味に合わない情緒の推定を抑制するという改良を行った [9]。

本研究では、評価極性の算出精度を向上させるため、極性表現を切り替える機能を実装し、夏と冬の状況に応じた評価極性の算出を行う。そして、得られた評価極性を用いて情緒推定を行う。

具体的には、まず、名詞句の種類として「AのB」型名詞句、「形容詞 A 名詞 B」型名詞句、「形容動詞 A 名詞 B」を想定し、修飾語句と被修飾語として $T = \{A, B\}$ を得る。

次に、 T の SO -score を算出するための極性表現対を極性表現対知識ベースから得る。ここでの SO -score は、一般の人がどう思うかの平均的な値として算出する。そして、「判断条件の接近／乖離の関係」と「名詞句の評価極性」を利用して判断条件の真偽判定を行う。判断条件が接近の関係かつ、名詞句が好評極性、または判断条件が乖離の関係かつ名詞句が不評極性ならば T と判定し、判断条件が接近の関係で名詞句が不評極性の時と、判断条件が乖離の関係で名詞句が好評極性の時は F と判定する。 F と判定された場合は、その判断条件が成り立たないものとし、情緒を出力しないことにする。

改良した手法を評価するため、テスト文 293 文に人手で正解情緒タグを付与して正解データを作成する。次に、同テスト文を自動で情緒推定したものと比較し、精度の調査を行う。

調査の結果、テスト文 293 文に対し、従来手法で出力された情緒 336 個のうち一致したものは 133 個となった。一方、提案手法で出力された情緒 311 個のうち一致したものは 139 個となった。情緒の出力数は減り、一致数が増えたため、従来手法に比べて精度が向上した。今後の課題は、本手法をより多くの状況、名詞の種類に適応させること、状況を入力文やその前後文から自動で判別することである。

目次

| | | |
|-------|----------------------------------|----|
| 第1章 | はじめに | 1 |
| 第2章 | 先行研究 | 3 |
| 2.1 | パターン辞書を用いた情緒推定 | 3 |
| 2.1.1 | 日本語語彙大系の結合価パターン | 3 |
| 2.1.2 | 結合価パターンを用いた意味解析 | 4 |
| 2.1.3 | 情緒属性付きパターン辞書 | 4 |
| 2.1.4 | 情緒属性付き結合価パターン辞書を用いた情緒推定方法 | 6 |
| 2.2 | 共起頻度を用いた評価極性の算出 | 7 |
| 2.2.1 | <i>SO-score</i> の算出式 | 7 |
| 2.2.2 | <i>SO-score</i> の算出に用いる好評表現と不評表現 | 7 |
| 第3章 | 提案手法 | 9 |
| 3.1 | 問題設定 | 9 |
| 3.2 | 名詞句の評価極性算出 | 9 |
| 3.2.1 | <i>SO-score</i> 算出式の拡張 | 9 |
| 3.2.2 | 評価極性算出までの流れ | 10 |
| 3.2.3 | 状況 | 11 |
| 3.2.4 | 極性表現対集合の取得 | 11 |
| 3.2.5 | コーパスの部分選択 | 14 |
| 3.3 | 極性表現対集合の知識ベース化 | 14 |
| 3.3.1 | 知識ベースの概要 | 14 |
| 3.3.2 | リソース | 14 |
| 3.3.3 | 構築手順 | 14 |
| 3.3.4 | 知識ベースの構築 | 15 |
| 3.3.5 | 構築結果 | 17 |
| 3.4 | 評価極性を用いた判断条件の判定 | 18 |

| | | |
|------------|----------------------------|-----------|
| 第4章 | 実装 | 24 |
| 4.1 | システムの概要 | 24 |
| 4.1.1 | システムの構成要素 | 24 |
| 4.1.2 | 各要素の概要 | 25 |
| 4.1.3 | システムの流れ | 25 |
| 4.2 | 各手順の説明 | 26 |
| 4.2.1 | 手順1：判断条件と格要素の取得 | 26 |
| 4.2.2 | 手順2：名詞句の認識 | 26 |
| 4.2.3 | 手順3：極性表現集合 P, N の取得 | 27 |
| 4.2.4 | 手順4：算出用コーパスの選択 | 27 |
| 4.2.5 | 手順5： SO -scoreの算出 | 27 |
| 4.2.6 | 手順6：判断条件の真偽判定 | 28 |
| 第5章 | 実験 | 29 |
| 5.1 | 評価極性の精度 | 29 |
| 5.1.1 | 実験の目的 | 29 |
| 5.1.2 | 実験方法 | 29 |
| 5.1.3 | 正解データ作成 | 29 |
| 5.1.4 | 評価方法 | 30 |
| 5.1.5 | 実験結果 | 34 |
| 5.2 | 情緒推定の精度 | 36 |
| 5.2.1 | 実験の目的 | 36 |
| 5.2.2 | 実験方法 | 36 |
| 5.2.3 | 正解データ作成 | 36 |
| 5.2.4 | 評価方法 | 38 |
| 5.2.5 | 実験結果 | 39 |
| 第6章 | 考察 | 45 |
| 6.1 | SO -scoreによる評価極性算出精度について | 45 |
| 6.2 | 複雑な名詞句への対応 | 46 |
| 6.3 | 夏と冬以外の状況への対応 | 47 |
| 第7章 | おわりに | 48 |

目次

| | | |
|-----|-----------------------------|----|
| 2.1 | 結合価パターンの例 | 3 |
| 2.2 | 用言「食べる」のレコード | 4 |
| 2.3 | 《期待》の情緒原因の特徴フレーム | 5 |
| 2.4 | 情緒主と情緒対象に名詞や ϕ が入力される例 | 6 |
| 3.1 | 評価極性を算出するための処理のフロー | 10 |
| 3.2 | 用言「食べる」のレコード | 22 |
| 4.1 | システムの構成図 | 24 |
| 4.2 | patlap 出力例 | 26 |
| 4.3 | patlap 出力の抑制例 | 28 |
| 5.1 | 用言「飲む」の生理・近のレコード | 40 |
| 5.2 | 用言「飲む」の生理・離のレコード | 40 |
| 5.3 | 用言「触る」の生理・近のレコード | 42 |
| 5.4 | 用言「触る」の生理・離のレコード | 42 |
| 5.5 | 用言「着く」の目標実現・近のレコード | 43 |

表 目 次

| | | |
|------|-----------------------------------|----|
| 2.1 | 判断条件の種類と意味 | 6 |
| 3.1 | 名詞句パターン | 12 |
| 3.2 | 名詞句の抽出結果 | 13 |
| 3.3 | 衣服, 食料, 居住施設への使用に適さないかもしれない用言パターン | 18 |
| 3.4 | 極性表現対知識ベース | 19 |
| 3.5 | 判断条件「生理・近」, 「生理・離」の真理値表 | 22 |
| 3.6 | 判断条件「目標実現・近」, 「目標実現・離」の真理値表 | 22 |
| 3.7 | 判断条件「対人関係・近」, 「対人関係・離」の真理値表 | 22 |
| 4.1 | 「極性表現対知識ベース」の冒頭部分 | 27 |
| 5.1 | 「衣服」のテスト名詞句と正解極性(被験者5名) | 31 |
| 5.2 | 「食料」のテスト名詞句と正解極性(被験者5名) | 32 |
| 5.3 | 「居住施設」のテスト名詞句と正解極性(被験者5名) | 33 |
| 5.4 | 実験結果(評価極性) | 34 |
| 5.5 | 入力データ | 34 |
| 5.6 | 評価極性算出結果(提案手法) | 35 |
| 5.7 | 評価極性算出結果(従来手法) | 35 |
| 5.8 | 入力データ | 35 |
| 5.9 | 評価極性算出結果(提案手法) | 36 |
| 5.10 | 評価極性算出結果(従来手法) | 36 |
| 5.11 | 入力文「熱いおでんを食べる」と文の詳細情報 | 37 |
| 5.12 | 「熱いおでんを食べる」に状況を付与する場合 | 37 |
| 5.13 | 「イチゴ味のかき氷を食べる」に状況を付与する場合 | 38 |
| 5.14 | 「熱いおでんを食べる」に正解情緒を付与する場合 | 38 |
| 5.15 | 実験結果(情緒9分類) | 39 |

| | |
|---------------------------------|----|
| 5.16 実験結果 (情緒 5 分類) | 39 |
| 5.17 実験結果 (情緒 3 分類) | 39 |
| 5.18 判断条件 : 生理の真理値表 | 41 |
| 5.19 判断条件 : 目標実現の真理値表 | 44 |

第1章 はじめに

言語の意味理解の一つとして、言語表現から登場人物や話者の情緒を推定する技術に期待が寄せられている。なぜなら、この技術はテキストマイニングとして応用できる可能性があるからである。例えば、ブログなどに蓄積されたテキストデータを情緒推定することにより、観光地の情報や、旅行者の気持ちや行動を分析するといったことが挙げられる。

情緒推定の方法には様々なアプローチがある。目良らは格要素に対して、情緒主が予め持っている好感度を数値として扱い、それらの数値を情緒計算式に代入し、情緒主に生起する情緒の程度を算出した [1]。徳久良子らは Web から情緒の生起する原因となる文を抽出し、「感情生起要因コーパス」に素性と共に蓄積し、SVM による情緒推定を行った [2]。しかし、情緒の成り立ちを踏まえた上で情緒を推定するためには文法構造を解析して情緒推定する必要がある。

そこで本研究室の田中らは、「情緒生起原因に着目した手法」を提案した。この手法は、結合価パターン辞書を構築することで、情緒生起原因を明記した特徴（情緒原因）を用言の語義から解析し、情緒を推定する手法である [3]。さらに、吾郷らは、不足する情緒原因の特徴を補うために、本辞書に「判断条件」を追加した [4]。それに加えて、滝川らは、判断条件において情緒主と情緒対象の関係の方向性である「接近」と「乖離」の関係に注目し、辞書を改良した [5]。野口らは、判断条件「保留」と付与された、判断条件が不明確なパターン 1,600 件に対し、再分析と補修を行った [8]。本辞書を用いた情緒推定方法は、もし、入力文と結合価パターンがマッチし、意味属性制約を充足し、かつ、判断条件が成立するならば、対応する情緒属性として「情緒主」、「情緒対象」、「情緒名」を出力するというものである。しかし、この手法は、判断条件における格要素同士の関係を判定する際に、格要素に係る修飾語句を読み捨てている。例えば、「私は美味しいご飯を食べる。」という文も「私は不味いご飯を食べる。」という文も、修飾語句である「美味しい」、「不味い」が捨てられてしまい、同じ情緒が推定されてしまうという問題がある。竹本らは、修飾語句の評価極性を「好評極性」、「不評極性」および「極性なし」の 3 分類で捉え、それらを利用することで、修飾語句の意味に合わない情緒の推定を抑制

するという改良を行った [9].

そこで本研究では，評価極性の算出精度を向上させるため，状況として気温を設定し，状況に応じて極性表現を切り替えて評価極性を算出する機能と算出用コーパスを可変化する機能を実装する．

第2章では，従来手法の情緒推定方法と評価極性算出方法を説明する．第3章では，従来手法の問題点と，本研究で行う手法の改良方法について説明する．第4章では，本手法の実装方法について説明する．第5章では，評価極性算出精度と情緒推定精度の実験をそれぞれ行う．第6章では，実験結果と，精度の向上方法について考察する．第7章では，まとめを行う．

第2章 先行研究

本章では、情緒推定の基礎および、評価極性の算出方法について説明する。

2.1 パターン辞書を用いた情緒推定

2.1.1 日本語語彙大系の結合価パターン

情緒属性付き結合価パターン辞書の構築には日本語語彙大系 [6] の結合価パターンを利用している。日本語語彙大系は、「意味体系」、「構文体系」によって構成されている。「意味体系」は、日本語の一般名詞、固有名詞、用言の意味的用法を合計約 3,000 の意味属性体系で体系づけている。「構文体系」は、日本語の用言約 6,000 語について表現構造を結合価パターン約 14,000 件にまとめたものである。

図 2.1 に例として「味わう」という用言の結合価パターンを示す。「味わう」には、3 つの結合価パターンがあり、それぞれに「用言意味属性」、「変形情報」、「日本語の文型パターン」、「英語の文型パターン」、および、「一般名詞意味属性制約」が定義されている。「味わう」の (1) であれば、「30 知覚動作 31 感情動作 (動作)」、「動作」、「N1 が N2 を味わう」、「N1 savor N2」、「N1 (4 人) N2 (838 食料 1253 感情)」が対応する。

| |
|--|
| 味わう (あじわう) |
| (1) 30 知覚動作 31 感情動作 (動作) |
| N1 が N2 を味わう N1 savor N2 |
| [N1 (4 人) N2 (838 食料 1253 感情)] |
| (2) 30 知覚動作 31 感情動作 (動作) |
| N1 が N2 を味わう N1 appreciate N2 |
| [N1 (4 人) N2 (1037 芸術 1236 人間活動 2422 抽象的関係)] |
| (3) 31 感情動作 (動作) |
| N1 が N2 を味わう N1 experience N2 |
| [N1 (4 人) N2 (1259 苦しみ 1270 心配 1380 苦心 1262 悲しみ 1239 感覚)] |

図 2.1: 結合価パターンの例

2.1.2 結合価パターンを用いた意味解析

結合価パターンを用いた意味処理では，入力文が結合価パターンにマッチし，一般名詞意味属性制約を充足した場合，結合価パターンに付随している情報が得られる．これにより，入力文の解析が可能となる．

具体的な結合価パターンを用いた意味解析の例を示す．例えば「祖父が寿司を味わう」という文の場合，「 $N1$ が $N2$ を味わう」とマッチする．そして意味属性制約より，図 2.1 の (1) だけが適合する．その結果，この文の「味わう」の語義は知覚動作および感情動作であることが分かる．また，語義は英語の文型パターンより「savor」であることが分かる．また，このように，入力文と結合価パターンが適合することで，入力文の表す言語規範としての意味（語義）を解析することができる．

2.1.3 情緒属性付きパターン辞書

本辞書は，日本語語彙大系の結合価パターンに，情緒属性として，「情緒主」，「情緒対象」，「情緒原因」，「情緒名」，および，「判断条件」を付与して構築されている．入力文と結合価パターンが適合することで，入力文の用言の語義が解析できるのだが，そこに情緒属性を付与することで，さらに入力文の表す状況において情緒生起原因が含まれているかどうか解析できるようになる．なお，入力文の表す言語規範の意味以上に意味を求める処理を意味理解という．情緒属性付き結合価パターン辞書のレコードの 1 つを図 2.2 に示す．

-
- ◆日本語パターン： $N1$ が $N2$ を食べる
 - ◆意味属性制約： $N1$ (4人 535動物) $N2$ (838食料 534生物)
 - ◆判断条件：生理・近($N1$, $N2$) ◆情緒原因：〈五感(味覚)〉
 - ◆情緒主： $N1$ ◆情緒対象： $N2$ ◆情緒名：《好ましい》
-

図 2.2: 用言「食べる」のレコード

情緒の種類 辞書に付与されている情緒は，文献 [7] で定義されている《好ましい》，《喜び》，《期待》，《悲しみ》，《恐れ》，《怒り》，《嫌だ》，《恐れ》および，《なし》の 9 種類である．

また，これらの 9 種類の情緒を 5 分類した場合は， P （好ましい，期待，喜び）， N （悲しみ，恐れ）， A （怒り，嫌だ）， S （驚き）および，なし（なし）に分類され，3 分

類した場合は, *Positive* (好ましい, 喜び, 期待), *Negative* (悲しみ, 嫌だ, 恐れ, 怒り) および, なし (驚き, なし) に分類される.

情緒の原因 文献 [7] には, 第 2.1.3 節で述べたそれぞれの情緒に対し, 情緒原因の特徴が階層的に定義されている. 階層構造の中では, 下位の特徴は上位の特徴を継承しながらより具体的な特徴となっている. 図 2.3 に《期待》の例を示す.

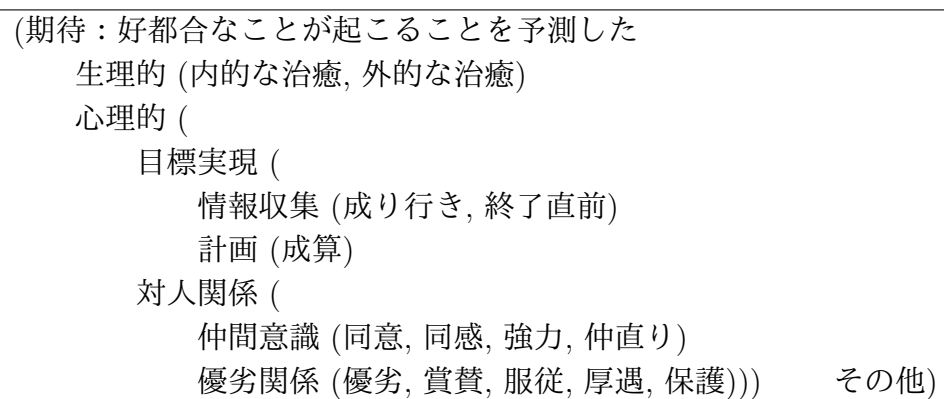


図 2.3: 《期待》の情緒原因の特徴フレーム

例えば, 《期待》の生起する原因となる事態には, 〈好都合なことが起こることを予測した〉という特徴がある. これは最も抽象的な特徴であり, 最上位特徴と呼ばれる. 最上位特徴の一段下は〈生理的〉と〈心理的〉とに分類される. さらに, 〈心理的〉の一段下は〈目標実現〉と〈対人関係〉とに分類される. また, 〈目標実現〉の一段下に 3 つの情緒原因があり, 〈対人関係〉は一段下に 2 つの情緒原因がある. これらの階層構造の中で, 例えば, 〈成算〉の特徴は「ある目標を実現するための企画・行為が成功すると予測した」という意味があるので, 上位特徴と比べて, より具体的であることが分かる.

情緒主と情緒対象 情緒主と情緒対象は, 情緒を持つ者と情緒が向けられる物事である. 主に, 結合価パターンの変数を利用して記述されるが, 具体的な格要素である名詞や, ϕ (don't care) を記入しても良い. 図 2.4 に例を示す.

判断条件 判断条件は, 情緒属性が成立するための前提条件である. 判断条件は, 命題関数であり, 引数に入力される格要素によって真偽を判定する. 表 2.1 は, 文献 [7] が示した情緒原因の階層関係を参考に設計された, 判断条件の種類と意味を示している [4],[5].

◇入力文：鯖味噌を食べる
◆適合パターン：N1がN2を食べる
◆意味属性制約：N1(4人 535動物) N2(838食料 534生物)
◇判断条件：生理・近(ϕ , 鯖味噌)
◇情緒原因：〈五感(味覚)〉 ◇情緒名：《好ましい》
◇情緒主： ϕ ◇情緒対象：鯖味噌

図 2.4: 情緒主と情緒対象に名詞や ϕ が入力される例

下位の判断条件は上位の判断条件を引き継いでいる。最上位の判断条件として「生理」, 「心理」, 「その他」があり, 「心理」の下位には「目標実現」と「対人関係」がある。

表 2.1: 判断条件の種類と意味

| 分類 | 判断条件 (命題関数) | 意味 |
|------|----------------------|----------------------------------|
| 生理 | 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 | 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 | 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 | 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 と 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 |

2.1.4 情緒属性付き結合価パターン辞書を用いた情緒推定方法

情緒属性付き結合価パターン辞書を用いた情緒推定方法について説明する。まず、情緒推定する入力文を情緒属性付き結合価パターン辞書に照合する。その際、入力文に対し、日本語パターンが適合し、かつ意味属性制約が充足するレコードを検索する。この時、判断条件の真偽は引数に入力される値（主に格要素）によって決まる。最後に、判断条件が真ならば、情緒名などの情緒属性を出力し、偽ならば情緒《なし》を出力する。

具体的な例を示す。「私はご飯を食べる。」という入力文の場合、図 2.2 の日本語パターンに適合し、かつ意味属性制約として“私”が(人 動物)を充足し、“ご飯”が(食料 生物)を充足するので、情緒属性 5 項目が得られる。得られた判断条件「生理・近 ($N1, N2$)」は、格要素同士の関係が、生理に「接近」の関係であることを意味する。具体的な格要素を含めた判断条件「生理・近 (私, ご飯)」はこの条件を満たすため、情緒名候補である《好ましい》を出力する。

2.2 共起頻度を用いた評価極性の算出

2.2.1 *SO-score* の算出式

名詞句に評価極性を付与する手法として Turney らの手法がある。Turney らは、次式を用い、コーパスから語句 x の評価極性を算出した [10]。 *SO-score* の算出式を (2.1) 式に、相互情報量 *PMI* の算出式を (2.2) 式に示す。ここで、 $p(a, b)$ はコーパス内において語句 a と語句 b が同一文で共起する確率を、 $p(a)$ は語句 a を含む文がコーパス内に出現する確率をそれぞれ表す。

また、(2.1) 式と (2.2) 式から変形して、コーパス中での出現文数 *hits* で表した式を (2.3) 式に示す。 *hits* に複数の代入がされている場合はそれらが同時に出現した文数を表す。

$$SO-score(x) = PMI(x, \text{“好評表現”}) - PMI(x, \text{“不評表現”}) \quad (2.1)$$

$$PMI(a, b) = \log_2 \frac{p(a, b)}{p(a)p(b)} \quad (2.2)$$

$$SO-score(x) = \log_2 \frac{hits(\text{“不評表現”})hits(x, \text{“好評表現”})}{hits(\text{“好評表現”})hits(x, \text{“不評表現”})} \quad (2.3)$$

SO-score を算出する際に Turney らは“好評表現”に“excellent”，“不評表現”に“poor”を用いている。また、 *hits* の値が 0 の時、 \log が計算できないため、 *hits* には最初から 0.01 を足している。

2.2.2 *SO-score* の算出に用いる好評表現と不評表現

本研究では、 *SO-score* の値が正の場合、 x は「好評極性」であり、負の場合、 x は「不評極性」であると解釈する。 *SO-score* が 0 の場合は「極性なし」とする。コーパスに出

現した共起頻度が4未満の語句については、共起頻度不足として扱い、評価極性を付与しない。日本語のコーパスで x の SO -score を求める場合、(2.1) 式の“好評表現”，“不評表現”には“excellent”，“poor”の代わりに“良い”，“悪い”がよく用いられる。これらを(2.3)式に代入した式を(2.4)式に示す。また、 $hits$ には、Turneyらと同様に最初から0.01を足す。

$$SO\text{-score}(x) = \log_2 \frac{hits(\text{“悪い”})hits(x, \text{“良い”})}{hits(\text{“良い”})hits(x, \text{“悪い”})} \quad (2.4)$$

第3章 提案手法

本章では、パターン辞書を用いた情緒推定手法の問題点と、従来手法の問題点および改良方法について説明する。

3.1 問題設定

パターン辞書を用いた情緒推定では、判断条件に入力する値に名詞変数 N を用いている。例えば、「私はまずいご飯を食べる。」という入力文の場合、名詞変数 N に“ご飯”が適合する。しかし、名詞として“ご飯”のみが判断条件に入力されるため、“まずい”という修飾語句の情報は捨てられ、“まずい”が“ご飯”に与える影響が無視されてしまう。従来手法では、名詞変数 N の代わりに修飾語句を含むことのできる名詞句変数 NP を用い、共起頻度から SO -score を算出することで、名詞句に評価極性を付与していた。しかし、名詞 A, B の SO -score の単純な足し合わせで評価極性を判定したため評価極性算出精度が悪かった。また、状況による評価極性の変化を考慮していない（例えば「冷たいかき氷を食べる」ことは、夏は嬉しいが冬はそれ程でもない）。従って、状況ごとの評価極性が必要となる。

そこで本研究では、名詞句の種類、判断条件、状況を考慮した評価極性を得るために、 SO -score 算出用の極性表現とコーパスを可変化するよう改良する。なお、求める評価極性は、個人依存の評価極性ではなく、一般的な人が好評と不評のどちらを選ぶかを推定するものである。

3.2 名詞句の評価極性算出

3.2.1 SO -score 算出式の拡張

複数の単語に対する評価極性を算出するため、 SO -score 算出式を、式 (3.1) の通りに拡張する。

$$SO\text{-score}(T) = PMI(T, P) - PMI(T, N) \quad (3.1)$$

ここで、変数 T は評価極性を算出される表現の集合を、変数 P は好評表現集合を、変数 N は不評表現集合をそれぞれ表す。例えば、表現集合 T は、「 A の B 」型名詞句や、「形容詞 A 名詞 B 」型名詞句の場合は、 $T = \{A, B\}$ となる。好評表現集合 P 、不評表現集合 N は、名詞の種類、判断条件、状況の3つで極性表現対知識ベースを参照して求める(3.3節)。次に、2つの表現集合 X, Y の間の PMI を式(3.2)の通りにして求める。

$$PMI(X, Y) = \log_2 \frac{p(X, Y)}{p(X)p(Y)} \quad (3.2)$$

$$p(X) = \frac{1}{C} \left| \bigcup_{x \in X} s(x) \right| \quad (3.3)$$

$$p(X, Y) = \frac{1}{C} \left| \bigcup_{x \in X, y \in Y} s(x, y) \right| \quad (3.4)$$

ここで、関数 $s(x)$ は単語 x を含む文の集合を返す関数、関数 $s(x, y)$ は単語 x および y の両方を含む文の集合を返す関数であり、これらはコーパス(3.2.5節)から求める。変数 C は算出に用いるコーパスの総文数である。

3.2.2 評価極性算出までの流れ

評価極性は図3.1の流れで算出する。以下に解説を述べる。

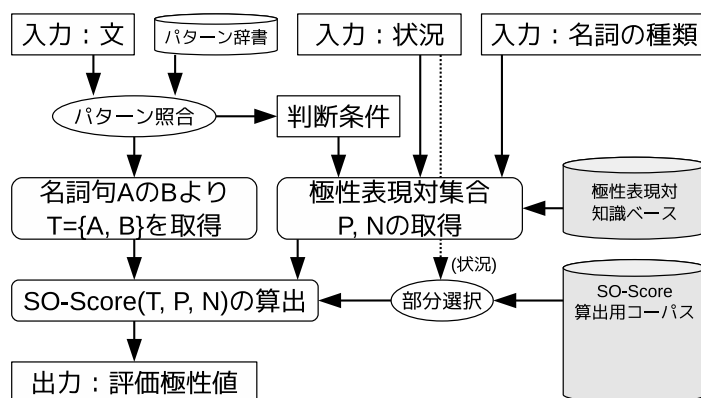


図 3.1: 評価極性を算出するための処理のフロー

まず最初に入力文をパターン辞書に照合することで、文中で使われている用言の語義を解析し、解析後の出力である判断条件を取得する。また、情緒対象としてあてはまった名詞句を取得する。得られた名詞句が「 A の B 」型名詞句、「形容詞 A 名詞 B 」型名詞

句,「形容動詞 A 名詞 B 」型名詞句のどれかにマッチしている場合, $T = \{A, B\}$ として表現集合 T を取得する.

次に, 極性表現対集合 P, N を取得するため, 先程取得した「判断条件」と, それとは別の入力「状況」, 「名詞の種類」の3つを用いて極性表現対知識ベース(3.3節)を参照する.

以上によって得られた T, P, N を SO -score の拡張式に代入する. このとき, SO -score 算出用コーパスは, 状況によって部分選択(3.2.5節)したコーパスを用いる.

扱う名詞句パターン 本手法では, 判断条件に代入する値に名詞句を扱うが, 名詞句には様々なパターンが存在するので, 全ての名詞句について研究を進めるのはコストがかかってしまう. そこで, ブログ文上でどのような名詞句が出現するかを調査する.

まず, 4単語以下で構成される名詞句のパターンを作成し, それらのパターンをブログ文コーパスに照合することで各パターンの出現率を求めた. この照合には2011年1月1日のブログ文コーパス583,827文を利用した. このコーパスは, ameba ブログ, ココログ, goo ブログ, Biglobe ブログより収集したものである. 作成した60個の名詞句パターンを表3.1に, 照合結果を3.2に示す.

名詞句を抽出した結果, このブログ文コーパス上で最も出現していたのは, 「 A の B 」型名詞句, 「形容詞 A 名詞 B 」型名詞句, 「形容動詞 A 名詞 B 」型名詞句であり, 72%以上を占めているものであったため, 本研究ではこの3つの名詞句を主に扱う.

3.2.3 状況

状況には「夏」, 「冬」など気温に関係するものがある. 例えば「いちごのかき氷」を食べる場合, 「夏」ならば冷たくて好評となるが, 「冬」ならば不評となる. このように, 気温の情報も評価極性の算出には必要となる.

本研究では, 気温を表す状況として「夏」, 「冬」, および「一般」の3つを取り扱う. 「一般」は, 「夏」と「冬」で共通している場合に用いるために追加したものである.

3.2.4 極性表現対集合の取得

SO -score を算出する極性表現対には「良い」, 「悪い」のような極性表現対がよく使われる. 本研究では名詞の種類, 判断条件, 状況の3つの情報で極性表現対知識ベース(3.3章で構築)を参照し, 極性表現対を決定する.

表 3.1: 名詞句パターン

| | | | | | | | |
|------|------|------|----|------|------|------|----|
| 形容詞 | 名詞 | | | 動詞 | ことが | 形容詞 | 名詞 |
| 形容詞 | 名詞 | の | 名詞 | 動詞 | ことが | 形容動詞 | 名詞 |
| 形容詞 | 形容動詞 | 名詞 | | 動詞 | のが | 形容詞 | 名詞 |
| 形容詞 | ことが | 形容詞 | 名詞 | 動詞 | のが | 形容動詞 | 名詞 |
| 形容詞 | ことが | 形容動詞 | 名詞 | 名詞 | 動詞 | 名詞 | |
| 形容詞 | ため | 形容詞 | 名詞 | 名詞 | から | 動詞 | 名詞 |
| 形容詞 | ため | 形容動詞 | 名詞 | 名詞 | が | 動詞 | 名詞 |
| 形容詞 | て | 形容詞 | 名詞 | 名詞 | に | 動詞 | 名詞 |
| 形容詞 | て | 形容動詞 | 名詞 | 名詞 | の | 動詞 | 名詞 |
| 形容詞 | のが | 形容詞 | 名詞 | 名詞 | で | 動詞 | 名詞 |
| 形容詞 | のが | 形容動詞 | 名詞 | 名詞 | と | 動詞 | 名詞 |
| 形容動詞 | 名詞 | | | 名詞 | へ | 動詞 | 名詞 |
| 形容動詞 | 名詞 | の | 名詞 | 名詞 | より | 動詞 | 名詞 |
| 形容動詞 | 形容詞 | 名詞 | | 名詞 | を | 動詞 | 名詞 |
| 形容動詞 | ことが | 形容詞 | 名詞 | 形容詞 | から | 動詞 | 名詞 |
| 形容動詞 | ことが | 形容動詞 | 名詞 | 形容詞 | ため | 動詞 | 名詞 |
| 形容動詞 | ため | 形容詞 | 名詞 | 形容詞 | て | 動詞 | 名詞 |
| 形容動詞 | ため | 形容動詞 | 名詞 | 形容詞 | ので | 動詞 | 名詞 |
| 形容動詞 | で | 形容詞 | 名詞 | 形容動詞 | から | 動詞 | 名詞 |
| 形容動詞 | で | 形容動詞 | 名詞 | 形容動詞 | ため | 動詞 | 名詞 |
| 形容動詞 | のが | 形容詞 | 名詞 | 形容動詞 | で | 動詞 | 名詞 |
| 形容動詞 | のが | 形容動詞 | 名詞 | 形容動詞 | ので | 動詞 | 名詞 |
| 名詞 | 形容詞 | 名詞 | | 形容詞 | 副詞 | 形容詞 | 名詞 |
| 名詞 | 形容動詞 | 名詞 | | 形容詞 | 副詞 | 形容動詞 | 名詞 |
| 名詞 | が | 形容詞 | 名詞 | 形容動詞 | 副詞 | 形容詞 | 名詞 |
| 名詞 | が | 形容動詞 | 名詞 | 形容動詞 | 副詞 | 形容動詞 | 名詞 |
| 名詞 | の | 名詞 | | 副詞 | 形容詞 | 名詞 | |
| 名詞 | の | 形容詞 | 名詞 | 副詞 | 形容詞 | 形容動詞 | 名詞 |
| 名詞 | の | 形容動詞 | 名詞 | 副詞 | 形容動詞 | 名詞 | |
| | | | | 副詞 | 形容動詞 | 形容詞 | 名詞 |
| | | | | 副詞 | 動詞 | 名詞 | |

表 3.2: 名詞句の抽出結果

| 抽出数 | 抽出率 | パターン | 抽出数 | 抽出率 | パターン |
|----------------|---------|---------------|-----|--------|------------------|
| 81290 | 52.41 % | 名詞-の-名詞 | 12 | 0.01 % | 動詞-ことが-形容動詞-名詞 |
| 19564 | 12.61 % | 形容詞-名詞 | 12 | 0.01 % | 形容詞-ので-動詞-名詞 |
| 11525 | 7.43 % | 形容動詞-名詞 | 9 | 0.01 % | 副詞-形容動詞-形容詞-名詞 |
| 7468 | 4.82 % | 名詞-を-動詞-名詞 | 8 | 0.01 % | 動詞-ことが-形容詞-名詞 |
| 6908 | 4.45 % | 名詞-に-動詞-名詞 | 6 | 0.00 % | 形容動詞-副詞-形容詞-名詞 |
| 5715 | 3.68 % | 名詞-動詞-名詞 | 4 | 0.00 % | 形容詞-副詞-形容詞-名詞 |
| 3404 | 2.19 % | 名詞-が-動詞-名詞 | 4 | 0.00 % | 副詞-形容詞-形容動詞-名詞 |
| 2313 | 1.49 % | 名詞-と-動詞-名詞 | 4 | 0.00 % | 動詞-のが-形容動詞-名詞 |
| 1870 | 1.21 % | 副詞-動詞-名詞 | 3 | 0.00 % | 形容詞-のが-形容動詞-名詞 |
| 1732 | 1.12 % | 名詞-形容動詞-名詞 | 3 | 0.00 % | 形容詞-のが-形容詞-名詞 |
| 1730 | 1.12 % | 名詞-で-動詞-名詞 | 1 | 0.00 % | 形容詞-ことが-形容動詞-名詞 |
| 1715 | 1.11 % | 副詞-形容詞-名詞 | 1 | 0.00 % | 形容詞-ため-動詞-名詞 |
| 1647 | 1.06 % | 名詞-の-動詞-名詞 | 1 | 0.00 % | 形容詞-ことが-形容詞-名詞 |
| 1526 | 0.98 % | 名詞-形容詞-名詞 | 1 | 0.00 % | 形容動詞-副詞-形容動詞-名詞 |
| 1238 | 0.80 % | 名詞-の-形容詞-名詞 | 1 | 0.00 % | 形容詞-副詞-形容動詞-名詞 |
| 880 | 0.57 % | 形容詞-名詞-の-名詞 | 1 | 0.00 % | 形容動詞-ので-動詞-名詞 |
| 819 | 0.53 % | 名詞-が-形容詞-名詞 | 0 | 0.00 % | 形容動詞-ことが-形容動詞-名詞 |
| 801 | 0.52 % | 名詞-から-動詞-名詞 | 0 | 0.00 % | 形容動詞-ことが-形容詞-名詞 |
| 712 | 0.46 % | 形容動詞-名詞-の-名詞 | 0 | 0.00 % | 形容動詞-ため-形容動詞-名詞 |
| 527 | 0.34 % | 副詞-形容動詞-名詞 | 0 | 0.00 % | 形容動詞-で-形容詞-名詞 |
| 507 | 0.33 % | 名詞-の-形容動詞-名詞 | 0 | 0.00 % | 形容動詞-で-形容動詞-名詞 |
| 364 | 0.23 % | 名詞-が-形容動詞-名詞 | 0 | 0.00 % | 形容動詞-のが-形容詞-名詞 |
| 312 | 0.20 % | 形容動詞-形容詞-名詞 | 0 | 0.00 % | 形容動詞-から-動詞-名詞 |
| 174 | 0.11 % | 形容詞-形容動詞-名詞 | 0 | 0.00 % | 形容動詞-ため-動詞-名詞 |
| 152 | 0.10 % | 名詞-へ-動詞-名詞 | 0 | 0.00 % | 形容詞-ため-形容動詞-名詞 |
| 34 | 0.02 % | 名詞-より-動詞-名詞 | 0 | 0.00 % | 形容詞-ため-形容詞-名詞 |
| 30 | 0.02 % | 形容詞-て-動詞-名詞 | 0 | 0.00 % | 形容動詞-のが-形容動詞-名詞 |
| 28 | 0.02 % | 形容詞-て-形容詞-名詞 | 0 | 0.00 % | 形容詞-から-動詞-名詞 |
| 25 | 0.02 % | 動詞-のが-形容詞-名詞 | 0 | 0.00 % | 形容動詞-で-動詞-名詞 |
| 15 | 0.01 % | 形容詞-て-形容動詞-名詞 | 0 | 0.00 % | 形容動詞-ため-形容詞-名詞 |
| 合計 155,096 名詞句 | | | | | |

3.2.5 コーパスの部分選択

SO-score で共起頻度を得るためのコーパスには2012年のブログコーパス1年分約4千万文を用いる。状況が夏の場合，夏に属するコーパスから共起頻度を求めることで，夏にブログ文中でよく出現する表現を集め易くなるという目的がある。状況を「夏」とすると7, 8月分の約7百万文に, 「冬」とすると12, 1月分の約6百万文にそれぞれ限定できる。

3.3 極性表現対集合の知識ベース化

3.3.1 知識ベースの概要

本知識ベースは，名詞の種類，状況，および，判断条件に応じて好評表現，不評表現を提供する。本研究では，衣食住の名詞, 「夏/冬」の状況，および，判断条件の「生理」, 「目標実現」, および, 「対人関係」を想定して表現集を手作業で作成する。

3.3.2 リソース

結合価パターン辞書から用言を抽出し，極性表現の連想に使用する。また，全ての種類の名詞句について表現集を作成するのは困難のため，本研究では衣食住に属する名詞句のみを対象とする。具体的には，日本語語彙大系の意味コードの「衣服 (NIコード:817) 以下の子孫」, 「食料 (NIコード:838) 以下の子孫」, および, 「居住施設 (NIコード:445) 以下の子孫」に属する名詞を対象とする。判断条件は結合価パターン辞書に記載されているもので，表2.1の分類から「生理」, 「目標実現」, および, 「対人関係」の3つを使用する。

3.3.3 構築手順

以下の手順に沿って知識ベースを構築する。

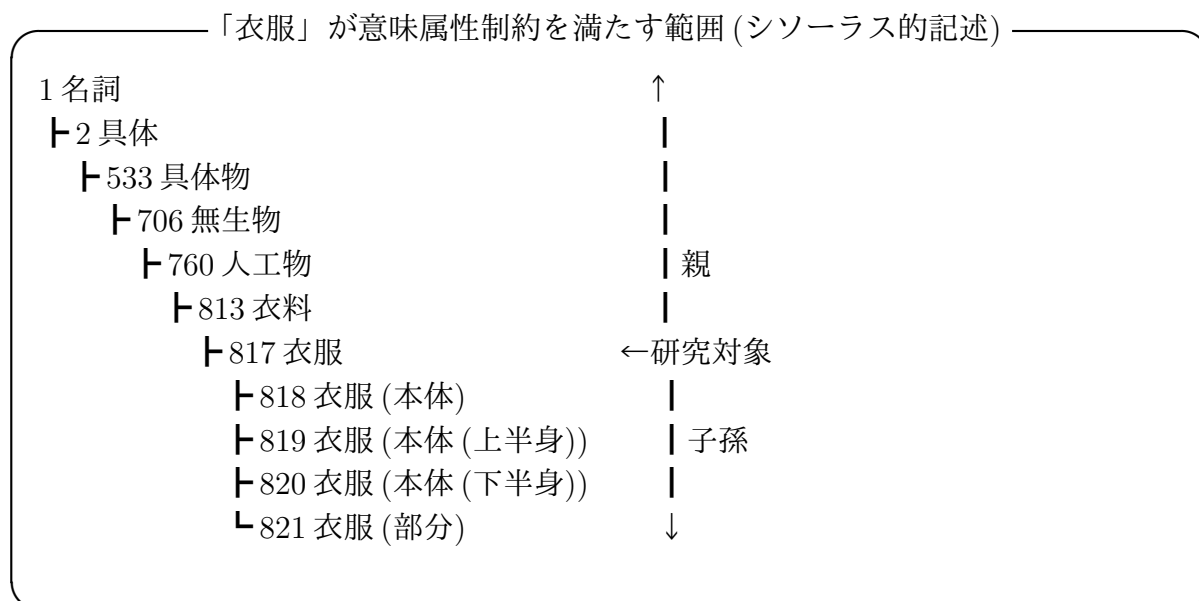
- (1) 名詞の種類，状況，判断条件，用言の組み合わせを収集
- (2) (1)の組み合わせに対して人手で極性表現を連想

3.3.4 知識ベースの構築

(1) 用言の収集 まず、結合価パターン辞書から、衣食住の意味属性を満たすパターンと用言を収集する。収集するパターンは、用言パターンに付与されている意味属性制約の内、一つでも衣食住の意味属性を満たすものとし、残りの意味属性制約には衣食住の意味属性、もしくは「人 (NI コード:4)」が充足するものとする。

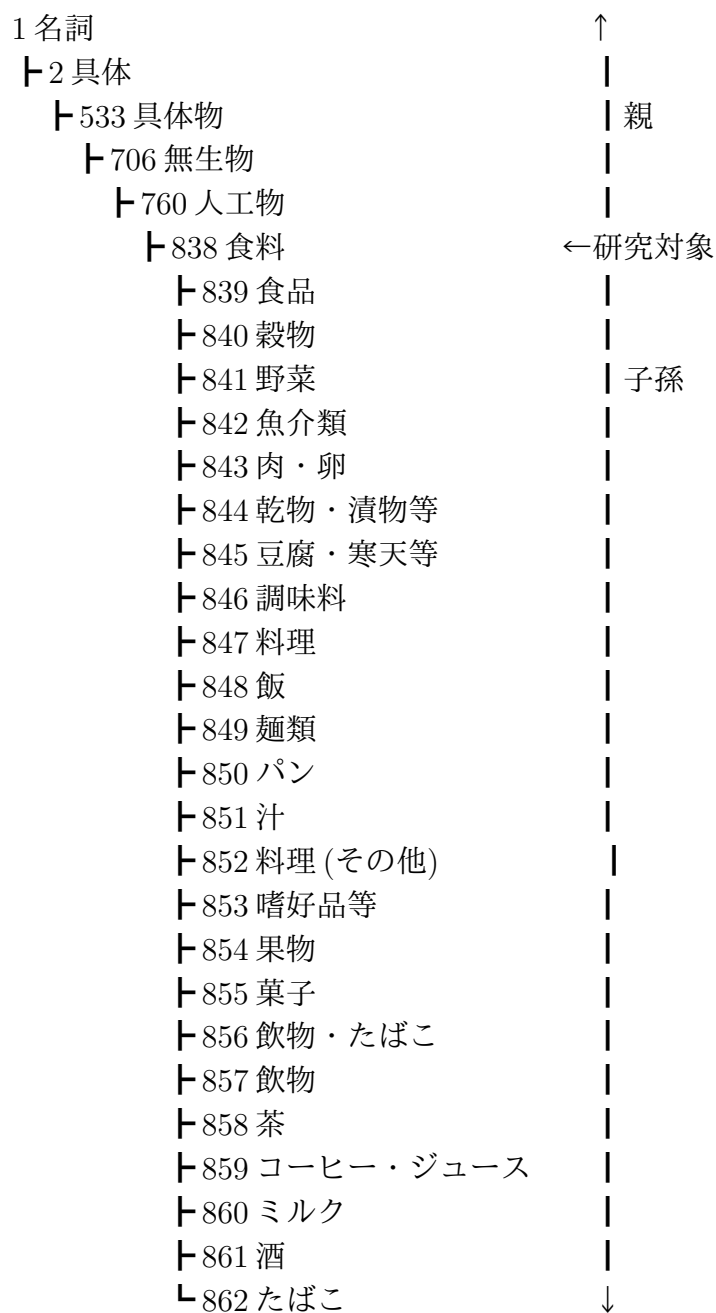
用言パターンを収集した結果、情緒属性付き結合価パターン辞書の所蔵する全 18,810 パターンの内、「衣服」に関して使用される用言パターンを 117 件、「食料」に関して使用される用言パターンを 148 件、「居住施設」に関して使用される用言パターンを 110 件収集した。

「衣服」の意味コードの範囲 日本語語彙大系の意味コードの内、衣服として意味属性にマッチできる範囲は、「衣服 (NI コード:817)」から親と子孫の (1, 2, 533, 706, 760, 813, 817, 818, 819, 820, 821) となる。



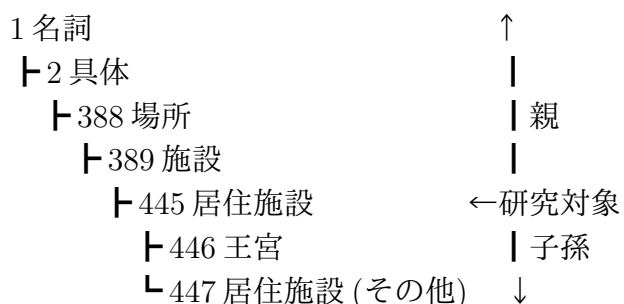
「食料」の意味コードの範囲 日本語語彙大系の意味コードの内、食料として意味属性にマッチできる範囲は、「食料 (NI コード:838)」から親と子孫の (1, 2, 533, 706, 760, 839, 840, 841, 842, 843, 844, 845, 846, 847, 848, 849, 850, 851, 852, 853, 854, 855, 856, 857, 858, 859, 860, 861, 862) となる。

「食料」が意味属性を満たす範囲(シソーラス的記述)



「居住施設」の意味コードの範囲 日本語語彙大系の意味コードの内，居住施設として意味属性にマッチできる範囲は，「居住施設(NIコード:445)」から親と子孫の(1, 2, 389, 388, 445, 446, 447)となる。

「居住施設」が意味属性を満たす範囲(シソーラス的記述)



(2) 極性表現の連想 収集した用言パターンに、状況と名詞の種類を想定しながら直感的に表現集を作成する。次のように条件とともに列挙できる。

- $(P, \text{状況: 夏}) = \{\text{涼しい}\}$
- $(P, \text{名詞の種類: 食料}) = \{\text{美味しい}\}$
- $(P, \text{判断条件: 対人関係}) = \{\text{高級}\}$
- $(N, \text{状況: 夏}) = \{\text{熱い}\}$
- $(N, \text{名詞の種類: 食料}) = \{\text{不味い}\}$
- $(N, \text{判断条件: 対人関係}) = \{\text{粗末}\}$

次に、条件の重なりや一般化を考慮しながら構成する。条件の重なりの例として、「夏」と「食料」を同時に想定するような表現には「あっさり」などがある。また、「美味しい」のように、「夏」と「冬」に共通するものは「一般」という状況カテゴリに記した。

衣食住の名詞句に使用できない用言パターン 抽出された用言パターンは、ほとんどが衣食住を扱えるものであったが、何点か衣食住に適さない用言が見られたので、使用に適さないかもしれない用言パターンを表 3.3 に残しておく。

3.3.5 構築結果

知識ベースを構築した結果、「衣服」について 40 件、「食料」について 40 件、「居住施設」について 28 件、合計で 108 件の極性表現対を作成した。これは、好評表現として

表 3.3: 衣服, 食料, 居住施設への使用に適さないかもしれない用言パターン

| 名詞の種類 | 判断条件 | 動詞 | パターン | 意味属性制約 |
|-------|--------|-------|--------------------|--------------------------------------|
| 衣服 | 目標実現・近 | 摂取する | N1がN2を摂取する | N1(534生物 4人) N2(706無生物) |
| 衣服 | 目標実現・近 | 調合する | N1がN2を調合する | N1(4人) N2(706無生物) |
| 衣服 | 目標実現・近 | 研ぐ | N1がN2をN3で研ぐ | N1(3主体 533具体物) N2(533具体物) N3(533具体物) |
| 衣服 | 生理・近 | 服用する | N1がN2を服用する | N1(4人 535動物) N2(533具体物) |
| 食料 | 対人関係・近 | 親しむ | N1がN2に/と親しむ | N1(4人) N2(3主体 533具体物 1001抽象物) |
| 食料 | 対人関係・近 | 忍ぶ | N1が人目を忍ぶ | N1(2具体 1001抽象物 1235事) |
| 食料 | 目標実現・近 | 洗濯する | N1がN2を洗濯する | N1(4人) N2(760人工物) |
| 食料 | 目標実現・離 | 掴まる | N1がN2に掴まる | N1(4人 535動物) N2(533具体物) |
| 食料 | 目標実現・近 | 研ぐ | N1がN2をN3で研ぐ | N1(3主体 533具体物) N2(533具体物) N3(533具体物) |
| 居住施設 | 目標実現・近 | 蒸発する | N1がN2から/より蒸発する | N1(4人) N2(388場所 2610場) |
| 居住施設 | 目標実現・近 | 吸い上げる | N1がN2をN3から/より吸い上げる | N1(4人 533具体物) N2(2具体) N3(2具体 2610場) |
| 居住施設 | 生理・離 | 跨る | N1がN2に跨る | N1(4人 535動物) N2(2具体) |
| 居住施設 | 対人関係・離 | 跨ぐ | N1がN2を跨ぐ | N1(4人) N2(2具体) |

108件, 不評表現として108件であり, 異なり数としては, 全部で79個の表現が収録されている. 構築結果を表3.4に示す.

3.4 評価極性を用いた判断条件の判定

判断条件の真偽判定について説明する. 本手法では, 判断条件の接近乖離の方向性と, 判断条件の第二引数 NP_y の評価極性から, 判断条件の真偽を判定する.

判断条件が接近の関係で名詞句が好評極性の場合と, 判断条件が乖離の関係で名詞句が不評極性の場合を T とする. また, 判断条件が接近の関係で名詞句が不評極性の場合と, 判断条件が乖離の関係で名詞句が好評極性の場合を F とする. T の時は判断条件が成り立つものとし, 情緒を出力させ, F の時は判断条件が成り立たないものとし, 情緒の出力を抑制する(情緒名《なし》を出力). 表3.5, 表3.6, 表3.7に, 各判断条件の真理値表を示す.

T が成り立つ例として, 「私は美味しいご飯を食べる。」という入力文があったとする. この文は図3.2のレコードにマッチする. この時, 従来手法では具体的な格要素が含まれる判断条件として「生理・近(私, ご飯)」が得られる. しかし, 提案手法では名詞句 NP_y として「美味しいご飯」が用いられるため, 「生理・近(私, 美味しいご飯)」が得られる. 「美味しいご飯」の SO -score の値は正なので, 好評極性が付与される. そして表3.5を参照すると T が成り立つため, 情緒名《好ましい》を出力する.

F が成り立つ例として, 「私は不味いご飯を食べる。」という入力文があったとする. この文でも同様に判断条件「生理・近(私, 不味いご飯)」が得られる. 「不味いご飯」の SO -score の値は負なので, 不評極性が付与される. そして表3.5を参照すると F が成り立つため, 情緒名《好ましい》の出力を抑制する.

表 3.4: 極性表現対知識ベース

| 名詞の種類 | 判断条件 | 状況 | 極性表現対 | 属性 |
|-------|------|----|----------|------|
| 衣服 | 生理 | 一般 | 軽い-重い | 重量 |
| 衣服 | 生理 | 一般 | 無臭-臭い | 臭い |
| 衣服 | 生理 | 一般 | 柔らかい-固い | 質感 |
| 衣服 | 生理 | 夏 | 薄い-厚い | 厚さ |
| 衣服 | 生理 | 夏 | 涼しい-暑い | 温度 |
| 衣服 | 生理 | 冬 | 厚い-薄い | 厚さ |
| 衣服 | 生理 | 冬 | 暖かい-寒い | 温度 |
| 衣服 | 生理 | 冬 | 温かい-寒い | 温度 |
| 衣服 | 目標実現 | 一般 | 格好いい-ダサイ | 見た目 |
| 衣服 | 目標実現 | 一般 | 可愛い-ダサイ | 見た目 |
| 衣服 | 目標実現 | 一般 | きれい-汚い | 清潔さ |
| 衣服 | 目標実現 | 一般 | ぴったり-大きい | サイズ |
| 衣服 | 目標実現 | 一般 | ぴったり-小さい | サイズ |
| 衣服 | 目標実現 | 一般 | 必要-不要 | 需要 |
| 衣服 | 目標実現 | 一般 | 軽い-重い | 重量 |
| 衣服 | 目標実現 | 一般 | 丁寧-適当 | 品質 |
| 衣服 | 目標実現 | 一般 | 多い-少ない | 量 |
| 衣服 | 目標実現 | 一般 | 安い-高い | 値段 |
| 衣服 | 目標実現 | 一般 | 高い-安い | 値段 |
| 衣服 | 目標実現 | 一般 | 高級-安物 | 値段 |
| 衣服 | 目標実現 | 一般 | 新しい-古い | 新しさ |
| 衣服 | 目標実現 | 一般 | 厚い-薄い | 厚さ |
| 衣服 | 目標実現 | 一般 | 薄い-厚い | 厚さ |
| 衣服 | 目標実現 | 一般 | 丈夫-ぼろい | 強度 |
| 衣服 | 目標実現 | 一般 | 有名-無名 | 知名度 |
| 衣服 | 目標実現 | 一般 | 人気-イマイチ | 人気さ |
| 衣服 | 目標実現 | 一般 | ブランド-安物 | ブランド |
| 衣服 | 目標実現 | 一般 | 柔らかい-固い | 質感 |
| 衣服 | 目標実現 | 一般 | 珍しい-ありがち | 珍しさ |
| 衣服 | 目標実現 | 一般 | 良質-粗悪 | 品質 |
| 衣服 | 目標実現 | 夏 | 涼しい-暑い | 温度 |
| 衣服 | 目標実現 | 冬 | 暖かい-寒い | 温度 |
| 衣服 | 目標実現 | 冬 | 温かい-寒い | 温度 |

| 名詞の種類 | 判断条件 | 状況 | 極性表現対 | 属性 |
|-------|------|----|-----------|------|
| 衣服 | 対人関係 | 一般 | 高い-安い | 値段 |
| 衣服 | 対人関係 | 一般 | 丈夫-ぼろい | 強度 |
| 衣服 | 対人関係 | 一般 | 人気-イマイチ | 人気さ |
| 衣服 | 対人関係 | 一般 | 新しい-古い | 新しさ |
| 衣服 | 対人関係 | 一般 | ブランド-安物 | ブランド |
| 衣服 | 対人関係 | 夏 | 涼しい-暑い | 温度 |
| 衣服 | 対人関係 | 冬 | 暖かい-寒い | 温度 |
| 衣服 | 対人関係 | 冬 | 温かい-寒い | 温度 |
| 食料 | 生理 | 一般 | 多い-少ない | 量 |
| 食料 | 生理 | 一般 | 大量-少量 | 量 |
| 食料 | 生理 | 一般 | 美味しい-不味い | 美味しさ |
| 食料 | 生理 | 一般 | 新鮮-陳腐 | 新しさ |
| 食料 | 生理 | 一般 | 新しい-古い | 新しさ |
| 食料 | 生理 | 一般 | きれい-汚い | 清潔さ |
| 食料 | 生理 | 一般 | 濃い-薄い | 濃さ |
| 食料 | 生理 | 一般 | 強い-弱い | 強さ |
| 食料 | 生理 | 一般 | 香しい-臭い | 臭い |
| 食料 | 生理 | 一般 | 芳しい-臭い | 臭い |
| 食料 | 生理 | 一般 | やさしい-しつこい | 味 |
| 食料 | 生理 | 夏 | 冷たい-熱い | 温度 |
| 食料 | 生理 | 冬 | 熱い-冷たい | 温度 |
| 食料 | 生理 | 冬 | 暖かい-冷たい | 温度 |
| 食料 | 生理 | 冬 | 温かい-冷たい | 温度 |
| 食料 | 目標実現 | 一般 | 美味しい-不味い | 美味しさ |
| 食料 | 目標実現 | 一般 | 香しい-臭い | 臭い |
| 食料 | 目標実現 | 一般 | 芳しい-臭い | 臭い |
| 食料 | 目標実現 | 一般 | 新しい-古い | 新しさ |
| 食料 | 目標実現 | 一般 | きれい-汚い | 清潔さ |
| 食料 | 目標実現 | 一般 | 多い-少ない | 量 |
| 食料 | 目標実現 | 一般 | 大量-少量 | 量 |
| 食料 | 目標実現 | 一般 | 濃い-薄い | 濃さ |
| 食料 | 目標実現 | 一般 | 強い-弱い | 強さ |
| 食料 | 目標実現 | 夏 | 冷たい-熱い | 温度 |
| 食料 | 目標実現 | 冬 | 熱い-冷たい | 温度 |
| 食料 | 目標実現 | 冬 | 暖かい-冷たい | 温度 |
| 食料 | 目標実現 | 冬 | 温かい-冷たい | 温度 |

| 名詞の種類 | 判断条件 | 状況 | 極性表現対 | 属性 |
|-------|------|----|----------|------|
| 食料 | 対人関係 | 一般 | 高級-安物 | 値段 |
| 食料 | 対人関係 | 一般 | 高い-安い | 値段 |
| 食料 | 対人関係 | 一般 | 美味しい-不味い | 美味しさ |
| 食料 | 対人関係 | 一般 | ブランド-安物 | ブランド |
| 食料 | 対人関係 | 一般 | 新しい-古い | 新しさ |
| 食料 | 対人関係 | 一般 | 人気-普通 | 人気さ |
| 食料 | 対人関係 | 一般 | 有名-無名 | 知名度 |
| 食料 | 対人関係 | 一般 | 珍しい-ありがち | 珍しさ |
| 食料 | 対人関係 | 夏 | 冷たい-熱い | 温度 |
| 食料 | 対人関係 | 冬 | 熱い-冷たい | 温度 |
| 食料 | 対人関係 | 冬 | 暖かい-冷たい | 温度 |
| 食料 | 対人関係 | 冬 | 温かい-冷たい | 温度 |
| 居住施設 | 生理 | 一般 | 安心-不安 | 安堵 |
| 居住施設 | 生理 | 一般 | 広い-狭い | 広さ |
| 居住施設 | 生理 | 一般 | 広い-窮屈 | 広さ |
| 居住施設 | 生理 | 一般 | 大きい-小さい | 大きさ |
| 居住施設 | 生理 | 一般 | 安全-危険 | 安全度 |
| 居住施設 | 生理 | 一般 | 頑丈-ぼろい | 強度 |
| 居住施設 | 生理 | 夏 | 涼しい-暑い | 温度 |
| 居住施設 | 生理 | 冬 | 暖かい-寒い | 温度 |
| 居住施設 | 生理 | 冬 | 温かい-寒い | 温度 |
| 居住施設 | 目標実現 | 一般 | 手頃-高い | 値段 |
| 居住施設 | 目標実現 | 一般 | 安い-高い | 値段 |
| 居住施設 | 目標実現 | 一般 | 快適-不快 | 快適さ |
| 居住施設 | 目標実現 | 一般 | 広い-狭い | 広さ |
| 居住施設 | 目標実現 | 一般 | 広い-窮屈 | 広さ |
| 居住施設 | 目標実現 | 一般 | 開放-閉鎖 | 開放感 |
| 居住施設 | 目標実現 | 一般 | 近い-遠い | 近さ |
| 居住施設 | 目標実現 | 夏 | 涼しい-暑い | 温度 |
| 居住施設 | 目標実現 | 冬 | 暖かい-寒い | 温度 |
| 居住施設 | 目標実現 | 冬 | 温かい-寒い | 温度 |
| 居住施設 | 対人関係 | 一般 | 高級-普通 | 高級感 |
| 居住施設 | 対人関係 | 一般 | 頑丈-ボロい | 強度 |
| 居住施設 | 対人関係 | 一般 | 開放-閉鎖 | 開放感 |
| 居住施設 | 対人関係 | 一般 | 広い-狭い | 広さ |
| 居住施設 | 対人関係 | 一般 | 広い-窮屈 | 広さ |
| 居住施設 | 対人関係 | 夏 | 涼しい-暑い | 温度 |
| 居住施設 | 対人関係 | 冬 | 暖かい-寒い | 温度 |
| 居住施設 | 対人関係 | 冬 | 温かい-寒い | 温度 |

表 3.5: 判断条件「生理・近」,「生理・離」の真理値表

| NP_y | 生理・近 (NP_x, NP_y) | 生理・離 (NP_x, NP_y) |
|--------|-----------------------|-----------------------|
| 好評 | T | F |
| 不評 | F | T |

表 3.6: 判断条件「目標実現・近」,「目標実現・離」の真理値表

| NP_y | 目標実現・近 (NP_x, NP_y) | 目標実現・離 (NP_x, NP_y) |
|--------|-------------------------|-------------------------|
| 好評 | T | F |
| 不評 | F | T |

表 3.7: 判断条件「対人関係・近」,「対人関係・離」の真理値表

| NP_y | 対人関係・近 (NP_x, NP_y) | 対人関係・離 (NP_x, NP_y) | 対人関係(上下)・近 (NP_x, NP_y) |
|--------|-------------------------|-------------------------|-----------------------------|
| 好評 | T | F | T |
| 不評 | F | T | F |

-
- ◆日本語パターン：N1がN2を食べる
 - ◆意味属性制約：N1(4人 535動物) N2(838食料 534生物)
 - ◆判断条件：生理・近 (N1, N2) ◆情緒原因：〈五感(味覚)〉
 - ◆情緒主：N1 ◆情緒対象：N2 ◆情緒名：《好ましい》
-

図 3.2: 用言「食べる」のレコード

これにより，“美味しいご飯”を食べることは《好ましい》と出力され，“不味いご飯”を食べることは《好ましい》と出力されない。

第4章 実装

本章では、提案手法の実装方法について説明する。

4.1 システムの概要

4.1.1 システムの構成要素

本システムの構成を図 4.1 に示す。本システムは、「情緒属性付き結合価パターン辞書」、「極性表現対知識ベース」、「SO-score 算出用コーパス」、「パターン照合部」、「名詞句認識部」、「極性表現選択部」、「コーパス部分選択部」、「評価極性算出部」、および、「判断条件判定部」から構成されている。

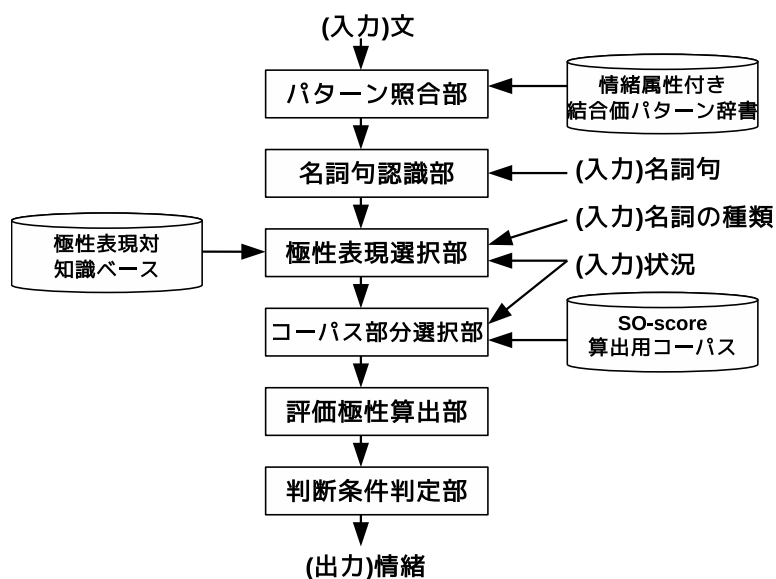


図 4.1: システムの構成図

4.1.2 各要素の概要

「情緒属性付き結合価パターン辞書」は、2.1.3節で説明したように、日本語語彙大系の結合価パターンに情緒属性を付与して作成された辞書を表す。「*SO-score* 算出用コーパス」は、3.2.5節で示したコーパスを表す。「パターン照合部」は、辞書を用いて入力文をパターン照合し、用言の意味を理解した上で具体的な格要素を含んだ情緒属性のレコードを出力する部分である。「名詞句認識部」は、レコードから格要素を取得し、その格要素から名詞句を取得する部分である。「評価極性算出部」は、得られた名詞句に対し、評価極性を算出する部分である。「極性表現選択部」は、入力である名詞の種類と状況、レコードから得た判断条件から極性表現を選択する部分である。「コーパス部分選択部」は、入力である状況から、算出用コーパスを切り替える部分である。「判断条件判定部」は、得られた評価極性と判断条件の接近乖離の方向性から真偽判定を行う部分である。

また、本システムは次の環境の上に実装する。OSにはVineLinux5.2を、プログラミング言語にはRubyを用い、patlap（本研究室で作成されたパターン言語処理ツール）を用いる。本研究は以上のものが使用できる環境で行う。

4.1.3 システムの流れ

本システムの大まかな流れを以下に示す。

- (手順1) パターン照合部：入力文と情緒属性付き結合価パターン辞書を照合すると、いくつかの出力が得られる。ここで、patlapにより出力候補が絞られる。その後、出力候補の情緒属性から「判断条件」と「具体的な格要素」を取得する。
- (手順2) 名詞句格納部：入力から得た名詞句を T として格納する。
- (手順3) 入力から得た状況と名詞の種類、手順1で得た判断条件の3つを用いて極性表現対知識ベースを参照し、好評表現集合 P と不評表現集合 N を求める。
- (手順4) 入力から得た状況で、予め用意しておいたコーパスを選択する。
- (手順5) 評価極性算出部：手順2、手順3で得た T 、 P 、 N と手順4で選ばれたコーパスを用いて *SO-score* の算出を行い、評価極性を算出する。
- (手順6) 判断条件判定部：手順1で得た判断条件と、手順5で得た評価極性を用いて判断条件の真偽判定を行い、偽の場合は出力候補の情緒属性の情緒名に変更を加える。

4.2 各手順の説明

4.2.1 手順 1：判断条件と格要素の取得

まず，入力文と情緒属性付き結合価パターン辞書を照合する．すると，レコードの出力候補がいくつか得られる．patlap はその複数の出力候補を順位付けすることができる．その結果，最上位の出力として図 4.2 のような情報を得ることができる．

```
INPUT=I000001
私はコンビニの弁当を食べる.
[6, "食べる", "原因", "emotion:好ましい", "feeler:私", "feelto:弁当",
"cause:五感(味覚)", "precond:生理・近(私, 弁当,nil)"]
```

図 4.2: patlap 出力例

ここで，一行目は文の ID 番号，二行目は入力文，三行目は左から順に「述語の位置」，「述語の字面」，「情緒過程」，「情緒名」，「情緒主」，「情緒対象」，「情緒原因名」，および，「判断条件」となっている．これらの情報のうち，「判断条件」と「情緒対象」を取得する．この例文でいえば，「生理・近」と「弁当」を取得する．

4.2.2 手順 2：名詞句の認識

入力から得た名詞句を，手順 5 に渡せる形に格納する．本手法は「A 名詞の B 名詞」型，「A 形容詞 B 名詞」型，「A 形容動詞 B 名詞」型の名詞句を想定しているので，名詞句の入力データは以下のように記述し，○○の部分で T として格納する．

「A 名詞の B 名詞」型 … 「○○-の-○○」

「A 形容詞 B 名詞」型 … 「○○-○○」(○○半角ハイフン2つ○○)

「A 形容動詞 B 名詞」型 … 「○○-○○」(○○半角ハイフン2つ○○)

例として「コンビニの弁当」でいえば，「コンビニ-の-弁当」となり，A に「コンビニ」，B に「弁当」が得られる．同様に，「美味しいかき氷」でいえば，「美味しい-かき氷」となり，A に「美味しい」，B に「かき氷」が得られる．

4.2.3 手順3：極性表現集合 P , N の取得

入力から得た名詞の種類と状況，手順1から得た判断条件を用いて「極性表現対知識ベース」を参照し，好評表現集合 P と不評表現集合 N を取得する。

「極性表現対知識ベース」は「名詞の種類」，「判断条件」，「状況」，「極性表現対」，および「属性」の5項目から成るタブ区切りテキストとなっており，名詞の種類，判断条件，状況で参照し，条件に当てはまる極性表現対で集合を作成することで，好評表現集合 P と不評表現集合 N を作成できる。極性表現対は「軽い-重い」のように表記されているがこれは半角ハイフンの左側が好評表現，右側が不評表現となっている。「極性表現対知識ベース」の冒頭部分を表4.1に載せる。

表 4.1: 「極性表現対知識ベース」の冒頭部分

| #名詞の種類 | 判断条件 | 状況 | 極性表現対 | 属性 |
|--------|------|----|----------|-----|
| 衣服 | 生理 | 一般 | 軽い-重い | 重量 |
| 衣服 | 生理 | 一般 | 無臭-臭い | 臭い |
| 衣服 | 生理 | 一般 | 柔らかい-固い | 質感 |
| 衣服 | 生理 | 夏 | 薄い-厚い | 厚さ |
| 衣服 | 生理 | 夏 | 涼しい-暑い | 温度 |
| 衣服 | 生理 | 冬 | 厚い-薄い | 厚さ |
| 衣服 | 生理 | 冬 | 暖かい-寒い | 温度 |
| 衣服 | 生理 | 冬 | 温かい-寒い | 温度 |
| 衣服 | 目標実現 | 一般 | 格好いい-ダサイ | 見た目 |
| 衣服 | 目標実現 | 一般 | 可愛い-ダサイ | 見た目 |
| ... | | | | |

4.2.4 手順4：算出用コーパスの選択

入力から得た状況を用いて，3.2.5節で述べたコーパスの中から適したものを選択する。コーパスを選択とあるが，実際にはそれぞれのコーパスで予め極性表現毎の出現数のデータベースを作っておき，手順5ではそのデータベースを参照して出現文数を得れば良い。

4.2.5 手順5： SO -scoreの算出

手順2で得た名詞句の評価極性を算出する。 SO -scoreの算出の際に必要なのは，以下の4項目である。

- (1) コーパス内での好評表現集合 P のどれかの出現文数合計
- (2) コーパス内での不評表現集合 N のどれかの出現文数合計
- (3) コーパス内で表現集合 T のどれかと好評表現集合 P のどれかが同時に出現する文数合計
- (4) コーパス内で表現集合 T のどれかと不評表現集合 N のどれかが同時に出現する文数合計

これらに 0.01 を足し, (2.4) 式に代入することで SO -score の算出を行う. 評価極性が得られた場合はそれを採用し, 手順 6 に移る. もし, 共起頻度不足で評価極性が算出出来なかった場合は, その名詞句には評価極性を付与せず, 「無効」と付与する.

4.2.6 手順 6 : 判断条件の真偽判定

判断条件の真偽判定に必要な項目は以下の 2 つである.

- (1) 判断条件の方向性が「接近」であるか「乖離」であるか
- (2) 名詞句の評価極性が「好評極性」であるか「不評極性」であるか

手順 1 で取得した判断条件の種類と, 手順 5 で得た名詞句の評価極性を用いて真偽判定を行う. 判断条件が「生理」の場合は表 3.5 をもとに真偽判定を行う. 判断条件「目標実現」と「対人関係」についても同様に実装を行う. 真偽判定の結果が F だった場合にのみ, 抑制を行う. 具体的には, 出力候補のレコード中の「情緒名」の項目を書き換えることで抑制されたことを確認できるように設計する. 本研究では真偽判定が F の場合に「《情緒》」を「《情緒》→抑制」と書き換えるよう設計した. 抑制結果を図 4.3 に示す.

名詞の評価極性が「極性なし」の場合と, 評価極性が付与されていない場合は, 抑制を行わない (真偽判定に用いることができないため).

```

INPUT=I000002
私は貧相な弁当を食べる.
[6, "食べる", "原因", "emotion:好ましい→抑制", "feeler:私", "feelto:弁当",
"cause:五感(味覚)", "precond:生理・近(私, 貧相な弁当,nil)"]

```

図 4.3: patlap 出力の抑制例

第5章 実験

5章では、2通りの実験を行う。1つ目は、提案手法である拡張した *SO-score* 算出式を用いた評価極性の判定精度を求める実験、2つ目は、その評価極性を用いて情緒推定した際の推定精度を求める実験である。

5.1 評価極性の精度

提案手法により、名詞句の評価極性が正しく推定できる割合を求める。

5.1.1 実験の目的

本実験で評価極性の判定精度を求めることにより、以下の2項目の有効性を示すことを目的とする。

- 名詞の種類、状況、判断条件に応じて極性表現を選択する機能の有効性
- コーパスの部分選択機能の有効性

5.1.2 実験方法

実験用のテスト名詞句に予め正解極性を付与しておき、本手法で求めた評価極性と比較することにより、評価極性の判定精度を求める。

5.1.3 正解データ作成

テスト名詞句の収集 まず、テスト名詞句となる名詞句の収集を行う。収集には、2012年の1年分のブログ文コーパスをリソースとして使い、「A名詞のB名詞」型、「A形容詞B名詞」型、および、「A形容動詞B名詞」型の名詞句パターンをマッチングさせ名詞句を抽出する。抽出した名詞句は、B名詞の意味属性(日本語語彙大系の意味コード)を用

いて「衣類」,「食料」, および,「居住施設」に分類する。その中から, 固有名詞をもつ名詞句は極力省き, 一般的に広く知られていそうな名詞句から, 衣食住それぞれについて人手で 50 個を選定した。

正解極性の付与 被験者 5 名に, 先程収集したテスト名詞句に対して, 好評か不評かの判定を依頼した。各名詞句には, 状況として「一般」,「夏」, および,「冬」の 3 つの状況が与えられており, それぞれに対して好評, 不評の判定を行う。つまり, 一つの名詞句に対して 3 回の判定を行うものである。判定作業の際には, 以下の点について注意を呼びかけた。

- 一つの名詞句に対して, 好評か不評かのどちらか 1 つを必ず判定する (二値分類)。
- 「夏」でも「冬」でも評価極性に变化無しと考えられた場合は, 「一般」のものと同じ正解極性を記述する。
- 被験者個人の視点から判定するのではなく, 一般的な人がどちらの評価極性と判定するかを想像して判定する。

判定結果の集計をとり, 満場一致で好評, 不評であるものを「一般的な正解極性」として付与した。

正解極性の付与結果 衣食住の各付与結果を表 5.1, 5.2, 5.3 に示す。各欄は「好評と判定した人数/不評と判定した人数」をそれぞれ表している。例えば, 衣服名詞句の「赤いアロハ」では状況が一般の時, 好評が 3 件, 不評が 2 件であったため, $3/2$ と表記している。この場合は意見が別れているので, 正解極性は付与できない。

好評, 不評のどちらかに満場一致で判定された場合には, その後ろに「(好評)」, 「(不評)」と付与している。評価極性の判定精度を求める実験では, 満場一致で判定されたもののみを用いる。状況毎に評価極性が別々に存在しているとすれば, 衣食住についてそれぞれ 150 個の正解極性の判定を行ったことになる。その結果, 「衣服」から 102 件, 「食料」から 110 件, 「居住施設」から 81 件の合計 293 語句の正解データを得た。

5.1.4 評価方法

評価極性の算出精度の評価は一致率で行う。計算方法を式 5.1 に示す。衣食住全ての正解データをまとめた 293 語を出題数とし, 正解データと機械が判定した評価極性が一致

表 5.1: 「衣服」のテスト名詞句と正解極性 (被験者 5 名)

| No. | 名詞句 | 好評/不評 (状況：一般) | 好評/不評 (状況：夏) | 好評/不評 (状況：冬) |
|-----|-------------|---------------|--------------|--------------|
| 1 | 沢山の洗い物 | 0/5(不評) | 0/5(不評) | 0/5(不評) |
| 2 | 赤いアロハ | 3/2 | 4/1 | 0/5(不評) |
| 3 | 舞台の衣装 | 5/0(好評) | 5/0(好評) | 5/0(好評) |
| 4 | 軽いインナー | 5/0(好評) | 5/0(好評) | 4/1 |
| 5 | お気に入りの上着 | 5/0(好評) | 4/1 | 5/0(好評) |
| 6 | いとこのお下がり | 1/4 | 1/4 | 1/4 |
| 7 | ウールのカーディガン | 5/0(好評) | 0/5(不評) | 5/0(好評) |
| 8 | 固いカッターシャツ | 3/2 | 2/3 | 3/2 |
| 9 | 予備の着替え | 5/0(好評) | 5/0(好評) | 5/0(好評) |
| 10 | 美しい着物 | 5/0(好評) | 5/0(好評) | 5/0(好評) |
| 11 | 毛皮のコート | 5/0(好評) | 0/5(不評) | 5/0(好評) |
| 12 | 可愛いコスチューム | 5/0(好評) | 5/0(好評) | 5/0(好評) |
| 13 | 汚い作業着 | 0/5(不評) | 0/5(不評) | 0/5(不評) |
| 14 | ユニクロのジーパン | 4/1 | 4/1 | 3/2 |
| 15 | 替えの下着 | 5/0(好評) | 5/0(好評) | 5/0(好評) |
| 16 | お気に入りの私服 | 5/0(好評) | 5/0(好評) | 5/0(好評) |
| 17 | 黒いジャージ | 5/0(好評) | 3/2 | 4/1 |
| 18 | 格好いいジャケット | 5/0(好評) | 2/3 | 5/0(好評) |
| 19 | 薄いシャツ | 5/0(好評) | 5/0(好評) | 1/4 |
| 20 | 厚いジャンパー | 4/1 | 0/5(不評) | 5/0(好評) |
| 21 | 白いスーツ | 2/3 | 1/4 | 2/3 |
| 22 | 短いスカート | 4/1 | 4/1 | 2/3 |
| 23 | 厚めのズボン | 5/0(好評) | 0/5(不評) | 5/0(好評) |
| 24 | 学校の制服 | 5/0(好評) | 5/0(好評) | 5/0(好評) |
| 25 | 分厚いセーター | 4/1 | 0/5(不評) | 5/0(好評) |
| 26 | 黒のタイツ | 5/0(好評) | 2/3 | 5/0(好評) |
| 27 | 軽いTシャツ | 5/0(好評) | 5/0(好評) | 1/4 |
| 28 | 柔道の道着 | 2/3 | 2/3 | 1/4 |
| 29 | 美しいドレス | 5/0(好評) | 5/0(好評) | 5/0(好評) |
| 30 | 高いトレンチコート | 5/0(好評) | 1/4 | 5/0(好評) |
| 31 | 暖かい長ズボン | 5/0(好評) | 0/5(不評) | 5/0(好評) |
| 32 | しまむらの長袖 | 3/2 | 0/5(不評) | 4/1 |
| 33 | 去年の夏服 | 2/3 | 3/2 | 0/5(不評) |
| 34 | フワフワの寝間着 | 5/0(好評) | 2/3 | 5/0(好評) |
| 35 | お気に入りのパジャマ | 5/0(好評) | 5/0(好評) | 5/0(好評) |
| 36 | 安い半ズボン | 2/3 | 3/2 | 0/5(不評) |
| 37 | お古の半袖 | 0/5(不評) | 2/3 | 0/5(不評) |
| 38 | 黒のビキニ | 5/0(好評) | 5/0(好評) | 2/3 |
| 39 | お気に入りの服 | 5/0(好評) | 5/0(好評) | 5/0(好評) |
| 40 | モコモコの冬着 | 5/0(好評) | 0/5(不評) | 5/0(好評) |
| 41 | 新しい冬服 | 5/0(好評) | 1/4 | 5/0(好評) |
| 42 | 初めての振袖 | 5/0(好評) | 4/1 | 5/0(好評) |
| 43 | 学校のブレザー | 5/0(好評) | 0/5(不評) | 5/0(好評) |
| 44 | チェックのポロシャツ | 5/0(好評) | 5/0(好評) | 2/3 |
| 45 | 新作の水着 | 5/0(好評) | 5/0(好評) | 1/4 |
| 46 | 紺色の浴衣 | 5/0(好評) | 5/0(好評) | 2/3 |
| 47 | サッカーのユニフォーム | 5/0(好評) | 5/0(好評) | 4/1 |
| 48 | 高い洋服 | 5/0(好評) | 5/0(好評) | 5/0(好評) |
| 49 | 男の和服 | 5/0(好評) | 5/0(好評) | 4/1 |
| 50 | 白のワンピース | 5/0(好評) | 5/0(好評) | 2/3 |

表 5.2: 「食料」のテスト名詞句と正解極性 (被験者 5 名)

| No. | 名詞句 | 好評/不評 (状況：一般) | 好評/不評 (状況：夏) | 好評/不評 (状況：冬) |
|-----|---------------|---------------|--------------|--------------|
| 1 | イチゴ味のかき氷 | 4/1 | 5/0(好評) | 0/5(不評) |
| 2 | 大きな豆 | 4/1 | 4/1 | 4/1 |
| 3 | 大量のご飯 | 4/1 | 2/3 | 4/1 |
| 4 | 熱い緑茶 | 4/1 | 0/5(不評) | 5/0(好評) |
| 5 | 美味しい納豆 | 5/0(好評) | 5/0(好評) | 5/0(好評) |
| 6 | 熱々のクラムチャウダー | 5/0(好評) | 0/5(不評) | 5/0(好評) |
| 7 | パサパサのレタス | 0/5(不評) | 0/5(不評) | 0/5(不評) |
| 8 | 美味しいアルフォート | 5/0(好評) | 5/0(好評) | 5/0(好評) |
| 9 | ショートケーキの生クリーム | 5/0(好評) | 5/0(好評) | 5/0(好評) |
| 10 | 汚いジャガイモ | 0/5(不評) | 0/5(不評) | 0/5(不評) |
| 11 | 温かいコーヒー | 5/0(好評) | 1/4 | 5/0(好評) |
| 12 | 大きな大根 | 5/0(好評) | 5/0(好評) | 5/0(好評) |
| 13 | 新鮮な野菜 | 5/0(好評) | 5/0(好評) | 5/0(好評) |
| 14 | 冷たいビール | 5/0(好評) | 5/0(好評) | 4/1 |
| 15 | 美味しい芋 | 5/0(好評) | 5/0(好評) | 5/0(好評) |
| 16 | 美味しい酒 | 5/0(好評) | 5/0(好評) | 5/0(好評) |
| 17 | 魚の肝 | 0/5(不評) | 0/5(不評) | 0/5(不評) |
| 18 | ぬるいお茶 | 1/4 | 2/3 | 2/3 |
| 19 | 熱い日本酒 | 5/0(好評) | 0/5(不評) | 5/0(好評) |
| 20 | 臭いチーズ | 0/5(不評) | 0/5(不評) | 0/5(不評) |
| 21 | 辛いキムチ | 3/2 | 3/2 | 3/2 |
| 22 | 固いニンジン | 1/4 | 1/4 | 0/5(不評) |
| 23 | 熱いおでん | 5/0(好評) | 0/5(不評) | 5/0(好評) |
| 24 | 高級なワイン | 5/0(好評) | 5/0(好評) | 5/0(好評) |
| 25 | 辛い寿司 | 0/5(不評) | 0/5(不評) | 0/5(不評) |
| 26 | 新作のお菓子 | 5/0(好評) | 5/0(好評) | 5/0(好評) |
| 27 | ローズマリーの紅茶 | 5/0(好評) | 4/1 | 5/0(好評) |
| 28 | ベタベタのチョコレート | 0/5(不評) | 0/5(不評) | 0/5(不評) |
| 29 | おでんの玉子 | 5/0(好評) | 4/1 | 5/0(好評) |
| 30 | 豚骨のラーメン | 5/0(好評) | 5/0(好評) | 5/0(好評) |
| 31 | 大きい饅頭 | 5/0(好評) | 5/0(好評) | 5/0(好評) |
| 32 | 冷たいジュース | 5/0(好評) | 5/0(好評) | 1/4 |
| 33 | 強い酒 | 2/3 | 2/3 | 2/3 |
| 34 | 甘いお菓子 | 5/0(好評) | 5/0(好評) | 5/0(好評) |
| 35 | カレーのニンジン | 4/1 | 4/1 | 4/1 |
| 36 | 温かいうどん | 5/0(好評) | 1/4 | 5/0(好評) |
| 37 | 安い肉 | 0/5(不評) | 0/5(不評) | 0/5(不評) |
| 38 | ぬるいビール | 1/4 | 0/5(不評) | 1/4 |
| 39 | 不味い酒 | 0/5(不評) | 0/5(不評) | 0/5(不評) |
| 40 | 小さなお菓子 | 2/3 | 2/3 | 2/3 |
| 41 | 冷たいご飯 | 0/5(不評) | 0/5(不評) | 0/5(不評) |
| 42 | 地元の酒 | 5/0(好評) | 5/0(好評) | 5/0(好評) |
| 43 | 美味しいワイン | 5/0(好評) | 5/0(好評) | 5/0(好評) |
| 44 | 大きな魚 | 5/0(好評) | 5/0(好評) | 5/0(好評) |
| 45 | 熱い鍋 | 5/0(好評) | 0/5(不評) | 5/0(好評) |
| 46 | 激辛のカレー | 1/4 | 1/4 | 1/4 |
| 47 | キンキンのビール | 5/0(好評) | 5/0(好評) | 3/2 |
| 48 | 卵の黄身 | 5/0(好評) | 5/0(好評) | 5/0(好評) |
| 49 | 鮭の皮 | 3/2 | 3/2 | 3/2 |
| 50 | 黒いバナナ | 0/5(不評) | 0/5(不評) | 0/5(不評) |

表 5.3: 「居住施設」のテスト名詞句と正解極性 (被験者 5 名)

| No. | 名詞句 | 好評/不評 (状況：一般) | 好評/不評 (状況：夏) | 好評/不評 (状況：冬) |
|-----|-------------|---------------|--------------|--------------|
| 1 | ぼろいアパート | 0/5(不評) | 0/5(不評) | 0/5(不評) |
| 2 | 大きい家 | 5/0(好評) | 5/0(好評) | 5/0(好評) |
| 3 | 寂しい一軒家 | 1/4 | 1/4 | 1/4 |
| 4 | 近所の屋敷 | 4/1 | 4/1 | 4/1 |
| 5 | 近くのマンション | 5/0(好評) | 5/0(好評) | 5/0(好評) |
| 6 | 海沿いの宿 | 5/0(好評) | 5/0(好評) | 2/3 |
| 7 | ヒノキの小屋 | 5/0(好評) | 5/0(好評) | 5/0(好評) |
| 8 | 自分の部屋 | 5/0(好評) | 5/0(好評) | 5/0(好評) |
| 9 | 遠い実家 | 1/4 | 1/4 | 1/4 |
| 10 | 安めのマンション | 4/1 | 4/1 | 3/2 |
| 11 | 小さい社宅 | 2/3 | 2/3 | 2/3 |
| 12 | 新しい住居 | 5/0(好評) | 5/0(好評) | 5/0(好評) |
| 13 | 沢山の住宅 | 2/3 | 2/3 | 2/3 |
| 14 | ぼろい宿舎 | 0/5(不評) | 0/5(不評) | 0/5(不評) |
| 15 | 近所の団地 | 2/3 | 2/3 | 2/3 |
| 16 | 安いアパート | 4/1 | 3/2 | 4/1 |
| 17 | 広い平屋 | 5/0(好評) | 5/0(好評) | 5/0(好評) |
| 18 | 海外の別荘 | 5/0(好評) | 5/0(好評) | 5/0(好評) |
| 19 | 暑い部屋 | 0/5(不評) | 0/5(不評) | 3/2 |
| 20 | 新築のマンション | 5/0(好評) | 5/0(好評) | 5/0(好評) |
| 21 | 田舎の民家 | 3/2 | 3/2 | 2/3 |
| 22 | 寒い山小屋 | 2/3 | 3/2 | 0/5(不評) |
| 23 | 大学の寮 | 3/2 | 3/2 | 3/2 |
| 24 | 安心の我が家 | 5/0(好評) | 5/0(好評) | 5/0(好評) |
| 25 | 狭い家 | 0/5(不評) | 0/5(不評) | 1/4 |
| 26 | 小さいアパート | 1/4 | 1/4 | 1/4 |
| 27 | 木造の山小屋 | 4/1 | 4/1 | 2/3 |
| 28 | ピカピカの新居 | 5/0(好評) | 5/0(好評) | 5/0(好評) |
| 29 | 地元の宿 | 4/1 | 4/1 | 4/1 |
| 30 | 狭いマンション | 0/5(不評) | 0/5(不評) | 1/4 |
| 31 | 嫁の実家 | 4/1 | 5/0(好評) | 4/1 |
| 32 | 沖縄の別荘 | 5/0(好評) | 5/0(好評) | 3/2 |
| 33 | 近くの宿 | 5/0(好評) | 5/0(好評) | 5/0(好評) |
| 34 | 芸能人の自宅 | 4/1 | 4/1 | 4/1 |
| 35 | 穴場のマンション | 5/0(好評) | 5/0(好評) | 5/0(好評) |
| 36 | 鳥取のアパート | 2/3 | 2/3 | 2/3 |
| 37 | 海の家 | 4/1 | 5/0(好評) | 0/5(不評) |
| 38 | 広いマンション | 5/0(好評) | 5/0(好評) | 5/0(好評) |
| 39 | 田舎の宿 | 4/1 | 4/1 | 4/1 |
| 40 | 北海道の別荘 | 5/0(好評) | 5/0(好評) | 2/3 |
| 41 | 日本の住宅 | 5/0(好評) | 5/0(好評) | 5/0(好評) |
| 42 | コンクリートのアパート | 4/1 | 4/1 | 4/1 |
| 43 | レンガの家 | 4/1 | 4/1 | 4/1 |
| 44 | 臭い部屋 | 0/5(不評) | 0/5(不評) | 0/5(不評) |
| 45 | 人気のマンション | 5/0(好評) | 5/0(好評) | 5/0(好評) |
| 46 | 秘密の隠れ家 | 5/0(好評) | 5/0(好評) | 5/0(好評) |
| 47 | 木造の宿 | 5/0(好評) | 5/0(好評) | 4/1 |
| 48 | 理想の家 | 5/0(好評) | 5/0(好評) | 5/0(好評) |
| 49 | 大きい住宅 | 5/0(好評) | 5/0(好評) | 5/0(好評) |
| 50 | 近所の集団住宅 | 2/3 | 2/3 | 2/3 |

した数の合計を一致数とする.

$$\text{一致率} = \frac{\text{一致数}}{\text{出題数}} \quad (5.1)$$

また, 比較手法として, 以下の機能を限定した評価も行う.

f1. 極性表現対選択機能 名詞の種類, 状況, 判断条件に応じて *SO-score* 算出に用いる極性表現対を選択する機能

f2. コーパス選択機能 状況に応じて使用するコーパスを可変化する機能

ベースラインには, Turney らの手法 [10] を用いる (すなわち, 極性表現に “良い”, “悪い” を用い, コーパスの選択は行わない).

5.1.5 実験結果

結果を表 5.4 に示す. 提案手法の f1 と f2 を両方使用した手法で最も高い F 値を得た.

表 5.4: 実験結果 (評価極性)

| 手法 | f1. 極性表現対選択機能 | f2. コーパス選択機能 | 一致数 | 出題数 | 一致率 |
|--------------|---------------|--------------|-----|-----|------|
| 提案手法 (f1+f2) | ○ | ○ | 215 | 293 | 0.73 |
| 比較手法 (f1 のみ) | ○ | × | 209 | 293 | 0.71 |
| 比較手法 (f2 のみ) | × | ○ | 199 | 293 | 0.68 |
| ベースライン手法 1 | × | × | 204 | 293 | 0.70 |

提案手法が優れていた例

- 入力データを表 5.5 に示す.

表 5.5: 入力データ

| 名詞句 | 名詞の種類 | 状況 | 判断条件 |
|-------|-------|----|------|
| 熱いおでん | 食料 | 夏 | 生理 |

- 提案手法での評価極性算出結果を表 5.6 に示す.
- 従来手法 (極性表現対に “良い”, “悪い” を使用) での評価極性算出結果を表 5.7 に示す.
- 従来手法では夏に熱いおでんが好評と出力され, 提案手法では不評と出力された. 正解極性は不評であるため, 提案手法の方が優れていた.

表 5.6: 評価極性算出結果 (提案手法)

| 入力 | | | | 出力 | 正解極性 |
|-------|-------|----|------|-----------|------|
| 名詞句 | 名詞の種類 | 状況 | 判断条件 | SO-score | |
| 熱いおでん | 食料 | 夏 | 生理 | -5.77(不評) | 不評 |

表 5.7: 評価極性算出結果 (従来手法)

| 入力 | | | | 出力 | 正解極性 |
|-------|-------|----|------|----------|------|
| 名詞句 | 名詞の種類 | 状況 | 判断条件 | SO-score | |
| 熱いおでん | 食料 | 夏 | 生理 | 0.87(好評) | 不評 |

提案手法が劣っていた例

- 入力データを表 5.8 に示す.

表 5.8: 入力データ

| 名詞句 | 名詞の種類 | 状況 | 判断条件 |
|----------|-------|----|------|
| お気に入りの上着 | 衣服 | 一般 | 対人関係 |

- 提案手法での評価極性算出結果を表 5.9 に示す.
- 従来手法 (極性表現対に “良い”, “悪い” を使用) での評価極性算出結果を表 5.10 に示す.
- 従来手法では一般的にお気に入りの上着が好評と出力され, 提案手法では不評と出力された. 正解極性は好評であるため, 提案手法の方が劣っていた. 調べてみると極性表現として評価極性の算出に使われた表現は, 好評表現集合 P が「高い, 丈夫, 人気, 新しい, ブランド」であり, 不評表現集合 N が「安い, ぼろい, イマイチ, 古い, 安物」であった. これらの表現の内, 出現回数で言えば P の単語の方が多くなっていたが, 出現確率の分母は N の方が小さかったため, 総合的に N の方に偏るという結果となっていた. 結果を改善させるためには, N の表現をコーパス内で出現し易い単語に変更するか, 「お気に入り」や「上着」で好評となる表現を P に追加する必要がある.

表 5.9: 評価極性算出結果 (提案手法)

| 入力 | | | | 出力 | 正解極性 |
|----------|-------|----|------|-----------------|------|
| 名詞句 | 名詞の種類 | 状況 | 判断条件 | <i>SO-score</i> | |
| お気に入りの上着 | 衣服 | 一般 | 対人関係 | -0.21(不評) | 好評 |

表 5.10: 評価極性算出結果 (従来手法)

| 入力 | | | | 出力 | 正解極性 |
|----------|-------|----|------|-----------------|------|
| 名詞句 | 名詞の種類 | 状況 | 判断条件 | <i>SO-score</i> | |
| お気に入りの上着 | 衣服 | 一般 | 対人関係 | 1.78(好評) | 好評 |

5.2 情緒推定の精度

本章では、3章で述べた提案手法により、状況を考慮した評価極性を用いた情緒推定の性能を調査するための実験を行う。

5.2.1 実験の目的

名詞句の評価極性を用いた判断条件の真偽判定の性能を評価し、それを用いた情緒推定の有効性を示すことを目的とする。

5.2.2 実験方法

- 5.1.3 節で作成した名詞句を含む文に状況を付与し、各文に人手で正解情緒タグを付与してテストデータを作成する。
- テスト文に各手法を用いて自動で情緒推定し、その出力を正解データと比較することで、再現率、適合率、F 値を調査する。

5.2.3 正解データ作成

正解極性を付与した名詞句 293 語と、3.3.4 節で収集した用言パターンを組み合わせることで入力文を作成する。その後、作成した入力文に人手で正解情緒を付与することで正解データを作成する。

入力文の作成 5.1.3 節で正解極性を付与した名詞句 293 語と、3.3.4 節で収集した用言パターンを組み合わせる入力文を作成する。実際に「熱いおでん」を使用した入力文を作成する場合を、順を追って解説する。

まず、「熱いおでん」は「食料」名詞句であるので、「食料」を扱うパターンとして抽出された用言パターン 148 件の中から、「熱いおでん」への使用に適した用言パターンを無作為に選択する（この時、できるだけ様々な種類の用言を使用するよう注意し、一部の用言に偏り過ぎないようにする）。例として「食べる」という用言を用いると、「熱いおでんを食べる」といった入力文が得られる。この入力文に、詳細情報として名詞の種類、用言、判断条件も付与したものをまとめて入力データとして扱い、表 5.11 のように表現する。また、入力分の作成の際には、以下の点を満たす文を作成することとする。

表 5.11: 入力文「熱いおでんを食べる」と文の詳細情報

| 入力文 | 名詞句 | 名詞の種類 | 用言 | 判断条件 |
|-----------|-------|-------|-----|------|
| 熱いおでんを食べる | 熱いおでん | 食料 | 食べる | 生理 |

- 否定語を含まない
- 動詞の時制が過去または現在
- 極性の反転として扱われる接続表現を含まない

入力データへの状況の付与 次に、入力データに状況を付与する。ここで付与される状況は、名詞句が正解極性を有する状況のみとする。例えば、「熱いおでん」は表 5.2 より、状況が「一般」のときは好評、「夏」の時は不評、「冬」の時は好評となっており、各状況全てで正解極性を有しているので、「一般」、「夏」、「冬」全ての状況を付与する。「イチゴ味のかき氷」の様に、「夏」と「冬」にしか正解極性を有していない場合は、付与される状況は「夏」、「冬」のみとなる。状況を付与した例を表 5.12, 5.13 示す。このとき、同時に正解極性の付与も行っておく。

表 5.12: 「熱いおでんを食べる」に状況を付与する場合

| 入力文 | 名詞句 | 名詞の種類 | 用言 | 判断条件 | 状況(付与した) | 正解極性(付与した) |
|-----------|-------|-------|-----|------|----------|------------|
| 熱いおでんを食べる | 熱いおでん | 食料 | 食べる | 生理 | 一般 | 好評 |
| 熱いおでんを食べる | 熱いおでん | 食料 | 食べる | 生理 | 夏 | 不評 |
| 熱いおでんを食べる | 熱いおでん | 食料 | 食べる | 生理 | 冬 | 好評 |

表 5.13: 「イチゴ味のかき氷を食べる」に状況を付与する場合

| 入力文 | 名詞句 | 名詞の種類 | 用言 | 判断条件 | 状況(付与した) | 正解極性(付与した) |
|--------------|----------|-------|-----|------|----------|------------|
| イチゴ味のかき氷を食べる | イチゴ味のかき氷 | 食料 | 食べる | 生理 | 夏 | 好評 |
| イチゴ味のかき氷を食べる | イチゴ味のかき氷 | 食料 | 食べる | 生理 | 冬 | 不評 |

入力データへの正解情緒の付与 最後に、各入力データに正解情緒を付与して正解データを作成する。今まで付与してきた情報から、一般的な人物が想像するであろう正解情緒を想像して付与する。ここでは例として、表 5.12 に対して正解情緒の付与を行う。この入力データでは、状況が「一般」の時に「熱いおでん」は好評であるので、好評なおでんを食べる時の情緒を推定する。推定する情緒は 2.1.3 節で述べた情緒 9 分類の中から 1 つを付与する。実際に付与を行った結果を表 5.14 に示す。以上の手順を「熱いおでん」

表 5.14: 「熱いおでんを食べる」に正解情緒を付与する場合

| 入力文 | 名詞句 | 名詞の種類 | 用言 | 判断条件 | 状況 | 正解極性 | 正解情緒(付与した) |
|-----------|-------|-------|-----|------|----|------|------------|
| 熱いおでんを食べる | 熱いおでん | 食料 | 食べる | 生理 | 一般 | 好評 | 好ましい |
| 熱いおでんを食べる | 熱いおでん | 食料 | 食べる | 生理 | 夏 | 不評 | 嫌だ |
| 熱いおでんを食べる | 熱いおでん | 食料 | 食べる | 生理 | 冬 | 好評 | 好ましい |

以外の名詞句でも行い、正解データを作成する。文の異なり数としては 116 個であり、状況の違いを含む事で 293 個の正解データを作成した。

5.2.4 評価方法

情緒の推定精度は、適合率、再現率、F 値によって評価する。適合率はシステム出力の正解率、再現率は問題に対する取りこぼしの指標であり、適合率は (5.2)、再現率は (5.3) の式で表される。F 値は適合率と再現率の調和平均であり、(5.4) の式を用いて算出する。

ここで、出題数は正解データの入力データ数 (293 個) を表し、出力数は機械が出力した情緒の総数を表し、一致数は正解データと機械が出力した情緒の一致数を表す。

$$\text{適合率} = \frac{\text{一致数}}{\text{出力数}} \quad (5.2)$$

$$\text{再現率} = \frac{\text{一致数}}{\text{出題数}} \quad (5.3)$$

$$F \text{ 値} = \frac{2 \times \text{適合率} \times \text{再現率}}{\text{適合率} + \text{再現率}} \quad (5.4)$$

ベースライン手法には以下の手法を用い、提案手法と比較する。

ベースライン手法 1 Turney らの手法である．つまり極性表現に“良い”，“悪い”を用い，コーパスの選択は行わない手法

ベースライン手法 2 判断条件の真偽判定に評価極性を用いず，全て T とする手法

5.2.5 実験結果

実験結果を表 5.15, 5.16, 5.17 それぞれに示す．表 5.15 が情緒 9 分類，表 5.16 が情緒 5 分類，表 5.17 が情緒 3 分類で実験した結果となっている．どの分類でも，提案手法が最も高い F 値を得られている．

提案手法の内，極性表現選択機能を f1，コーパス選択機能を f2 と表記している．

表 5.15: 実験結果 (情緒 9 分類)

| 手法 | f1 機能 | f2 機能 | 判断条件の充足判定 | 一致数 | 出題数 | 出力数 | 再現率 | 適合率 | F 値 |
|--------------|-------|-------|-----------|-----|-----|-----|------|------|------|
| 提案手法 (f1+f2) | ○ | ○ | ○ | 139 | 293 | 311 | 0.47 | 0.45 | 0.46 |
| 比較手法 (f1 のみ) | ○ | × | ○ | 139 | 293 | 311 | 0.47 | 0.45 | 0.46 |
| 比較手法 (f2 のみ) | × | ○ | ○ | 127 | 293 | 310 | 0.43 | 0.41 | 0.42 |
| ベースライン手法 1 | × | × | ○ | 133 | 293 | 336 | 0.45 | 0.40 | 0.42 |
| ベースライン手法 2 | × | × | × | 185 | 293 | 531 | 0.63 | 0.35 | 0.45 |

表 5.16: 実験結果 (情緒 5 分類)

| 手法 | f1 機能 | f2 機能 | 判断条件の充足判定 | 一致数 | 出題数 | 出力数 | 再現率 | 適合率 | F 値 |
|--------------|-------|-------|-----------|-----|-----|-----|------|------|------|
| 提案手法 (f1+f2) | ○ | ○ | ○ | 196 | 293 | 311 | 0.67 | 0.63 | 0.65 |
| 比較手法 (f1 のみ) | ○ | × | ○ | 192 | 293 | 311 | 0.66 | 0.62 | 0.64 |
| 比較手法 (f2 のみ) | × | ○ | ○ | 184 | 293 | 310 | 0.63 | 0.59 | 0.61 |
| ベースライン手法 1 | × | × | ○ | 196 | 293 | 336 | 0.67 | 0.58 | 0.62 |
| ベースライン手法 2 | × | × | × | 255 | 293 | 531 | 0.87 | 0.48 | 0.62 |

表 5.17: 実験結果 (情緒 3 分類)

| 手法 | f1 機能 | f2 機能 | 判断条件の充足判定 | 一致数 | 出題数 | 出力数 | 再現率 | 適合率 | F 値 |
|--------------|-------|-------|-----------|-----|-----|-----|------|------|------|
| 提案手法 (f1+f2) | ○ | ○ | ○ | 204 | 293 | 311 | 0.70 | 0.66 | 0.68 |
| 比較手法 (f1 のみ) | ○ | × | ○ | 199 | 293 | 311 | 0.68 | 0.64 | 0.66 |
| 比較手法 (f2 のみ) | × | ○ | ○ | 190 | 293 | 310 | 0.65 | 0.61 | 0.63 |
| ベースライン手法 1 | × | × | ○ | 200 | 293 | 336 | 0.68 | 0.60 | 0.64 |
| ベースライン手法 2 | × | × | × | 269 | 293 | 531 | 0.92 | 0.51 | 0.65 |

| 入力文 | 名詞句 | 名詞の種類 | 用言 | 判断条件 | 状況 | 正解極性 | 正解情緒 |
|-----------|-------|-------|-----|------|----|------|------|
| 熱い日本酒を飲んだ | 熱い日本酒 | 食料 | 食べる | 生理 | 夏 | 不評 | 嫌だ |

提案手法が従来手法より優れていた例

- 入力：「熱い日本酒を飲んだ。」

step1 入力文を結合価パターン辞書に照合する

この文は図 5.1 と図 5.2 のレコードにマッチする。

| |
|--------------------------------------|
| ◆日本語パターン：N1 が/は N2 を/は 飲む |
| ◆意味属性制約：N1(4人 535動物) N2(746液体 857飲物) |
| ◆判断条件：生理・近(N1, N2) ◆情緒原因：〈五感(味覚)〉 |
| ◆情緒主：N1 ◆情緒対象：N2 ◆情緒名：《好ましい》 |

図 5.1: 用言「飲む」の生理・近のレコード

| |
|--------------------------------------|
| ◆日本語パターン：N1 が/は N2 を/は 飲む |
| ◆意味属性制約：N1(4人 535動物) N2(746液体 857飲物) |
| ◆判断条件：生理・離(N1, N2) ◆情緒原因：〈五感(味覚)〉 |
| ◆情緒主：N1 ◆情緒対象：N2 ◆情緒名：《嫌だ》 |

図 5.2: 用言「飲む」の生理・離のレコード

step2 引数に具体的な格要素が代入された判断条件を得る

- ベースライン手法 1(Turney らの手法)
 - (1) 生理・近(ϕ , 熱い日本酒)
 - (2) 生理・離(ϕ , 熱い日本酒)
- 提案手法
 - (1) 生理・近(ϕ , 熱い日本酒)
 - (2) 生理・離(ϕ , 熱い日本酒)

step3 評価極性で判断条件の真偽判定を行う

- ベースライン手法 1(Turney らの手法)

$SO\text{-score}(\text{熱い日本酒}) = 1.14(\text{好評})$

表 5.18: 判断条件：生理の真理値表

| NP_y | 生理・近 (NP_x, NP_y) | 生理・離 (NP_x, NP_y) |
|--------|-----------------------|-----------------------|
| 好評 | T | F |
| 不評 | F | T |

- 提案手法

$$SO\text{-score}(\text{熱い日本酒}) = -5.34(\text{不評})$$

step4 正解情緒と各手法の出力情緒を比較する

- ベースライン手法 1(Turney らの手法) では《好ましい》のみが出力される
 - (1) 生理・近 ($\phi, \text{熱い日本酒}$) = $T \rightarrow$ 《好ましい》
 - (2) 生理・離 ($\phi, \text{熱い日本酒}$) = $F \rightarrow$ 《嫌だ》 (抑制)
- 提案手法では《嫌だ》のみが出力される
 - (1) 生理・近 ($\phi, \text{熱い日本酒}$) = $F \rightarrow$ 《好ましい》 (抑制)
 - (2) 生理・離 ($\phi, \text{熱い日本酒}$) = $T \rightarrow$ 《嫌だ》

この文からは出力情緒の候補として《好ましい》と《嫌だ》の2つが結合価パターン辞書より得られている。正解情緒が《嫌だ》なので、図 5.2 のレコードの判断条件だけが成り立つ必要がある。生理・離 ($\phi, \text{熱い日本酒}$) が真となるには、表 5.18 より、「熱い日本酒」の評価極性が不評となる必要がある。提案手法では不評と出力されていたので、提案手法の方が優れていた。

提案手法と従来手法が同じ結果になった例

- 入力：「汚い作業着に触る。」

| 入力文 | 名詞句 | 名詞の種類 | 用言 | 判断条件 | 状況 | 正解極性 | 正解情緒 |
|----------|-------|-------|----|------|----|------|------|
| 汚い作業着に触る | 汚い作業着 | 衣服 | 触る | 生理 | 一般 | 不評 | 嫌だ |

step1 入力文を結合価パターン辞書に照合する

この文は図 5.3 と図 5.4 のレコードにマッチする。

step2 引数に具体的な格要素が代入された判断条件を得る

-
- ◆日本語パターン：N1 が/は N2 に/は 触る
 - ◆意味属性制約：N1(4人 535動物) N2(2具体)
 - ◆判断条件：生理・近 (N1, N2) ◆情緒原因：〈五感 (触感)〉
 - ◆情緒主：N1 ◆情緒対象：N2 ◆情緒名：《好ましい》
-

図 5.3: 用言「触る」の生理・近のレコード

-
- ◆日本語パターン：N1 が/は N2 に/は 触る
 - ◆意味属性制約：N1(4人 535動物) N2(2具体)
 - ◆判断条件：生理・離 (N1, N2) ◆情緒原因：〈五感 (触感)〉
 - ◆情緒主：N1 ◆情緒対象：N2 ◆情緒名：《嫌だ》
-

図 5.4: 用言「触る」の生理・離のレコード

- ベースライン手法 1(Turney らの手法)
 - (1) 生理・近 (ϕ , 汚い作業着)
 - (2) 生理・離 (ϕ , 汚い作業着)
- 提案手法
 - (1) 生理・近 (ϕ , 汚い作業着)
 - (2) 生理・離 (ϕ , 汚い作業着)

step3 評価極性で判断条件の真偽判定を行う

- ベースライン手法 1(Turney らの手法)

$SO\text{-score}(\text{汚い作業着}) = -3.46(\text{不評})$
- 提案手法

$SO\text{-score}(\text{汚い作業着}) = -3.03(\text{不評})$

step4 正解情緒と各手法の出力情緒を比較する

- ベースライン手法 1(Turney らの手法) では《嫌だ》のみが出力される
 - (1) 生理・近 (ϕ , 汚い作業着) = $F \rightarrow$ 《好ましい》 (抑制)
 - (2) 生理・離 (ϕ , 汚い作業着) = $T \rightarrow$ 《嫌だ》
- 提案手法では《嫌だ》のみが出力される
 - (1) 生理・近 (ϕ , 汚い作業着) = $F \rightarrow$ 《好ましい》 (抑制)

(2) 生理・離 (ϕ , 汚い作業着) = $T \rightarrow$ 《嫌だ》

この文からは出力情緒の候補として《好ましい》と《嫌だ》の2つが結合価パターン辞書より得られている。正解情緒が《嫌だ》なので、図5.4のレコードの判断条件だけが成り立つ必要がある。生理・離 (ϕ , 汚い作業着) が真となるには、表5.18より、「汚い作業着」の評価極性が不評となる必要がある。ベースライン手法1と提案手法で共に不評と出力されていたので、ベースライン手法1と提案手法が同じ結果になった。

提案手法が従来手法より劣っていた例

- 入力：「沖縄の別荘に着く。」

| 入力文 | 名詞句 | 名詞の種類 | 用言 | 判断条件 | 状況 | 正解極性 | 正解情緒 |
|----------|-------|-------|----|------|----|------|------|
| 沖縄の別荘に着く | 沖縄の別荘 | 居住施設 | 着く | 目標実現 | 一般 | 好評 | 好ましい |

step1 入力文を結合価パターン辞書に照合する

この文は図5.5のレコードにマッチする。

| |
|--|
| ◆日本語パターン：N1 が/は N2 に/へ/は 着く |
| ◆意味属性制約：N1(4人 986乗り物 760人工物) N2(388場所 2610場) |
| ◆判断条件：目標実現・近 (N1, N2) ◆情緒原因：〈補助〉 |
| ◆情緒主：N1 ◆情緒対象：N2 ◆情緒名：《好ましい》 |

図5.5: 用言「着く」の目標実現・近のレコード

step2 引数に具体的な格要素が代入された判断条件を得る

- ベースライン手法1(Turneyらの手法)
 - (1) 目標実現・近 (ϕ , 沖縄の別荘)
- 提案手法
 - (1) 目標実現・近 (ϕ , 沖縄の別荘)

step3 評価極性で判断条件の真偽判定を行う

- ベースライン手法1(Turneyらの手法)

$SO\text{-score}(\text{沖縄の別荘}) = 0.83(\text{好評})$

- 提案手法

$$SO\text{-score}(\text{沖縄の別荘}) = -0.33(\text{不評})$$

表 5.19: 判断条件：目標実現の真理値表

| NP_y | 目標実現・近 (NP_x, NP_y) | 目標実現・離 (NP_x, NP_y) |
|--------|-------------------------|-------------------------|
| 好評 | T | F |
| 不評 | F | T |

step4 正解情緒と各手法の出力情緒を比較する

- ベースライン手法1(Turney らの手法)では《好ましい》が出力される
 - (1) 目標実現・近 (ϕ , 沖縄の別荘) = $T \rightarrow$ 《好ましい》
- 提案手法では《好ましい》が抑制され,《なし》が出力される
 - (1) 目標実現・近 (ϕ , 沖縄の別荘) = $F \rightarrow$ 《好ましい》(抑制) \rightarrow 《なし》

この文からは出力情緒の候補として《好ましい》が結合価パターン辞書より得られている。正解情緒が《好ましい》なので、図 5.5 のレコードの判断条件が成り立つ必要がある。生理・近 (ϕ , 沖縄の別荘) が真となるには、表 5.19 より、「沖縄の別荘」の評価極性が好評となる必要がある。提案手法では不評と出力されていたので、ベースライン手法1と比べると提案手法が劣る結果になった。評価極性の算出に用いられた極性表現は、好評表現集合 P が「手頃, 安い, 広い, 近い」、不評表現集合 N が「高い, 狭い, 遠い」であった。この中では、特に「遠い」が「沖縄」と共起することで不評極性に影響を与えていた。「沖縄」や「別荘」が好評となる極性表現を追加することで改善できる。ただし、極性表現を追加すると他の名詞句の評価極性にも影響するので、一部の名詞句の評価極性のみを改善するのは難しいと考えられる。

第6章 考察

本章では、第5章の結果から、本手法を見直す。具体的には、提案手法である *SO-score* 算出式の拡張による評価極性算出が正しく動作していたか、そして精度の向上がわずかであった原因について考察する。

6.1 *SO-score*による評価極性算出精度について

*SO-score*の算出において、ベースライン手法からの評価極性算出精度の向上はわずかであった。考えられる原因として、極性表現対知識ベースの所有する極性表現対のデータ不足が考えられる。

極性表現対知識ベースの有する 210 対の極性表現対は、本稿の著者が手作業で付与していったものであり、名詞の種類、状況、判断条件の3つの組み合わせを手がかりに、連想する極性表現を随時追加したものである。そのため、見落としした極性表現が多く、物事の評価のベクトル(味, 臭い, 重さ, etc...)を知識ベースに列挙しきれていないという問題が残っている。この問題は、新しく極性を一つ追加するだけで、今まで好評であった単語が急に不評となるといった可能性も含んでいる。絶対数を増やすために、形容詞用法辞書等に記載されている表現を極性表現とみなし、機械的に追加する方法が考えられる。日本語の表現は形容詞だけでも 3000 語程存在するので、絶対数の不足は解消できる。この絶対数を増やすことで、物事を評価する際の切り口が増え、様々な要素について評価して評価極性を算出することが可能になるので、信憑性の向上にも繋げることができる。

また、他にも評価極性を算出する場合、コーパス内での出現頻度が多いほど評価極性算出精度の信憑性を向上させられる。しかし、コーパスを増やしたとしても、出現頻度が1桁台となる語句は存在し、そのような語句は評価極性が不安定になってしまう。本研究では頻度不足で *SO-score* が算出出来なかった名詞句はほとんど無かったため、提案手法での *SO-score* 算出式の拡張により、共起頻度の集め易さは向上していると考えられる。

極性表現を連想してゆく作業において、コーパスの種類によって極性表現対の表記を

使い分けた方がよい場合があった。簡単に言えば、ブログ文では口語的、現代風な表現が多く出現するため、作成する極性表現対もそれに沿った表現にした方がよいということである。例えば、極性表現対に「格好いい-格好悪い」があったとすると、「イケた-ダサイ」のような極性表現を用いる方が共起頻度が集め易い場合もある。本研究ではブログ文コーパスを *SO-score* の算出に使っているが、知識ベース作成の際にはその点に注意して作成するよう心がけた。因みに、コーパスに新聞記事等の体裁が整っている物ではなく、ブログ文を用いているのは、ブログ文に、筆者による直接的な感情表現や、一般的な人の赤裸々な意見や経験談が記述されているためであり、より人間らしいコーパスとして利用できるためである。

最後に提案手法の精度を求める際に、本研究では全ての極性表現対を使用した精度をまとめて求めるようにした。しかし、効果のある極性表現対を見分けるためには、極性表現対を一つ追加する度にそれによる影響を実験して求める必要があった。

表 5.4 の結果を見ると、コーパス絞り込み機能だけではベースラインよりも精度が下がっているが、極性表現選択機能と組み合わせることで、極性表現選択機能単体よりも精度が向上していることが確認できる。これは、“良い”、“悪い”以外として連想した極性表現が、絞り込んだコーパス先で上手く出現していることを表しており、絞り込んだ少量のコーパスでも、より正確な評価極性が算出可能になるということを示していると考えられる (リソースの削減効果)。

6.2 複雑な名詞句への対応

従来手法では、名詞句が複雑になればなるほどコーパス中での出現頻度が得られなくなるため、名詞句の評価極性の算出に適していなかった。本研究では「*A* 名詞の *B* 名詞」型名詞句、「*A* 形容詞 *B* 名詞」型名詞句、「*A* 形容動詞 *B* 名詞」型名詞句を主な研究対象として実験を行ったが、提案手法の拡張方法であれば、名詞句を構成する品詞毎にバラして式に代入することができるため、3 単語以上からなる名詞句にも適用は可能である。しかし、「夏休みの最初の日」と「夏休みの最後の日」のように、仮に「夏休み」が強い好評極性を持っていたとしても、極性を反転するような単語が含まれている場合、名詞句全体の評価極性が上手く算出できないかもしれない。

6.3 夏と冬以外の状況への対応

本研究では、状況として「夏」と「冬」の2つを取り扱った。これにより、主に気温に関する要素に変化が得られ、物事の温度に関する極性表現対知識ベースを作成できた。

気温の状況以外に極性表現対を作成する必要があるとすれば、物の値段に関係する状況等があれば対人関係において極性表現を付与し易くなる。例として状況「裕福/貧乏」、状況「量(安さ)優先/質(ブランド性)優先」などが考えられる。この2セットは組み合わせて一つの状況となり、4つの状況が生じる。例えば、「貧乏でも質優先」や、「裕福で質優先」といった状況が挙げられる。これらの状況を使えば極性表現である「安い/高い」を付与し易くなると考えられる。

他にも状況を挙げ出すときりがないが、今後の課題として状況は入力文から自動で判別したり、入力文の前後文から求めたいと考えている。

第7章 おわりに

本研究では、情緒推定精度を向上させるため、評価極性を算出する際に *SO-score* 算出式を拡張し、極性表現対を可変化し、*SO-score* 算出の際のコーパスも可変化するという方法で改良を行った。

極性表現対を収録した極性表現対知識ベースを作成することにより、名詞句の種類、状況、判断条件を考慮した極性表現対の取得が可能になった。

そして、名詞句の評価極性の算出精度を評価した結果、テスト名詞句 293 語に対し、従来手法 (極性表現に “良い”, “悪い” を使用) では一致数が 204 件であったが、提案手法では 215 件と 11 件増加し、従来手法に比べて精度が 0.70 から 0.73 に向上した。

また、評価極性を用いた情緒推定の性能を評価した結果、テストデータ 293 個に対し、従来手法 (極性表現に “良い”, “悪い” を使用) で出力された情緒 336 個のうち一致したものは 133 個となった。一方、提案手法で出力された情緒 311 件のうち一致したものは 139 個となった。これはつまり、より少ない出力でより多い一致数を得たということであり、精度が向上している。しかし、精度の向上は 0.42 から 0.46 という結果であった。

今後の課題は、本手法をより多くの状況、名詞の種類に適応させること、状況を入力文やその前後文から自動で判別することである。

謝辞

本研究を進めるにあたり，徳久雅人講師には，始終に渡り研究の進め方や論文の書き方など，細部に渡る御指導を頂きました．ここに深く感謝致します．また，ご多忙の中，本論文をまとめるにあたって岩井儀雄教授に御助言を頂きました．ここに厚くお礼申し上げます．研究を進めるにあたって種々の御助言を頂きました村田真樹教授に心から御礼申し上げます．3年間に渡り，ご指導頂きました村上仁一准教授に心から御礼申し上げます．

また，情緒推定実験など，様々な場面で御助力を頂きました計算機工学C講座の皆様に深く感謝致します．

参考文献

- [1] 目良和也, 市村匠, 相澤輝昭, 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.
- [2] 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.
- [3] 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.
- [4] 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.
- [5] 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.
- [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] 徳久雅人, 岡田直之: “パターン理解的手法に基づく知能エージェントの情緒生起”, 情報処理学会論文誌, 39(8), pp.2440-2451, 1998.
- [8] 野口和樹, 滝川晃司, 徳久雅人: “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”, 電子情報通信学会技術研究報告, 思考と言語, TL2011-32, pp.63-68, 2011.
- [9] 竹本翔: “修飾語句の評価極性を利用した情緒推定”, 鳥取大学工学部知能情報工学科卒業論文, 2011.
- [10] Peter D. Turney: “Thumbs Up or Thumbs down? Semantic Orientation Applied to Unsupervised Classification of Reviews”, Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, pp.417-424, 2002.