

常識的感情判断システムと主体語処理

Commonsense Emotion Judgment System and Processing of the Subject

風間勇志† 植野義孝† 渡部広一† 河岡司†
KAZAMA Yuji, UENO Yoshitaka, WATABE Hirokazu and KAWAOKA Tsukasa

1. はじめに

現在、我々を取り巻く環境におけるコンピュータの存在は、大きくなる一途を辿っているが、その“これから”を考えたときに人とのインターフェイスに普段、人と人との間に存在する意思疎通をモデル化、適用することが望まれることは必然であり、その主たるものは会話である。本稿では、会話システムを構築する上で必要不可欠となる感情に関する判断を行うためのシステム[1]に、新たな主体語処理の導入による精度向上の過程と結果を報告する。

2. 感情判断システム

感情判断システム[1]とは会話文から発話者の感情を判断するシステムであり、入力情報として{主体語, 目的語, 動詞}の、感情判断に関係する3要素が、適切な自然言語処理の結果抽出されたことを前提としている。即ち抽出された情報は極めて基本的な文([主体語]+[目的語]+[動詞]のSOV構造)となる。

このシステムの位置付けを以下の図1に示す。

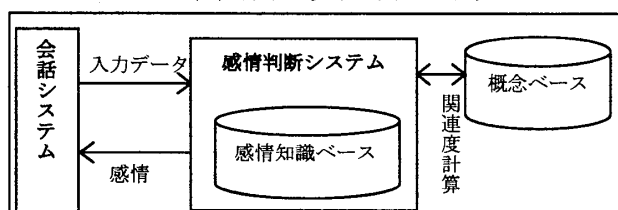


図1 感情判断システムの位置付け

3. 主体語処理導入の必要性

従来の感情判断システム[1]において、先述の入力情報は以下のように定義されていた。

[主体語] … 一人称の名詞(“私”)に固定

[目的語] … 修飾語(形容詞・形容動詞)+名詞, 又は名詞のみ

[動詞] … 複合動詞を除く原型の動詞

即ち、従来の感情判断システムでは、自分自身が主体であった場合のみの感情判断のみを可能としているのである。

ここで発話者が以下の文を発した場合を考える。

[文1] 父親が沢山のお金を得る

[文2] 泥棒が沢山のお金を得る

[文1], [文2]とも沢山のお金を得た主体自身は“喜び”を感じている。しかし、この文の発話者の感情はそうはいかないのである。

確かに[文1]であれば話者自身も“喜び”を感じていると判断できる、しかし、[文2]であれば、常識的に考えて発話者は“怒り”を感じている。

このような、主体とその文の発話者との感情の間に何らかの“ずれ”の表現の実現がもたらす感情判断の発展を考えたとき、主体語処理の導入は必要不可欠となるのである。なお、主体語処理導入において、本稿では[主体語]を以下のように定義する。

[主体語] … 主体語となり得る”人”を表す名詞(二人称は含まない)

4. 新たな感情判断システムの構成

従来の感情判断システムが主体語の感情判断を行っているということは既に述べたが、そこに図2のように主体語処理を導入する。

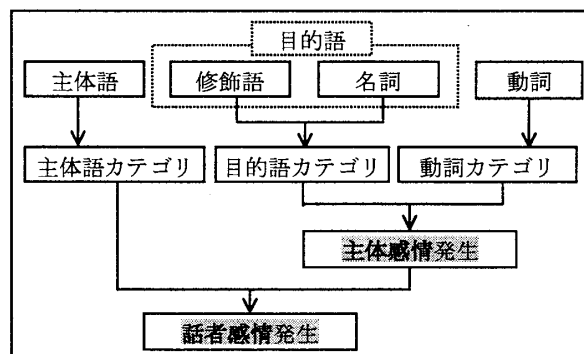


図2 主体語処理の導入過程

図2を見ると、主体語、目的語、動詞の各々を“カテゴリ”に処理の過程で置き換えているがこれらは種々の語の膨大な組み合わせを有用な代替語(カテゴリ)へ変換したものであり、変換の際に語の連想を行い、その手法としては概念ベースによる関連度計算を利用している。

5. 概念ベースと関連度計算

概念ベースとは語(概念)に複数の関係のある語(属性)を持たせた知識ベースであり、表1に示すようなデータが約10万語(概念数)格納されている。

関連度計算[2]とはこの概念ベースを用いて二つの概念AとBの関連性を定量的に評価するものである。即ち、関連度の値が概念間の繋がり(強さ)をあらわし、人が普段行っている言葉による連想の働きを担っている。

表1 概念ベースの構造

概念	属性
夏	肝試し, ビーチパラソル, かき氷, 冷麺...
医者	医者, 診察, 往診, 名医, 診る, 医師...
お茶	緑茶, 抹茶, 番茶, 茶道, 喫茶, 飲み物...

† 同志社大学大学院工学研究科
Graduate School of Engineering, Doshisha University

6. 主体語処理手法の決定

主体語処理において、主体語となり得る“人”を表す名詞を扱うことは既に述べたが、それらを分類するにあたって各名詞を3つの角度から常識的範疇の内で評価し、各々にA B Cのランクを設け、その組み合わせ、即ち計27通りに分類した。それらを以下に記述する。

- (1) 精神的距離 好き嫌い (好き - A B C - 嫌い)
- (2) 感覚的距離 身近さ (近い - A B C - 遠い)
- (3) 社会的尺度 善悪 (善 - A B C - 悪)

[文1], [文2]へこの分類を適用すると, [文1]の父親は(1, 2, 3) = (A, A, B) [文2]の泥棒は(C, B, C)となる。この分類で話者に対する主体語の印象を区別する。主体語と話者との間の感情の“ずれ”の表現は[話者感情生成テーブル]で定義している。

7. 主体語知識ベースの構築

主体語知識ベースには主体語となり得る“人”を表す名詞を、シソーラス[3]を基に約9000語格納した。また、概念ベースによる関連度計算、及び未知語処理を効率的に行うために、シソーラスから階層関係を継承した構造とし、これらの名詞の上位階層に約440語の語の代表語を付加した。

なお、シソーラスとは一般名詞の意味的用法を表す意味属性の上位下位関係、全体部分関係が木構造で示された辞書である。

8. 主体語処理における未知語処理手法の確立

先述の主体語知識ベースを持ってしても、主体語となり得る“人”をあらわす名詞全てを網羅することは出来ない。知識ベース内のどの語にも当てはまらない語は未知語として扱い、知識ベース内の語との概念ベースを用いた関連度計算を利用することによって、主体語カテゴリのいずれかにあてはめる。この処理が未知語処理である。

なお、未知語処理としては、関連のある語群でパラメータ毎に、ランク単位のグループを作成し、それらのグループのうち、未知語と最も高関連にあるグループのランクを未知語に付加することで主体語カテゴリへ分類する手法を用いた。

9. 話者感情生成テーブル

ここで、主体感情と主体語カテゴリとの組み合わせから発生する話者感情を定義した。

また、話者の感情に関しては、“喜び”、“安心”、“怒り”、“恐れ”、“悲しみ”、“落胆”、“後悔”、“罪悪感”、“恥”の計9種類の感情それぞれに対して最小0, 最大10の値を割り振っており、複数の感情の強弱をつけた表現を実現した。この、話者感情生成テーブルを表2に示す。

表2 話者感情生成テーブルの構造(例示)

精神的 距離	感覚的 距離	社会的 尺度	主体 感情	喜 び	安 心	怒 り	...
A	A	A	喜び	10	7	0	...
A	A	B	喜び	8	5	0	...
B	A	C	怒り	7	3	0	...
...	...	...	...	...	...	...	...

10. 評価

常識的な主体感情が判断された入力データを採用したアンケートにより収集した約530文による話者感情の正誤判定を行った。その結果を図3に示す。

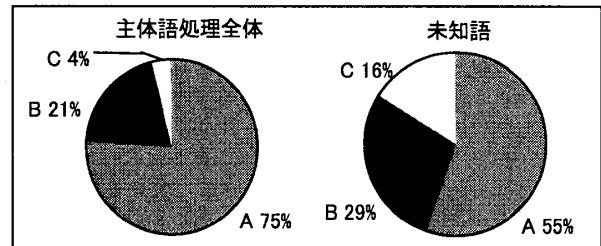


図3 評価結果

なお、Aは常識的に正しいもの、Bは非常識な答えではないもの、Cは非常識な答えであったものであり、未知語に関しては全体から未知語を抽出し、注目した値であり未知語の数は80語である。評価に使用した例文の一部を以下に示す。

[例文]

- ・泥棒がお金を拾う。
- ・私がお金を拾う。
- ・私が大切な約束を破る。
- ・息子が受験に失敗する。

11. 考察と今後の展望

評価から、この主体語処理手法はほぼ十分な精度を得られたと言える。従ってこの手法の利用は感情判断システムの精度向上に貢献出来ると考えられる。さらなる精度向上を考えたとき、複合名詞を主体語として扱えない、ということの解決する必要があることが明らかになったが、現在複合名詞を扱う研究が進んでいるため、その結果によってこれは解消される問題であると考えられる。

未知語処理に関しても改善の余地はあるが、未知語が全体に占める割合を考えると、その精度向上よりも先述の複合名詞の処理を改善することが先決であると考えられる。

また、現在のシステムの入力情報は主語・目的語・動詞であり、それらの組み合わせにより感情判断を行っている。しかし、より自然な会話文に対応できるようにする必要もある。

謝辞

本研究は文部科学省からの補助を受けた同志社大学の学術フロンティア研究プロジェクトにおける研究の一環として行った。

参考文献

- [1]土屋 誠司, 渡部 広一, 河岡 司 常識的感情判断メカニズムの構築 同志社大学理工学研究報告 Vol.43, No1, pp1-11(2002)
- [2]渡部広一, 河岡司 常識的判断のための概念間の関連度評価モデル, 自然言語処理, Vol.8, No2, pp39-54 (2001)
- [3]NTT コミュニケーションズ科学研究所 日本語語彙体系, 岩波書店 (1997)