

未知語処理に基づく感情判断メカニズムの実現

3J-8

馬場秀樹 土屋誠司 渡部広一 河岡司
同志社大学大学院 工学研究科

1. はじめに

人間とコンピュータの会話をより人間的なものにするため、コンピュータが会話を通じ話し手の感情を判断することが望まれる。しかし、自然言語の構文解析や語の意味理解だけでは感情を判断することは困難である[1]。

コンピュータに感情知識ベースをもたせ、入力された言葉（刺激）に対して人間と同じような感情を発生させるメカニズムの基本構成法については既に報告している[2]。本稿では、この方式を拡張し、知識ベースにない未知の語が入力された場合にも対応できる未知語処理のメカニズムについて報告する。

2. 感情判断メカニズム

感情判断は、入力された言葉からコンピュータが話し手の感情を類推するものである。本メカニズムは、入力情報に対する適切な自然言語処理の結果として、感情判断に関わる刺激語（主体語（名詞）、対象語（名詞）、変化語（動詞））3要素の抽出を前提としている。主体語は感情発生要因の主体（感情の度合いを判断）、対象語は要因となる事物、変化語は要因となる行為・作用である。

例えば「私はプレゼントを貰った」であれば刺激語は{私, プレゼント, 貰う}となり、「弟が試験に合格した」であれば{弟, 試験, 合格}となるが、いずれの場合も話し手の感情を予測する。尚、扱う感情は喜び・悲しみ・怒り・恐れ・恥・安心・落胆・嫌悪・罪悪感の9種類とした。

図1に感情判断の流れを示す。対象語と変化語はそれぞれ対応する分類語に集約されており、発生する感情は感情生成テーブル（{対象語の分類語, 変化語の分類語} → 感情）により定義されている。

変化語については語数（約1200語）が少ないため全ての語を分類語に分け感情知識ベースに登録している。一方、対象語については語数が多いため全てに登録しておくことは困難である。このため、代表となる対象語（象徴語）を定義しこれらを分類語に登録しておき、登録されていない語（未知語）が入力された場合、語の意味的な近さを関連度として計算し、最も関連度の高い象徴語で代替する。この際、感情生成に関わる観点（美しいなど）での未知語処理が必要となるため、分類語は象徴語の性質を表す語（形容詞）としている[2]。例えば、象徴語「宝石」は「美しい」という分類語に分類されている。

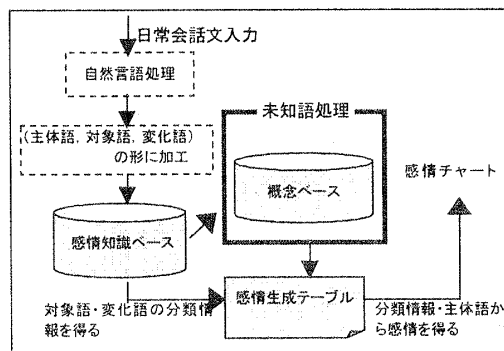


図1 感情判断の流れ

3. 未知語処理と問題点

対象語の未知語処理では、約5万語の概念ベースから未知語が抽出され、これと感情知識ベースに登録されている全ての代表語との関連度を計算し[3]、代替できる象徴語を決定する。未知の対象語を適切な分類語に対応付けるには、概念ベースの語属性として感情判断に関わる十分な属性を含むことが望まれるが概念ベースの語属性は国語辞書等から機械的に獲得しておりあまり期待できない。

既報告の未知語処理方式[2]の問題は次の二点である。

[問題I]: 象徴語と未知語の一般的な共通性のため関連度は高くなるが感情判断として期待している分類語の性質「美しい」などでは正しい評価となっていない。

例えば、概念ベースの語「宝石」には感情発生に有効な情報である「美しい」「装飾」「珠玉」などの属性と共に一般的な「無機物」「硬い」「鉱石」「鉱脈」等の属性も存在する。対象語の関連度計算では後者の属性が雑音となり期待する象徴語を見いだすことができない。

[問題II]: 名詞によっては鳥のように一部の鳥のみが分類語「美しい」に属するものもあり、鳥全体を象徴語とすると判断を誤る。

例えば白鳥を象徴語とした場合、白鳥の持つ一般属性（翼など）からその他の鳥（雀）も白鳥に分類されてしまい、結果として雀は「美しい」となってしまう。

表1 感情知識ベース構成

分類語	象徴語	代表語（反例）
美しい	宝石	白鳥（カラス）
	美術品	玉虫（ゴキブリ）
	花	流星（隕石）
痛い	怪我	感電（電気）
	疾患	火傷（火） 落馬（馬）

注. 代表語は今回の改良で付け加えた

4. 象徴語による自動精練と反意語付き代表語の導入

問題 I を解決するため、関連度計算の前に以下の方法で象徴語の自動精練を行う。

①特徴語の導入：象徴語を特徴づける特徴語 5 つを自動精練のキーとして人手で用意する。象徴語が「宝石」ならば「瑠璃・玉・珠玉・真珠・光沢」を特徴語とする。

②自動精練：概念ベース内の語「宝石」の 30 個の属性から宝石の特徴語と一致する属性 (5 個) を抽出する。次に、宝石の残りの属性 (25 個) と各特徴語の属性 (概念ベースから抽出) を候補として、各特徴語との関連度平均を求め高い順に宝石の残り 25 個の属性を決定する。

③関連度計算：「宝石」の自動精練された新しい属性で未知語「x」との関連度を計算する。

次に、問題 II を以下の方法で解決する。

①反例付き代表語の導入：問題 II の例に示すように白鳥のような語を象徴語とすると問題が発生する語は (代表語, 反例語) の形式で反例語と組にする (表 1)。

②関連度計算：未知語はまず各代表語との関連度計算を行う。高い関連度が得られた場合、その代表語の反例語との関連度計算を行い、これが低い場合のみ未知語はその代表語が属する分類語に決定される。

改良された未知語処理の一連の流れを図 2 に示す。これらの方式により未知語処理の大幅な改善が可能となる。

5. 実験と評価

新しい未知語処理方式の効果を以下の実験により評価した。本稿で行ったのは分類語「美しい」に関する実験である。

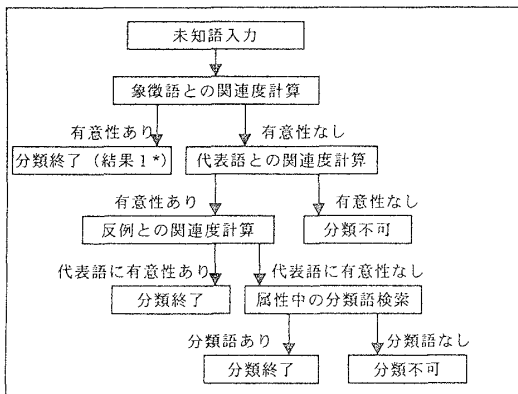


図 2 未知語処理の流れ (*結果 1 は図 4 に対応)

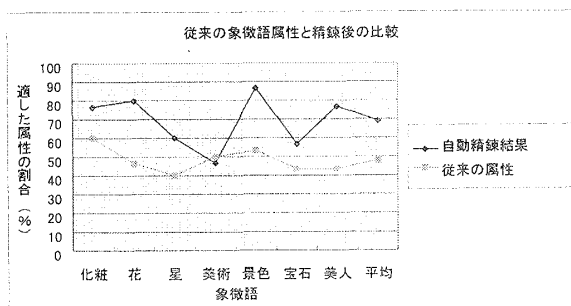


図 3 象徴語の自動精練結果 (一部)

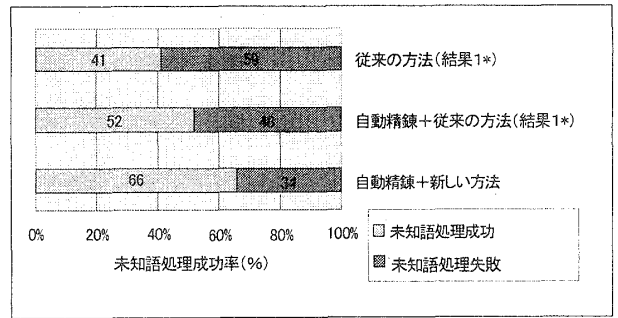


図 4 自動精練前後及び新旧未知語処理法での実験結果 (分類語「美しい」について)

[実験 1]：自動精練は全象徴語を対象とし、それぞれの特徴語を各 5 語ずつ手作業で用意した。4 節の方法で精練を行い、元の概念ベースから得られる属性情報からどれくらい精練されたかを検証した。評価は人間から見て各象徴語の属性 30 語のうち、いくつの属性が「美しい」と関連があるかで行った。結果の一部を図 3 に示す。一部精練後の結果が悪化した語があるが、自動精練後の方が平均 20% 程度、適した属性の割合が増加しているのが分かる。

[実験 2]：自動精練結果の検証、及び新しい未知語処理法に関する実験では、辞書から美しいと思われる名詞 50 語と美しいと思われない名詞 50 語を無作為に選んで、計 100 語を実験データとした。そして自動精練の前後で従来の未知語処理を行った場合と、自動精練後さらに新しい未知語処理法 (図 2) を行った場合の 3 通りで改善効果 (どれだけ正しく分類されているか) を評価した。評価法は 100 語中正しく判断できた語数の割合とした (図 4)。

実験結果から、自動精練によって未知語処理性能は 11% 向上し、さらに新しい未知語処理法で 14% 向上しているため、最終的に 25% 向上している。

以上の実験から、未知語処理は当初の意味的な処理から、感情判断を念頭に置いた処理にかなり修正でき、その性能も向上したことが伺える反面、未だ 34% の未知語は正しく処理できていない。これは今後、概念ベース関連の向上と本稿で述べた代表語等の改善によって解決していく。

6. 終わりに

今回提案した未知語処理方法は、人手による少量の知識をキーとして機械生成された大量の知識を自動精練する方法で種々の観点での一般的な未知語処理に適用可能である。

参考文献

[1]板橋：“知能・知識と情報” 近代科学社 p.110(1992)
 [2]馬場 渡部 河岡：“知的コミュニケーションのための語概念に基づく感情予測”, 情報処理学会第 58 回全国大会 4J-02(1999)
 [3]入江 渡部 河岡 松澤：“知的メカニズムのための概念間の類似度定量化方式”, 情報処理学会第 58 回全国大会 1M-06(1999)