常識的判断システムにおける未知語処理方式
～ニューラルネットワークと関連度偏差値による方式～

1P−4

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

常識を踏まえて物事を扱うことができる「常識的判断システム」の実現において，既知語についての判断は，知識を参照することができ，問題ではないが，未知語が入力された場合，その扱いは非常に難しい問題となる．

そこで，本稿では，概念ベースや関連度計算，ニューラルネットワークを用いることにより，意味的な関連の深さを考慮して，入力された未知語を，判断知識ベースに登録された代表語に関連付ける新しい未知語処理方式を提案し，その有効性を実験により検証する．

2. 概念ベース

本稿に用いた概念ベースは，複数の国語辞書などの語義文から自立語を抽出したもので，構成する各概念の構造は，重み情報のない属性語 a_i の集合のみの最も単純な構成のものを基に，質の向上を目的にした精錬操作を施し，ルールにより適切な重みを付加した約 12 万の概念からなるものを使用する．尚，属性語 a_i の語数は重みの大きいものから 30 個を上限とする[1]．

概念 X:{a_1, a_2, ･･･, a_i, ･･･}

3. 関連度

関連度とは，概念の関連性を定量的に評価するものであり，概念連鎖により概念を 2 次属性まで展開したところで，一致する属性の個数を評価することにより算出するものである [2]．具体的には，2つの概念がもつ各 30 語の 1 次属性の対応をとり，それらの 2 次属性の一致数を基に評価している．

概念 A と概念 B の2つの n 次属性 a_i^n と b_j^n の一致度を Match(a_i^n , b_j^n)，1 次属性数を N_A, N_B とし，関連度を Rel(A, B)とすると，以下のように表すことができる．

$$Rel(A,B)=\left(\left(\sum_{i=1}^{N_A}Match\left(a_i^1,b_{xi}^1\right)\right)/N_A+\left(\sum_{i=1}^{N_B}Match\left(a_i^1,b_{xi}^1\right)\right)/N_B\right)/2$$

4. 未知語処理

「常識的判断システム」全体は，量や感覚，感情に関する基本的なごく少数の知識（約5千語：物事の大小関係，夕焼け− 赤いなど）で構成する判断知識ベースサブシステムと，語概念間の関連度を評価する概念連想メカニズムで構成している．

しかし，システムに入力される語の多くは，判断知識ベースには陽に表現されていない未知語となるため，判断知識を利用するためには，概念連想メカニズムを

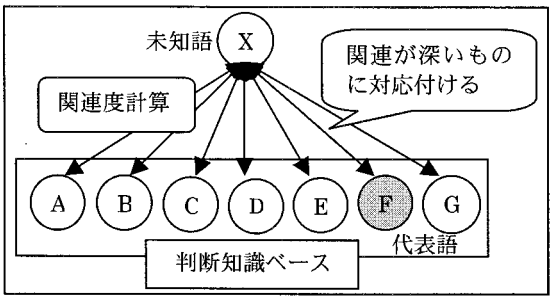

図1．未知語処理のイメージ

構成する概念ベースや関連度計算を用いることにより，これらの未知語について，意味的な関連やその強さの度合いを評価し，最も関連の強い代表語を決定する必要がある．この処理を総称して「未知語処理」と呼ぶ（図1）．次に未知語処理の方法を示す．

4. 1. 関連度偏差値による未知語処理方式

従来の未知語処理方式[3]は関連度偏差値によるものである．まず，常識的判断システムに未知語が入力されると，判断知識ベースに登録されている全代表語との意味的な近さを関連度計算により算出する．そして，全代表語を母集団としてその関連度の値を偏差値に変換する．算出された偏差値がある設定された値を超えていれば，その代表語に未知語は意味的に近いと判断し，対応付けることができる．尚，意味的な近さを表すある値は，実験的に算出したものであり，その値は偏差値 88 と設定している．

しかし，比較対象が代表語に限定されるため，無理やりにでもある代表語に関連付けようとする傾向が強くなってしまう．つまり，ある未知語が全代表語に対して関連が少なく低関連度を出力したにもかかわらず，その中で少し差があると高偏差値になってしまう欠点がある．

4. 2. ニューラルネットワークによる未知語処理方式

4．1．の欠点を解決するために，ニューラルネットワークによる未知語処理方式を提案する．この方式では，ある未知語と同義・類義，密，疎の各関係の間に現れる一致度（関連度計算過程に用いる各属性間の一致割合）を，降順に並び替えた時のパターンの差に着目し，同義・類義の関係と判断された代表語の中で，関連度の値が最高のものに未知語を対応付ける．

ニューラルネットワークの学習に使用したデータは，概念ベースに存在する語で，ある概念 X に対して，同義・類義，密，疎の各関係にある語を人手で抽出したものを 200 セット（計 600 パターン）使用している．

Unknown Word Processing Method for The Common Sense Judgment System Using Neural Network and the Deviation of Degree of Association.
TSUCHIYA Seiji, WATABE Hirokazu, KAWAOKA Tsukasa
Graduate School of Engineering, Doshisha University.

この未知語処理方式では，無作為に概念ベースから学習データを抽出し，一致度をパターンとしてとらえるため，概念ベースの雑音の影響を最小限に抑える効果があり，また，無理やり代表語に対応付けることはない．

5. 実験と考察

本稿で提案した未知語処理方式を評価・検証するために，常識的感情判断メカニズムに適用する．このメカニズムにおいて，未知語処理が必要な部分は名詞と動詞であるが，本論文では，名詞について処理を行う．

メカニズムに入力された名詞が感情判断知識ベースになければ，つまり，未知語であった場合，続いて日常使用される名詞を階層的に体系化した日本語語彙体系[5]を基に，感情発生に関係する語を人手で抽出し，再構築した感情シソーラス（約3万語）を検索する．そして，この処理で対処できなければ，今回問題にしている未知語処理方式を適用することになる．

5. 1. 評価方法

入力される未知語には，感情発生に寄与する語と，関係のない語の2種類がある．表 1 にそれぞれの語において，常識・非常識を評価する方法を示す．

表 1. 評価方法

	関係あり	関係なし
常識的	正しく分類	分類しない
非常識	誤って分類 分類しない	誤って分類
非常識 ではない	誤りとはいえない 語に分類	誤りとはいえない 語に分離

評価は常識度（常識的と評価する割合）と非常識度（非常識と評価する割合）の2つの尺度で，人手で行う．感情判断知識ベースの代表語は 99 語であり，反対の意味（分類）であるものは約半分存在するので，でたらめに未知語処理を行った場合，常識度は約1%であり，非常識度は約 50%となる．

実験データとして用いた未知語は，我々が日常的に使用している語をランダムに 500 語選出したものである．その内訳としては，感情発生に関係する語が 201 語（40.2%），関係ない語が 299 語（59.8%）である．

5. 2. 未知語処理の評価と考察

未知語処理を関連度偏差値方式（DAD），ニューラルネットワーク方式（NN）と複合型方式（NN+DAD）で行い検討した．結果を図 2 に示す．

本稿で問題にしている未知語処理のみの常識度については約 11%，非常識度については約 6.8% 精度が向上している．尚，常識的感情判断メカニズムにおける未知語処理全体での評価としては，常識度が 20.3%，非常識度は 15.5% 精度が向上している（図3）．

これより，感情発生に寄与しない語を無理やり関連付けず，自然な関連を見出すニューラルネットワーク方式の効果が十分に発揮されているのがわかる．また，実際のシステムとしては，図3に示す結果を使用するため，複合型方式が非常に有効な方式であるといえる．

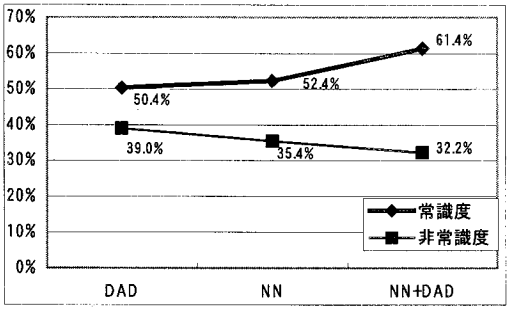

図2. 未知語処理のみの評価

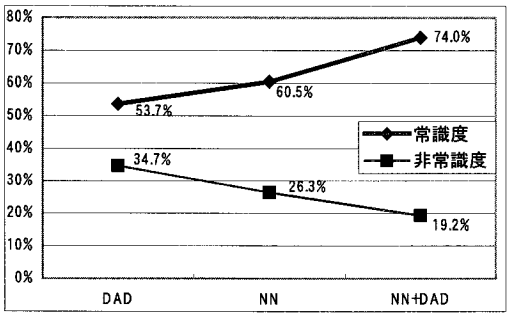

図3. 常識的感情判断メカニズムにおける未知語処理の評価

6. まとめ

本論文では，概念ベースや関連度計算，ニューラルネットワークを用いることにより，意味的な近さを考慮して，未知語を，（代表語）既知語に対応付ける新しい未知語処理方式を提案し，その有効性を常識的感情判断メカニズムに適用した実験により検証した．また，他の常識的判断メカニズムにおいても，基本的に常識的感情判断メカニズムと同様の構造をとっているため，本未知語処理方式は，非常に有効であると考えられる．

尚，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ものである．

参考文献

[1]眞鍋康人，小島一秀，渡部広一，河岡司：“概念間の関連度やシソーラスを用いた概念ベースの自動精錬法”，同志社大学理工学研究報告，Vol.42，No.1，pp.9-20，2001.

[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，No.2，pp.39-54，2001.

[3]土屋誠司，馬場秀樹，渡部広一，河岡司：“入力文から感情を判断するシステムにおける未知語の処理”，電子情報通信学会，信学技報，AI99-107，2000.

[4]馬場秀樹，渡部広一，河岡司：“知的コミュニケーションのための感情判断メカニズム”，同志社大学理工学研究報告，Vol.41，No.1，pp.16-24，2000.

[5]NTT コミュニケーション科学研究所監修：“日本語語彙体系”，岩波書店，1997