

知的判断メカニズムにおける概念間の関連度計算方式

3 J-7

入江 毅 東村 貴裕 渡部 広一 河岡 司

同志社大学大学院 工学研究科 知識工学専攻

1. はじめに

知的メカニズムは今後の情報処理システムのキーとなる技術であり、その中核となる機構は概念ベース^[1]と概念の関連性を利用した連想機能と考えられる。ここで、概念ベースは約5万語の概念と各概念の意味特徴を表現する単語（属性）と概念に対する重みの集合で構成されており、属性数は各概念によって異なる。しかし、維持管理の面から概念ベースは可能な限り単純な構造、つまり全ての属性が等しい値の重みを持ち、さらに全ての概念が同じ属性数を持つことが望まれる。本稿では、このような条件を満たす、より適切な固定属性数を概念間関連度計算方式により評価した。

2. 概念間関連度

概念間関連度とは、概念の関連性を定量的に評価するものであり、具体的には概念連鎖により概念をn次属性まで展開したところで、一致する属性個数を評価することにより算出するものであるが、今回の評価実験においては2次属性まで展開する方式^[2]を用いた。

3. 評価対象品詞

概念ベースは名詞・動詞・形容詞など、複数の品詞により構成されているが、今回は名詞のみを評価の対象とした。また、名詞ははっきりした形態を表現する「具体名詞」と、概念的な意味を表現する「抽象名詞」に分類され、これら2種類の名詞は意味・性質的に異なるため、別々に評価した。

4. 評価実験における前処理

評価実験に際して、次の前処理を行った。

4.1 属性数の均一化

概念ベースに登録されている語概念の属性数は各概念によって異なるため、固定属性数を評価することは困難である。よって、全ての概念が100個の属性を持つように概念ベースを再構成した。この再構成された属性集合はその概念の意味をそれなりに適切に表現するものと仮定する。

Measuring Semantic Relation Model between Concepts for Intelligent Mechanism
Takeshi Irie, Takahiro Higashimura, Hirokazu Watabe, Tsukasa Kawaoka
Faculty of Engineering, Doshisha University

4.2 属性の関連度による順位付け

評価実験を行う際、100個の属性からの属性選択法は無数にある。そこで、仮の固定属性数を“30”として概念とその属性100個との関連度を算出し、属性を関連度順に並べた。

5. 評価実験方法

固定属性数の評価法として次の方法を用いた。

5.1 同義語・無関係語による評価

具体・抽象名詞それぞれ100組の同義語・無関係語(例を表1に示す)を用意し、重みの降順に10通りの属性数 $N(N=10, 20, \dots, 100)$ における関連度の平均を算出する。ここで、同義語においては関連度が“1”に、また無関係語においては“0”に近い程、優れた結果といえる。

表1：同義語・無関係語の例

同義語		無関係語	
具体名詞	抽象名詞	具体名詞	抽象名詞
自動車 * 車	悪運 * 不運	筆箱 * 罎	決心 * 発展
医者 * 医師	改正 * 改善	猫 * 机	完了 * 着想
辞典 * 辞書	功績 * 業績	魚 * 書籍	規約 * 天性

5.2 尺度による評価

ある基本概念に対し、人間の感覚において「関連が深い(Rank_A)」、「関連がある(Rank_B)」、「関連がない(Rank_C)」と思われる3つの概念を想定し、それら4つの概念を1組の尺度として、具体・抽象名詞それぞれについて100組(例を表2,3に示す)を用意し、5.1と同様に10通りの属性数における関連度を算出する。さらに先の人間の感覚に適合すれば1点、そうでない場合は0点の評価点を与え、平均評価得点を算出する。

表2：具体名詞を用いた尺度の例

基本概念	Rank_A	Rank_B	Rank_C
自動車	車	機械	魚
本	書籍	柵	蜜柑
鳥	翼	動物	絵

表3：抽象名詞を用いた尺度の例

基本概念	Rank_A	Rank_B	Rank_C
解決	決着	判断	応対
決行	敢行	行動	健康
天才	俊才	才能	退職

6. 評価実験結果

6.1 同義語・無関係語における実験結果

図1は無関係語における実験結果であり、具体・抽象名詞ともにNの増加に伴ってほぼ線形に関連度が増加することがわかる。ここで、理想的な関連度はNに関わらず“0”であり、“0”より大きくなるのはNの増加に伴い、その概念にとって不適切な属性が増加し、それら不適切な属性同士が一致するために関連度が大きくなると考えられる。よって各Nにおける関連度は信頼できない関連度、つまり誤り関連度： $Rel_F(N)$ とみなせる。

また、図2は同義語における実験結果であるが、示されている関連度の全てが信頼できる関連度であるとはいえず、無関係語における関連度と同様、各Nには相応分の誤り関連度が含まれていることが考えられる。よって、信頼できる関連度： $Rel_T(N)$ は式(1)で表すことができ、信頼できる関連度のみを補正されたものを図3に示す。

$$Rel_T(N) = Rel(N) \times (1 - Rel_F(N)) \dots (1)$$

$Rel(N)$: 図2の各Nにおける関連度

図3より、具体名詞においてはN=20で最大値を取り、Nが増加するに伴って関連度が減少していることがわかる。しかし抽象名詞においてはほとんど関連度の変化が見られない。

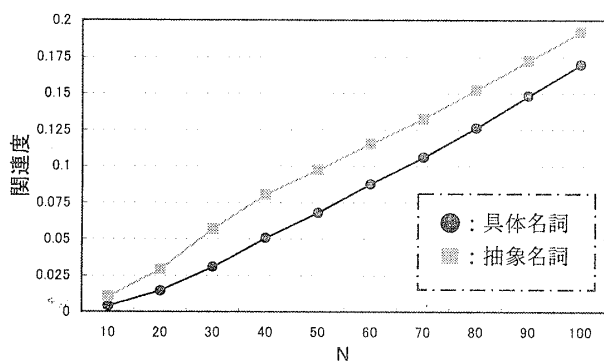


図1：各Nにおける平均無関係語関連度

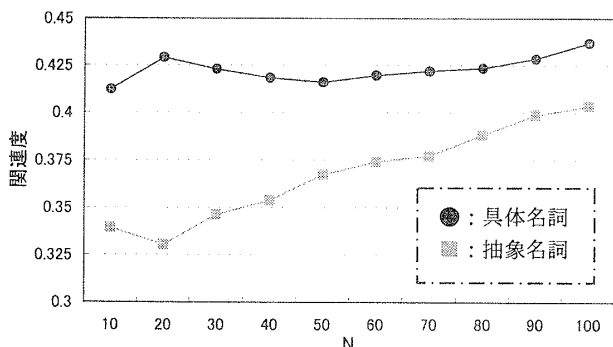


図2：各Nにおける平均同義語関連度

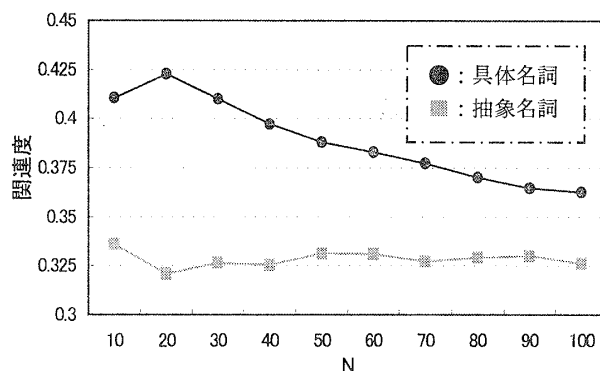


図3：各Nにおける平均同義語補正関連度

6.2 尺度を用いた実験結果

図4より、具体名詞においてはN=30で最大値を取り、N=30~50までは減少、N=50以降はほぼ一定となった。また、抽象名詞ではN=30~70ではほぼ一定となり、N=70以降減少していることがわかる。よって尺度による評価においては、具体・抽象名詞に関わらず、N=30が最適な固定属性数であることがわかった。

また、2つの評価方法を考察すると、具体・抽象名詞に関わらず、N=20~30がより適切な固定属性数であることがわかった。

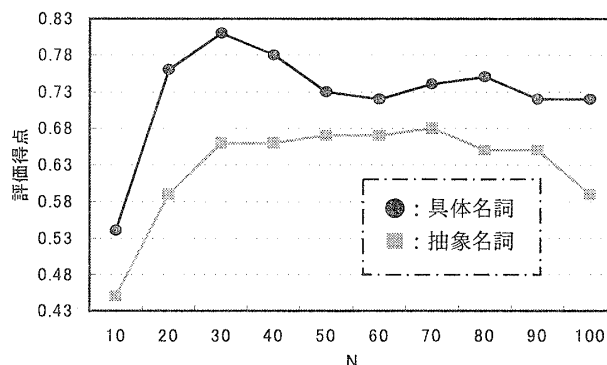


図4：各Nにおける平均評価得点

7. おわりに

今回の評価実験により、概念ベースの名詞においては、概念間関連度計算によってN=20~30がより適切な固定属性数であることがわかった。今後は動詞・形容詞など他の品詞についても同様な調査をする必要がある。

参考文献

- [1] 笠原要, 松澤和光, 石川勉: 国語辞書を利用した日常語の類似性判別, 情報処理学会論文誌, Vol. 38, No. 7, pp. 1272-1283 (1994)
- [2] 入江毅, 渡部広一, 河岡司, 松澤和光: 知的判断メカニズムのための概念間の類似度評価モデル, 信学技法, Vol. 98, No. 499, p47-54 (1999)