

# F-051 関連度を用いた概念ベースの新属性への重み付け手法 A Weight Generation Method to The New Attribute of The Concept-Base Using The Degree of Association

奥村 紀之\*  
Noriyuki Okumura

渡部 広一\*  
Hirokazu Watabe

河岡 司\*  
Tsukasa Kawaoka

## 1. はじめに

近年、コンピュータの発展はめざましく、より知的なインタフェースを持つコンピュータが望まれるようになることは明らかである。そこで、知的なインタフェースを実現する手段として、常識判断メカニズムが提案されている。常識判断メカニズムとは、人間の常識をコンピュータ上で模倣し、例えば、「キムチ→辛い」のような判断を行うメカニズムのことである。

本研究では、常識判断メカニズムを実現するために必要となる、連想メカニズムにおける概念ベースの学習手法について検討を行う。概念ベースは、概念をその意味・特徴を表す属性と、その重要性を示す重みの対の集合によって定義しており、本稿では、従来の属性信頼度 [1] を利用することなく大域的情報を利用し、簡易に重み付けを行う手法を提案する。

## 2. 概念ベース

概念ベースは、電子辞書等から機械的に構築され、約9万語の概念を蓄えた大規模知識ベースである。概念は、ある語  $A$  をその語と関連の強いと考えられる語 (属性)  $a_i$  と重み  $w_i (> 0)$  の対の集合として定義する。属性数  $n$  は概念ごとに異なる。

### • 概念の定義

概念  $A = \{(a_1, w_1), (a_2, w_2), \dots, (a_n, w_n)\}$

概念自身の属性を1次属性と呼び、任意の1次属性  $a_i$  は、必ず概念ベース内に含まれる語で定義する。つまり、属性を表す語もまた概念として定義されている。従って、1次属性それぞれを1つの概念と見た場合、1次属性からさらにそれぞれの属性を導くことができる。これを、概念に対する2次属性と呼び、同様に属性を導くことで概念は  $n$  次元までの属性連鎖集合により定義されている。

## 3. 関連度計算方式

### 3.1 意味関連度計算方式 [2]

2つの概念  $A, B$  の意味関連度は、概念  $A, B$  の二次属性を利用し一致度 [2] を求め、一致度の和が最大になるように一次属性の組み合わせを作る。ここで、一次属性が完全一致する場合は特殊処理とする。概念ベースには約9万の概念が存在し、属性が一致することは稀である。従って、属性の一致を特別処理にすることにより、属性が一致した場合の評価を大きくすることができる。2つの概念  $A, B$  の意味関連度を  $MR(A, B)$  と定義する。

### 3.2 意味共起関連度計算方式 [3]

意味関連度計算方式と共起関連度計算方式 [3] を合成して評価する方式を意味共起関連度計算方式と呼び、2つの概念  $A, B$  の意味共起関連度を  $MCR(A, B)$  と定義する。

## 4. 属性数による影響

関連度計算方式では、演算に使用する属性数を30に限定してある。そのため、30以上属性を持つ概念については、重み上位30個までしか演算に影響しない。したがって、大量の属性を含む概念に対しては、適切と考えられる属性が上位となるように重み付けを行う必要がある。また、30個に満たない概念に関しては、すべての属性が関連度計算に使用されるため、極力雑音属性を排除し、適切な属性のみを付与する必要がある。図1に概念ベースに含まれる概念を属性数の階級別に示す。図1か

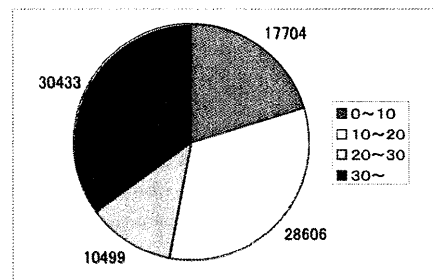


図1: 各概念の持つ属性数の分布

ら、概念ベース中に含まれる概念のおよそ60%の概念については属性数が30に満たないことがわかる。

## 5. 重み付け手法

現在の概念ベースに付与されている重みは、辞書の構造解析を行った際に得られる頻度情報や、シソーラスを利用した上位・下位関係等、国語的な情報を利用した属性信頼度によって付与されている。しかし、今後展開される概念学習において、国語的な情報を的確に取得することは困難であると考えられる。

そこで、以下の実験により、属性信頼度による重み付けと比較し、簡易に重み付けを行う手法を提案する。

### 5.1 関連度と $idf$ による重み付け

情報検索分野における  $tf \cdot idf$  による重み付け手法を利用し、関連度を  $tf$  とみなした重み付け手法を以下の式

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

で定義する. ある概念  $A$  に対する属性  $a_i$  の重み  $w_i$  は,

$$w_i = MR(A, a_i) \times idf_{a_i}$$

$$w_i = MCR(A, a_i) \times idf_{a_i}$$

で与えるものとする.

## 5.2 関連度と $idf$ を再帰的に利用した重み付け

5.1 を再帰的に行うことにより, 重みを定義する. 関連度  $\times idf$  を重みとし, その重みを用いて新たに算出した関連度を再び  $tf$  として利用する. この作業を繰り返して行うことで重みの変化を考察する. また, 繰り返し回数は 5 回である.

なお, 初期重みとして, 現行の概念ベースに付与されている重みを用いた場合, すべての重みを 1 とした場合, すべての重みを概念ベースを 3 次属性まで展開した際に得られる  $idf$  値とした場合について実験を行った.

## 6. 評価方法

( $X$ : 図書館,  $A$ : 書物,  $B$ : 勉強,  $C$ : 気軽い) のような 4 つの概念の組を用意する. ここで,  $X$  は任意の概念 (基準概念) であり,  $A$  は  $X$  と極めて密な概念,  $B$  は概念  $X$  に密な概念,  $C$  は概念  $X$  に疎な概念である. このような各評価用データの組に関して

$$Rel(X, A) > Rel(X, B) > Rel(X, C)$$

( $Rel(A, B)$ : 概念  $A$  と概念  $B$  の関連度)

を満たせば, その関連度計算結果を正解とし, それ以外は不正解とする. このような評価用データ 2370 組で, 何組のデータが正解したかによって概念ベースの精度評価を行う.

## 7. 実験と考察

### 7.1 意味関連度計算方式による重み付け

従来の関連度計算方式である意味関連度計算方式を用いた実験では, 図 2 に示すとおり, 概念ベースの重みを使用した場合は, 処理を加えることにより属性信頼度による重み付けよりも精度が低下することに対し, 重みを 1 および  $idf$  にした場合は, 一回目の処理で属性信頼度での重み付けと比較して差が見られなかった. また, 繰り返し処理によって徐々に精度が低下する傾向にある.

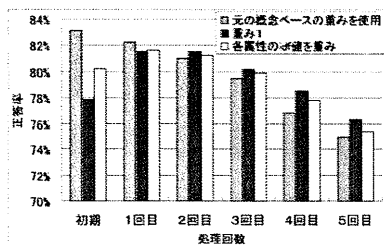


図 2: 意味関連度計算方式による重み付け

### 7.2 意味共起関連度計算方式による重み付け

意味共起関連度計算方式を用いた実験では, 図 3 に示すとおり, 傾向として意味関連度計算方式の実験と同様の結果が得られた. ただし, 意味関連度計算方式による操作と比較すると, 精度の低下幅は小さく, なだらかな傾向を示している.

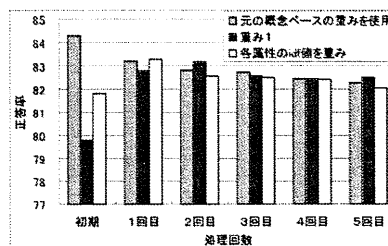


図 3: 意味共起関連度計算方式による重み付け

## 8. おわりに

本稿では, 関連度を利用し, 国語的な情報を利用せず, 大域的な情報を用いて簡易に重み付けを行う手法について提案した. 本手法は関連度計算の方式に依存せず, 有効に利用できると言えるだろう. また, 精度が現行の概念ベースと比較し遜色ないことから, 属性数の多い概念に対して, 適切な重みを付与することができ, 上位 30 個に適切な属性を与えられたと考える. そして, 属性数の少ない概念に対する新属性の追加の際の重み付け手法として利用可能と思われる. 今後, 初期に与える重みを 1 でよいのか,  $idf$  を与えるべきであるかを検討していく必要がある.

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“知能情報科学とその応用”における研究の一環として行った.

## 参考文献

- [1] 小島一秀, 渡部広一, 河岡司. 連想システムのための概念ベース構成法—属性信頼度の考え方に基づく属性重みの決定. 自然言語処理, Vol. 9, No. 5, pp. 93-110, 2002.
- [2] 井筒大志, 渡部広一, 河岡司. 概念ベースを用いた連想機能実現のための関連度計算方式. 情報科学技術フォーラム FIT2002, pp. 159-160, 2002.
- [3] 青田正宏, 渡部広一, 河岡司. 概念の意味・表記と共起情報を用いた関連度計算方式. 同志社大学理工学研究報告, Vol. 45, No. 1, pp. 23-34, 2004.