

# 概念の意味と表記を用いた関連性評価法

5X-3

東村 貴裕 渡部 広一 河岡 司  
同志社大学 工学研究科 知識工学専攻

## 1 はじめに

人間は曖昧な情報を受け取り適宜に解釈して適切に会話を進めることが出来る。これは、人間が長年にわたって蓄積してきた言語やその基本となる語概念に関する「常識」を持っているからである。本研究では、語と語の関連性について、コンピュータにも人間の常識的な感覚に近い判断をさせることをねらうものである。このような常識的判断を可能とするメカニズムは、利用者の意図を汲み取ることの出来る人間的な情報処理システムの基盤として役立つと考えている。

本稿では、連想メカニズムの基盤となる概念ベースを利用し、意味を表す属性の集合を用いる意味関連度計算方式と概念の表記を用いる漢字関連度計算方式による概念間の関連性評価法について提案している。

## 2 概念ベースと評価尺度

本研究では、関連性評価法の1つとして用いる関連度計算方式[1](以降、意味関連度計算方式と呼ぶ)において概念ベース[2]を利用している。また、各関連性評価法の評価実験には評価尺度を利用している。本章では、概念ベースと評価尺度について述べる。

### 2.1 概念と概念ベース

ある概念Cは、その概念の意味特徴を表す単語の集合で表現し、このような単語の集合(最大30語)を概念Cの1次属性と呼ぶ。また、概念Cの表記も単語により表現されている。例えば、概念“飛行機”は{推進, 噴射, 翼, 航空機, …}という1次属性で表現されている。

概念ベースには、このように定義された概念が約4万語収録されている。

また、概念Cの1次属性は単語であるので、その単語を概念と見なせば、1次属性はさらにその1次属性(概念)の意味特徴を表す単語の集合で表現できる。これらの単語の集合を概念Cの2次属性と呼ぶ。

さらに、概念Cの2次属性もその意味特徴を表す単語の集合で表現できる。このように、概念Cは、 $n$ 次属性まで定義することができる。

### 2.2 評価尺度

本研究では、表1に示すように人間の感覚において「基本概念MXと関連が深い(MA)」、「基本概念MXと関連がある(MB)」、「基本概念MXと無関連である(MC)」と思われる3つの概念を一組の尺度(MX-MA, MB, MC)として590組を人手により用意した。

表1. 評価尺度

MX	MA	MB	MC
ご飯	飯	米	青空
安易	簡易	気持ち	経済
意図	志向	内心	帰宅
飲料	飲み物	喉	反省
羽	翼	鳥	返還
…	…	…	…

## 3 意味による関連性評価法

概念Xに対して、2つの概念Aおよび概念Bのどちらがより概念Xと関連があるものであるかを評価する方法として、概念Xと概念A、概念Xと概念Bの関連性をそれぞれ数値で表し、その大小関係で関連の深さを判断する方法が考えられる。この関連の深さを表す数値を関連度と呼ぶ。

本章では、概念ベースを用いて意味的な関連性を判断する方法として意味関連度計算方式を用いた関連性評価法を考える。

### 3.1 意味関連度計算方式

二つの概念Aと概念Bの関連度は、概念A、Bをそれぞれ2次属性まで展開し、概念A、Bの持つ最大900語の属性のうち単語として一致する属性の一致度を評価することにより関連度を算出する。これを意味関連度と呼ぶ。意味関連度 $MR(A, B)$ は、概念A、概念Bそれぞれの2次属性列を一致度の和が最大になるように対応を決め、関連度を求める。

### 3.2 評価実験

評価尺度を用いて、意味関連度計算方式の評価実験を行う。

評価尺度(MX-MA, MB, MC)を(MX-MA, MB), (MX-MB, MC), (MX-MA, MC)の3つに分ける。この形式の評価尺度を評価尺度(MX- $M_a, M_b$ )と表す。この評価尺度(MX- $M_a, M_b$ )に対して意味関連度計算方式を適用し、 $M_a, M_b$ のどちらがMXとより関連が深いかを求め、 $M_a$ の方がMXと関連が深いという結果になったものを正解とし、最終結果は(MX- $M_a, M_b$ ), (MX-MB, MC), (MX- $M_a, M_c$ )においてそれぞれMA, MB,  $M_a$ がMXとより関連が深い、つまりMA, MB, MCの順でMXと関連が深いという結果になったものの割合を出す。

この評価実験の結果は評価尺度(MX-MA, MB, MC)590組に対し、正答率81.5%であった。

## 4 表記による関連性評価法

概念の表記から関連度を求める方法として、1文字でも意味を持っている漢字を用いて関連度を求める。

### 4.1 漢字関連度計算方式

漢字関連度計算方式では、次の2つの方法を用いて一致度を算出し漢字関連度KRを求める。

1) 概念Aと概念Bの表記で同じ漢字の文字数から一致

度 $S$ を算出する.

$$S = \left( \frac{Sa}{n} + \frac{Sb}{m} \right) / 2 \quad (1)$$

ここで,  $Sa, Sb$ をそれぞれ概念 $A, B$ の表記のうちもう一方の概念と一致した漢字数,  $n, m$ をそれぞれ概念 $A, B$ の漢字文字数とする.

- 2) 概念の表記に用いられている漢字のうち対象概念の漢字と熟語となる関係にある漢字に対して一致度を与える. 概念 $A$ の表記に用いられている漢字を $(Ak_1, Ak_2, \dots, Ak_n)$  ( $n$ は表記に用いられている漢字文字数)とし, 同様に概念 $B$ も $(Bk_1, Bk_2, \dots, Bk_m)$ と定義する. このとき,  $Ak_i$ と $Bk_j$ が熟語の中に存在する場合に式(2)によって一致度を与える.

$$N = \left( \frac{Na}{n} + \frac{Nb}{m} \right) / 2 \quad (2)$$

ここで,  $Na, Nb$ をそれぞれ概念 $A, B$ の表記のうちもう一方の概念と熟語になるものの漢字数とする.

上の2つの一致度にそれぞれ重み $Sw, Nw$ を掛けて漢字関連度 $KR$ を求める.

$$KR = \frac{S*Sw + N*Nw}{Sw + Nw} \quad (3)$$

#### 4.2 漢字関連度のパラメータ決定

ここで, 漢字関連度計算方式に用いている重み $Sw, Nw$ を決定するための評価実験を行う.

評価実験では,  $Sw$ を1.0に固定し,  $Nw$ を0.0から2.0まで0.05間隔で変化させ, 評価尺度の正解率が最大になる $Nw$ を求めた.

この実験の結果から,  $Nw$ が0.05から0.30の範囲にある場合に漢字関連度計算方式の正解率が29.2%がもっともよくなることがわかった. そこで, 本稿では漢字関連度計算方式に用いる重みを $Sw=1.0, Nw=0.3$ とする.

### 5 意味漢字関連度計算方式

#### 5.1 意味関連度と漢字関連度の合成

本章で提案する関連性評価法では, 従来から関連性評価法として用いている意味関連度計算方式を改良することを目的とし, 概念の意味だけでなく, 表記からも関連性を判断をする. 提案手法である意味漢字関連度計算方式では, 意味関連度計算方式, および, 漢字関連度計算方式の2つの関連度を合成して利用する.

意味漢字関連度計算方式では, 式4で表すように, 漢字関連度 $KR$ に重み $Kw$ を付与し, 意味関連度に足しあわせ意味漢字関連度を求める.

$$MKR(A,B) = \frac{MR(A,B) + Kw \cdot KR(A,B)}{1 + Kw} \quad (4)$$

#### 5.2 意味漢字関連度計算方式の評価実験

意味漢字関連度計算方式の評価実験を行った.

この評価実験では, 漢字関連度 $KR$ に付与する重み $Kw$ を決定し, 意味漢字関連度計算方式の精度を確かめた.

評価方法は, 評価尺度に対し, 重み $Kw$ を0.0から2.0まで0.05間隔で変化させ, 評価尺度の正解率が最大になる $Kw$ を求めた. 実験の結果は図1のようになり,  $Kw=0.25$ の場合に意味漢字関連度計算方式の精度が最大の87.1%になることがわかった.

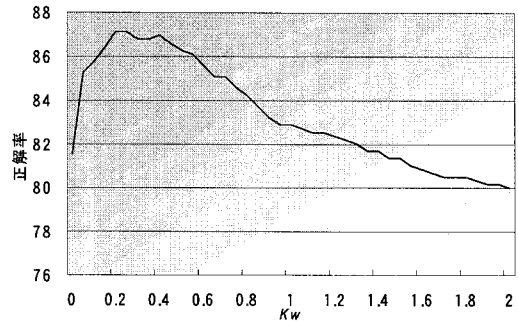


図1. 意味漢字関連度計算の重み $Kw$ による評価結果

#### 5.3 意味漢字関連度計算方式の考察

意味関連度 $MR$ と重みを付与した漢字関連度 $KR$ を合成することで, 意味関連度計算方式のみで概念間の関連性を評価するよりも, より高い精度で関連性の評価を行うことができた. また, 4.2で決定した重み $Kw$ は, 重みを決定する場合で用いたデータ以外の評価データを用いても, 概ね最もよい評価になることから,  $Kw=0.25$ という値は意味関連度計算方式の誤差を漢字関連度計算方式が補正するための値として適切であると言える.

### 6 おわりに

本稿では, コンピュータに常識的な感覚に近い判断をさせるための連想機能の基盤となる, 語と語の関連性の強さを判断させるための, 概念間の関連性評価法に関して, 従来から用いてきた意味属性による関連性評価法である意味関連度計算方式より精度の高い評価を行う手法を提案した.

本研究では意味関連度 $MR$ , 漢字関連度 $KR$ を合成して利用した. この手法では, 意味関連度計算方式より約6%精度が向上した.

#### 謝辞

本研究は文部科学省から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] 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, pp39-54, 2001
- [2] 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, 1997