知能と複雑系　123－17
（2001. 1. 11）

意味属性と漢字属性を用いた概念間の関連性評価法

東村　貴裕　　小島　一秀　　渡部　広一　　河岡　司

〒 610－ 0394　京都府京田辺市多々羅都谷 1-3
同志社大学工学研究科　知識情報処理研究室
Tel：0774-65-6429
E-mail：{ thigashi , kkojima , watabe , kawaoka } @indy.doshisha.ac.jp

あらまし

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と概念の関連性を利用した連想機能と考えられる．本稿では，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い，概念の意味属性から概念間の関連性を評価する方法として，意味関連度計算方式を利用した関連性評価法を提案し，その問題点を解決し，関連性評価法を改良する手段として，語の意味分類，概念の表記から関連性を判断するシソーラス，漢字関連度計算方式を用いている．また，これらを連携利用することで関連性評価法の改良を試みた．

キーワード　関連度，概念ベース，意味属性，漢字，シソーラス

Evaluation Method for Relation between Concepts with Attributes of the Concepts, Kanji Characters and Thesaurus

HIGASHIMURA Takahiro, KOJIMA Kazuhide, WATABE Hirokazu, KAWAOKA Tsukasa

Department of Knowledge Engineering and Computer Sciences
Graduate School of Engineering , Doshisha University
1 － 3 Miyakodani , Tatara , Kyotanabe , Kyoto , 610 － 0394 , Japan
Phone：0774-65-6429
E-mail：{ thigashi , kkojima , watabe , kawaoka } @indy.doshisha.ac.jp

Abstract

Intelligent mechanism is the technology which becomes the key of the future information processing system, and its kernel is the association function with Concept-Base and the relation between two concepts.

In this paper, the purpose is to make a computer which judges relation between two concepts like a human being ordinarily. The method to evaluate the relation between two concepts using attributes of the concepts, named MDA, has already been proposed. MDA method, however, is not good enough comparing with human's judgements. Therefore, in this paper, new method is proposed which is constructed by three different methods. One of those is MDA method. The second is the method using Kanji characters which express the part of meanings of the concept. And the third is the method using the thesaurus which represents the concept categories. By experimental results, it is shown that the combination of those three methods can evaluate the relation between two concepts like a human being.

key words　degree of association, Concept-Base, attributes of the concepts, Kanji characters, thesaurus

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．すなわち，ある単語から概念を想起し，さらに，その概念に関係のある様々な概念を連想できる能力が重要な役割を果たしていると考えられる．

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，語概念間の関連性を評価する連想メカニズムの基盤となる概念ベースの構造，すなわち，語とその意味を表す属性の集合の構成とそれを用いた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簡単に説明をする．

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，そのような単語の集合を概念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．例えば，概念"飛行機"を2次属性まで表現すると図1のようになる．

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559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
…	…	…	…

本研究では，この評価尺度を用いて関連性評価法の評価を行う．

3 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．

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，

概念	属性							
	推進	噴射	翼	航空機	…	重力	力	← 1次属性
飛行機	進める	霧	鳥類	飛行船		物体	物体	
	飛行機	噴出	飛行機	飛行機		地球	能力	
	汽船	燃料	翼	気球		重さ	働き	
	前	油	左右	飛行		自転	作用	2次属性
	⋮	⋮	⋮	⋮		⋮	⋮	
	高める	急激	変形	翼		時間	実力	
	対象	微笑	地層	出航		水平線	仕事	

図1. "飛行機"の属性

Bをそれぞれ2次属性まで展開し，概念A，Bの持つ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*Rm*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．(図2)

概念	二次属性				
自動車	**輪**	**巡る**	**車輪**	**進む**	機械
	車	**回転**	**自動車**	前進	歯車
	歯止め	**商品**	輪	走る	軸
	熱演	**展開**	車	始まり	移動
	車輪	**図形**	軸	行く	機械

概念	二次属性				
自転車	**輪**	**巡る**	走る	動かす	年輩
	車	**回転**	**自動車**	早い	年
	歯止め	**商品**	当たり屋	走る	年頃
	熱演	**展開**	**車輪**	移動	年回り
	車輪	**図形**	動く	**進む**	十代

一致度： 5/5=1　5/5=1　2/5=0.4　1/5=0.2　0/5=0

関連度*Rm*：(1+1+0.4+0.2+0)/5=0.52

図2. 意味関連度計算方式

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つにわける．この形式の評価尺度を評価尺度（MX-Ma,Mb）と表す．この評価尺度（MX-Ma,Mb）に対して意味関連度計算方式を適用し，Ma，Mbのどちらがと関連が深いかを求め，Maの方がMXと関連が深いという結果になったものを正答とし，最終結果は（MX-MA,MB），（MX-MB,MC），（MX-MA,MC）においてそれぞれMA，MB，MAがMXとより関連が深い，つまりMA，MB，MCの順でMXと関連が深いとなったものの割合で出す．

この評価実験の結果は評価尺度（MX-MA,MB,MC）559組に対し，正答率83.36%(466組)であった．表2に誤りであったもの一部を示す．この表のうち，斜体で表している概念の結果が誤っていた．例えば，（演技-***芝居***，***俳優***，灯油）のうち”俳優”が”芝居”より”演技”との関連が深いという結果になった．

表2. 意味関連度の評価実験結果

MX	MA	MB	MC
演技	***芝居***	***俳優***	灯油
花	***花弁***	***花瓶***	弁別
観客	***聴衆***	***公演***	火
議員	***代議士***	***選挙***	解析
式典	式	***入学***	***突起***

3.3 評価実験の考察

表2を見ると，”演技-芝居”，”観客-聴衆”，”議員-代議士”のように「同義・類義」の関係にある概念が「関連がある」概念より関連が薄いとなることが多々あることがわかった．これは意味関連度を算出する元になっている概念ベースが辞書等から作成されたものであり，「同義・類義」に関して概念が必ずしも同じような属性を持っているとはいえないことによるものと思われる．例として”議員”，”代議士”の1次属性の一部を表3に示す．この中で一致した1次属性は30語中5語であった．

また，概念ベースは機械的に作成されたものであり，人間の感覚では必要な属性が抜け落ちていたり，あきらかにおかしな属性が雑音として含まれていることによる関連度の誤差により，関連の深さを正しく評価できないことがあると考えられる．

表3. ”議員”，”代議士”の1次属性

概念	1次属性(30語)						
議員	***議員***	**人**	議決	国会	***選挙***	議会	…
代議士	***議員***	**人**	***選挙***	衆議	国政	国民	…

4 関連性評価法の改良

3.3節で述べた意味関連度計算方式の問題点を解消することを考える．

「同義・類義」に関する問題はシソーラスを用い，概念ベースの雑音による関連度の誤差については概念の意味から関連度を求めるのに対し，概念の表記から関連度を求めそれらを複合的に利用することにより解決することを試みた．

4.1 シソーラスを用いた関連性評価法

意味関連度計算方式では「同義・類義」の関係にある概念より「関連がある」概念の方が関連が深いという結果が出る場合がある．これを解決するために，シソーラスを利用する．本研究ではシソーラスにより同義・類義を判断するために親子関係，兄弟関係を利用することとする．

評価尺度（MX-Ma,Mb）の（MX-Ma）または(MX-Mb)から，シソーラスによりどちらか一方に対し親子または兄弟関係が得られた場合，その組合わせが他方より関連が深いとする．

例えば，（乗り物-自動車,馬）の場合（乗り物-自動車）は親子関係，（乗り物-馬）は親子，兄弟関係では無いので，”自動車”の方が”乗り物”と関連が深いと判断する．

シソーラスを用いた関連性判断では，明らかな関係の違い，つまり，（MX-Ma），（MX-Mb）の一方に親子，兄弟関係があり，他方にはその関係が無い場合でないと関連性を判断できないので，評価尺度をすべて判断できるとは限らない．

4.2 漢字関連度計算方式

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．

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Rk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．

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

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の中で隣接す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,… 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,… Bk_m）と定義する．このとき，Ak_iとBk_jが熟語のなかで隣接する場合に一致度を与える．

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

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Rkを求める．

$$Rk = \frac{S*Sw+N*Nw}{Sw+Nw} \qquad (3)$$

本研究では，Sw=1.0，Nw=0.5とする．

4.3 評価実験

シソーラス，漢字関連度計算方式について，評価尺度（MX-Ma,Mb）1677組を用い評価実験を行った．評価実験の結果は，図3のようになった．また，表4に評価結果の一部を示す．シソーラスの結果のうち，斜体の概念はシソーラスに存在しなかった．

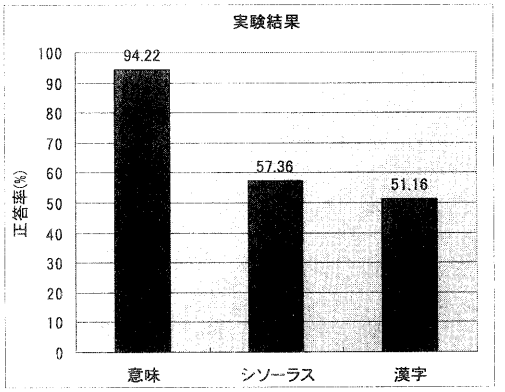

図3. 意味関連度，シソーラス，漢字関連度の評価実験結果

表4. 漢字関連度，シソーラスの評価実験結果（一部）

漢字

X	a	b	評価	XA	XB
ご飯	飯	米	○	0.5	0.17
安易	簡易	気持ち	○	0.58	0.06
意図	志向	内心	×	0.17	0.17
羽	翼	鳥	×	0	0

シソーラス(誤り)

X	a	b	評価	xa	xb
積	***掛け算***	和	×	—	兄弟
提携	連携	協賛	×	—	兄弟
応対	応答	接客	×	—	親子

この結果から，シソーラス，漢字関連度共に意味関連度による関連性評価法より精度が悪いとわかった．その理由として，シソーラスでは関係の違いがとれない場合が多いこと，漢字関連度では計算対象となるのが表記の漢字のみであるので，計算対象が少なく，同じ漢字が含まれていたら関連度が高く，同じ漢字がなければ関連度0が頻出することがわかった．これらのことから，シソーラス，漢字関連度による評価は評価対象を限定する事とする．評価対象は，シソーラスでは関係の違いのとれるもの，漢字関連度ではMXに含まれている漢字をMaもしくはMbの一方のみが持つ場合に限る．

評価対象を限定することによる評価実験の結果，シソーラスでは972組，漢字関連度では

868組が評価対象となり，正答率は図4となった．

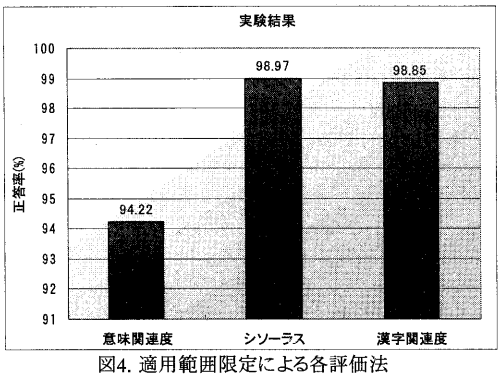

図4. 適用範囲限定による各評価法の評価実験結果

4.4 評価実験の考察

シソーラス，漢字関連度を用いた関連性評価法では，評価尺度全体に対してはシソーラス，漢字関連度共に意味関連度より評価値が悪くなることがわかった．しかし，評価尺度に対して適用範囲を限定することにより99%程度の正答率を得ることが出来た．このことから，シソーラス，漢字関連度をそれぞれの適用範囲で用い，意味関連度と連携して利用することで評価尺度全体に対して評価が可能となり，かつ，評価値を上げられるのではないかと考えられる．

4.5 各評価法の連携利用

前節で述べたように，シソーラス，漢字関連度による評価は，適用範囲を限定しないと評価値が極端に悪くなることがわかった．そこで，3つの評価方法を連携して利用する．

連携利用の仕方としては，

Ⅰ）ある評価方法で適用できないものについては他の評価方法で判断する．

Ⅱ）各評価方法で同じ結果となるものは，正答である可能性が高いので，適用範囲内で他の評価方法と同じ結果となれば，その結果で判断する．組み合わせとしては次の4つが考えられる．

1. 意味・漢字・シソーラス
2. 意味・漢字
3. 意味・シソーラス
4. 漢字・シソーラス

Ⅲ）意味関連度計算方式は機械的に作った概念ベースを用いているので，関連度に誤差が含まれていると考えられる．そこで，（MX-Ma）と（MX-Mb）の意味関連度の差がd未満であり，漢字関連度で（MX-Ma）と（MX-Mb）に差がある場合，漢字関連度の結果で関連性を判断する．本研究では，d=0.03とする．

これらの連携方法をどのように組み合わせるかについて評価尺度（MX-Ma,Mb）を用いて予備評価実験を行ったところ，図5のような結果を得た．この結果から，評価値の高い順に適用していくことが考えられる．

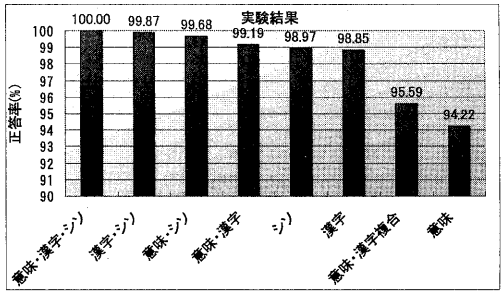

図5.評価法の組み合わせによる予備評価実験結果

1. 意味・漢字・シソーラス同順位
2. 漢字・シソーラス同順位
3. 意味・シソーラス同順位
4. 意味・漢字同順位
5. シソーラス（適用範囲限定）
6. 漢字関連度（適用範囲限定）
7. 意味・漢字複合（d=0.03）

ここで，上位5つの評価方法に注目してみると，表5から上位3つの評価方法は用いても用いなくても同じであることから，上位3つの評価方法は用いる必要がないことがわかる．

表5. 連携利用の比較

シソーラス	漢字	意味	1～5	4, 5
○	○	○	○	○
○	○	×	○	○
○	×	○	○	○
○	×	×	×	×
△	○	○	○	○
△	○	×	△	△
△	×	○	△	△
△	×	×	×	×
×	○	○	○	○
×	○	×	×	×
×	×	○	×	×
×	×	×	×	×

○：適用でき，正答である
×：適用でき，誤りである
△：適用できない

そこで，図6のように各評価方法を利用し関連性判断を行う．

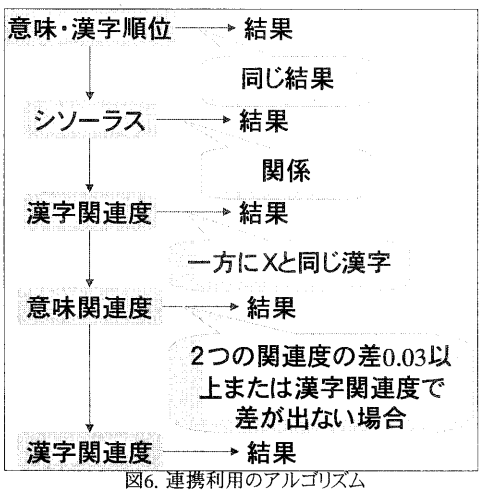

図6. 連携利用のアルゴリズム

4.6 連携利用の評価実験

4.5節の連携利用による関連性評価法について評価尺度（MX-MA,MB,MC）を用いて評価実験を行った．その結果を図7に，具体例の一部を表6に示す．

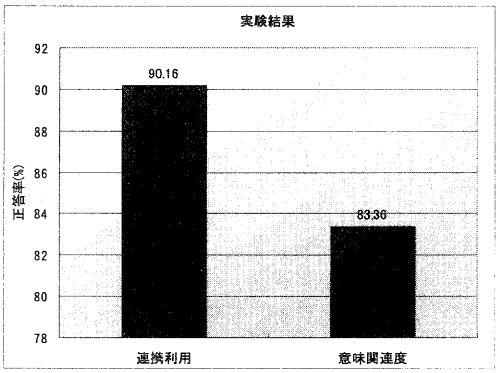

図7. 連携利用の評価実験結果

表6. 連携利用の評価実験結果（一部）

X	A	B	C	評価	xab	xbc	xac	ab	bc	ac
演技	芝居	俳優	灯油	×	3	3	3	×	○	○
羽	翼	鳥	返還	○	1	3	1	○	○	○
指揮	指導	引率	風	○	2	3	0	○	○	○
四角	長方形	三角	経済	×	0	0	0	×	○	○

0:意味漢字同順　1:シソーラス
2:漢字(限定)　3:意味・漢字複合

4.7 連携利用の考察

評価実験の結果から，シソーラス，漢字関連度を適用範囲を限定して意味関連度と連携利用することにより，意味関連度のみで関連性評価をおこなうより正答率が高くなることがわかった．しかし，表6からもわかるように，「同義・類義」の関係で救えていないものがあることもわかった．

5 おわりに

本稿では，意味関連度計算方式による関連性評価法を改良する手段として，語の意味分類，概念の表記から関連性を判断する方法として，シソーラス，漢字関連度計算方式を用いることとしたが，シソーラス，漢字関連度単体では意味関連度による関連性評価法より悪い結果となる．そこで3つの評価法を連携利用することにより評価値を上げることができた．

しかし，「同義・類義」の関係をシソーラスで救えていないものがあり，"帽子"，"梯子"の"子"のように多義である漢字による誤りなどの問題も残っている．

これらから，今後の研究課題として，同義語辞書の活用，漢字関連度計算方式のアルゴリズムの改良があげられる．また，連携利用に関しても，評価実験の結果を詳しく分析することで適用方法の最適化を行っていく．

謝辞

本研究は文部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]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，信学技報 Vol.98, No.499, p.47-54(1999)

[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, p.1272-1283(1997)