

Dynamic Calculation Method of Degree of Association Considering the Common Attributes of Target Concepts

Takayoshi ARAKI*, Noriyuki OKUMURA*, Hirokazu WATABE* and Tsukasa KAWAOKA*

(Received June 19, 2007)

We human beings associate when we are talking with each other. For example, someone says 'cars', we can associate relative words, 'Tire', 'Engine', 'Accident' ...and so on. It is important to give association mechanism to a computer for conversation with human and computer. If computers can evaluate the depth of relationship between concepts, association mechanism is established by present relational words to input-word from various words.

As a method of quantifying depth of relationship with words using concept-base, vector space model and a calculation method of degree of association have been proposed. In these methods, each method quantifies depth of relationship with words by evaluating weights of static attributes of target concepts. However, like multisense words, the concept has various meanings, and attributes are different each meaning generally. So degree of association between concepts depends on selection of attributes. In this paper, it is proposed that calculation method of degree of association by dynamic selection of attributes on target concepts, based on calculation method of degree of association that can evaluate depth of relationship between concepts by various points of view, and it is shown by experiment that proposed method more closely evaluate relationship between concepts like human beings evaluation.

Key words: concept-base, degree of association, degree of match

キーワード: 概念ベース, 関連度, 一致度

比較対象概念の共通属性を重視する動的関連度計算方式

荒木 孝允, 奥村 紀之, 渡部 広一, 河岡 司

1. はじめに

我々人間は, 日常生活において様々な会話の中から必要に応じて情報を取捨選択し, 会話の流れに即して語の意味を適宜解釈し適切な応答を行っている. 人間が自然な会話ができる背景として, 語と語に関する膨大な知識を有し, 語と語の関連性から新たな語を想起する連想能力が備わっていることが挙げられる. コンピュータが人間とより柔軟に常識

的な会話を行うためには連想機能が必須であり, そのためには, 多様な観点において語と語の関係の強さを定量的に評価する方法の開発が必要となる.

語と語の関係の定量的評価手法として広くベクトル空間モデル^{1,2)}が用いられている. これは語の意味の分類体系であるシソーラスを背景とする概念ベースを用い, 語と語の「意味の類似性」を定量的に評価している. しかし, 実際の会話では「自転車と

*Department of Knowledge Engineering and Computer Sciences, Graduate School of Engineering, Doshisha University, Kyotanabe, Telephone:+81-774-65-6944

自動車」のように、類似性という観点での評価のみならず、「車と運転」「赤ちゃんと玩具」「雨と傘」のように広い意味での関連性の評価が必要となる場合が多く存在する。

本稿では、人間とコンピュータの自然な会話の実現に必要な連想機能を念頭に置き、語と語の類似性のみならず幅広い関連の強さを出来るだけ人間の感覚に合った評価のできる新しい関連度計算方式を提案する。従来、コンピュータ会話処理の重要な要素の一つとして、語と語の類似度に関する研究がなされてきた。類似度は語と語が意味的にどの程度似ているかの定量評価を目的としている。例えば、「自転車と自動車」の類似度と「自動車と本」の類似度を定量化し、人間の感覚的判断と同じように「自転車と自動車」の方が意味的に似ていると判定する。一般的に、類似度は、会話において知らない言葉が出てきた時にそれを知っている類義語に置き換えて会話の意味を理解する時などに役立つ。

ベクトル空間モデルは、シソーラスのノードを語空間を形成する一次独立な基底ベクトルと見なし、語の意味（概念）をこれらの語の集合（属性）で表現した概念ベースに基づき2つの概念ベクトルの内積を類似度として計算している。自転車と自動車は、シソーラス上でどちらも「車」ノードに属する語であることから共通属性も多く適切なベクトル類似度を得ることが出来る。しかし、「車と運転」はシソーラス上では殆ど関係のないノードに属しており両者に共通する属性は少なくなり人間が関係が深いと判断出来るベクトル類似度を得ることができない。本研究で対象としている関連度計算方式は、語の意味をベクトルとして定義するのではなく、その語に対して容易に連想可能な語の集合（重み付き属性集合）として定義した概念ベースに基づき、対象となる語（概念）の重み付き属性集合の総合的な一致度を求め、これを関連度としている。この方式では、語の意味（属性集合）をシソーラスにとられることなく単に関連の深い語の集合として容易に定義することができ、かつ、語と語の関連度を広い意味で評価することが可能となる。関連度は類義語を見つけるだけでなくコンピュータ会話

において「雨が降っていますよ」と言う表現に対し「雨」と関連度の高い「傘、濡れる、天気予報、・・・」などの語を使い「それでは傘を持って行きます」というような人間らしい応答を作り出すのに役立つ。また、既発表の関連度計算方式³⁾が語（概念）に固定された属性集合により語と語の関連度を求めているのに対し、本論文で提案する方式は、人間が事物の比較を行う際に両者の類似点に着目して評価を行うことに習い、比較対象語が決まった時点で、対象語を特徴づける多くの属性語の中から動的に適切な属性を抽出する方式であり、より人間の感覚にあった関連度評価が得られると考えられる。提案する方式は、概念ベースを構築する個々の概念は出来るだけ多様な観点から多くの属性で定義しておく、比較対象語が決まった時点でそれらの属性の中から対象語に共通あるいは類似する属性を優先的に選択し、即時的にそれらの属性の重みを再割付けし関連度を計算する新しい方式である

2. 概念ベース

2.1 概念ベースの定義と構成

本稿で対象とする概念ベース^{4,5)}は、概念を基底ベクトルの集合として定義するのではなく、概念(A)を、その概念を意味的に特徴づける概念や、その概念から容易に連想できる語（概念）の集合として定義する。概念を特徴づける概念を、概念に対する属性 (a_1, a_2, \dots, a_n) と呼ぶ。また、各属性にはそれら属性がどの程度概念を特徴づけているかを示す重み (w_1, w_2, \dots, w_n) を付与している（式1）。また、各属性に対する重みは、総和が1.0となるよう正規化している。

$$A = \{(a_1, w_1), (a_2, w_2), \dots, (a_n, w_n)\} \quad (1)$$

各概念を特徴づける属性 (a_1, a_2, \dots, a_n) もまた、概念ベースにおいて概念として定義されているため、 (a_1, a_2, \dots, a_n) もまた概念である。従って、属性の属性を二次属性、二次属性の属性を三次属性、といったように、無限次元の属性連鎖集合によって概念を定義している。Fig.1に概念「雪」を三次属性まで展

開した例を示す。また、これ以降、語の意味（概念）とは、 n 次の属性連鎖集合によって定義された属性空間における重み付き属性のことを示し、概念表記とは、語の意味を識別するラベル（例えば車や飛行機など）を示す。

2.2 概念ベースの構成法

現在、概念ベースは国語辞書より機械的に構築したもの（初期概念ベース^{6,7)}）に加え、新聞記事から概念と属性を取得し概念ベースに登録することにより、約9万語の概念で構成されている。

各概念表記に付与されている属性は、概念ベースに概念表記として登録されている語で構成されるため、各属性を一つの概念表記としてみなした場合、さらにそれを表す属性を導くことができる (Fig.1)。

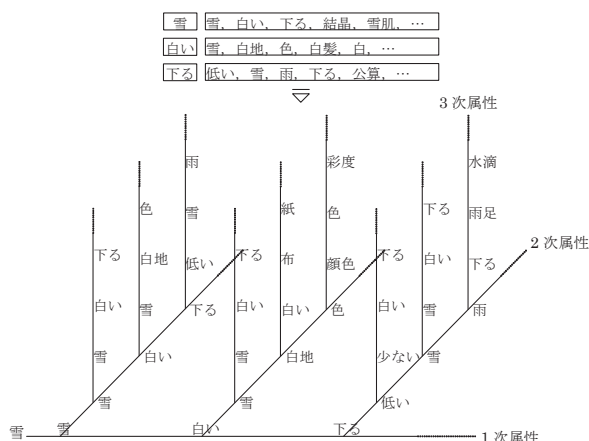


Fig. 1. Concept-base.

このように、概念ベースにより、概念 n 次の属性連鎖集合として定義する。 n 次の属性集合を n 次属性と呼ぶ。なお、次節に述べる関連度計算方式では、属性連鎖により概念表記を 2 次属性まで展開し演算を行う。以下「1 次属性」を単に「属性」とも表記する。

本研究で対象とする概念ベースは、ベクトル空間モデルによる類似性評価を前提としておらず、概念の属性をシソーラス等で意味的に圧縮することなく、そのまま使用している点に特徴がある。これは、意味的に圧縮を行うことにより概念の持つ細かい意味や定義が欠落する問題や、適切に圧縮すること

が困難であることが挙げられる。また、類似性評価に加え、語と語の関連性を同時に評価することを目的とするため、概念から容易に連想し得る語を属性としている。この結果、類似語だけではなく、一般的に関連のある連想語を出力することが可能となる。また、各概念に付与される属性に対し、その重みを以下の式で付与している⁸⁾。

$$w(A, a_i) = r_{A, a_i} \times \sqrt{idf_{cb}} \tag{2}$$

$$idf_{cb} = \log_2 \frac{V_{ALL}}{V_{a_i}} \tag{3}$$

r_{A, a_i} は概念 A に対する属性 a_i の属性信頼度⁵⁾の値であり、 idf_{cb} は概念ベースを属性連鎖により 3 次属性まで展開した空間を仮想文書空間とみなし算出した idf 値であり、 V_{ALL} は全概念数、 V_{a_i} は概念 a_i を 3 次属性空間内に持つ概念の総数である。

3. 概念間の関連度計算方式

3.1 意味関連度計算方式

概念ベースを利用した概念と概念の間にある関連の強さを定量的に評価する手法として、ベクトル空間モデルが広く用いられている。しかし、本稿では、概念を定義する属性集合の重みを含めた一致度を基本とした関連度計算方式³⁾を利用し、概念間の関連性評価を行っている。本稿で使用する概念ベースは、柔軟な連想に役立たせることを前提としており、ベクトル空間モデルに代表される語間の類似度評価では困難となる語と語の関連度の評価(ex.赤ちゃんとおもちゃ)を行えることを目的としている。そのため、意味関連度計算方式を基盤とし、人間の判断により適合する拡張方式を提案する。本節では意味関連度計算方式に用いる重み比率付き一致度と、一致度より算出される意味関連度の定義について述べる。

3.1.1 重み比率付き一致度

任意の概念 A, B について、それぞれ一次属性を a_i, b_j とし、対応する重みを u_i, v_j とする。また、概念 A, B の属性数を L 個、 M 個 ($L < M$) とする。

$$A = \{(a_i, u_i) | i = 1 \sim L\}$$

$$B = \{(b_j, v_j) | j = 1 \sim M\}$$

このとき、概念 A 、 B の重み比率付き一致度 $MatchWR(A, B)$ を以下の式で定義する。

$$MatchWR(A, B) = \sum_{a_i=b_j} \min(u_i, v_j) \quad (4)$$

$$\min(\alpha, \beta) = \begin{cases} \alpha(\beta > \alpha) \\ \beta(\alpha > \beta) \end{cases} \quad (5)$$

概念 A 、 B の属性 a_i 、 b_j に対し、 $a_i = b_j$ (概念 A 、 B に共通する属性がある) となる属性があった場合、共通する属性の重みの共通部分、つまり、重みの小さい方だけ有効に一致すると考え、このように一致度を定義する。定義から明らかな様に両概念の属性と重みが完全に一致する場合に一致度は1.0となる。

3.1.2 意味関連度

概念 A と概念 B の関連度 $MR(A, B)$ は、

1. 概念 A と概念 B の属性を重み順に上位 t 個抽出する。
2. 属性の少ない方の概念を A とし、概念 A の属性を基準とする。

$$A = \{(a_1, w_1), (a_2, w_2), \dots, (a_L, w_L)\}$$

3. 属性どうしが完全に一致する属性 ($a_i = b_j$) は別扱いとする。すなわち、 $\max(u_i, v_j)$ となる a_i または b_j の重みを $\max(u_i, v_j) - \min(u_i, v_j)$ とし、その重みを u_i' あるいは v_j' とする。
4. 完全に一致する属性を除いた概念 A' と概念 B' は、

$$A' = \{(a'_1, u'_1), \dots, (a'_i, u'_i), \dots, (a'_L, u'_L)\}$$

$$B' = \{(b'_1, v'_1), \dots, (b'_i, v'_i), \dots, (b'_M, v'_M)\}$$

のように表せる。ここで、概念 A' と概念 B' には完全に一致する属性が存在しないので、類似する属性について一致度を計算する。一次属性の一致度により全体の一致度の和が最大となるように対応を決め、定まった各属性の一致度とそれぞれの属性の重みを用い、関連度を求める。

5. 完全一致する属性が α 個あったとすると、概念 A と概念 B の意味関連度を以下に示す式で定義する。

$$MR(A, B) = \sum_{i=1}^{\alpha} \{\max(u_i, v_j) - \min(u_i, v_j)\} + \sum_{i=1}^{t+\alpha} MatchWR(a'_i, b'_j) \times \frac{u'_i + v'_j}{2} \times \frac{\min(u'_i, v'_j)}{\max(u'_i, v'_j)} \quad (6)$$

このように関連度を定義しているため、関連度の妥当性は如何に正しい属性が付与され、適切な重みが与えられているかに大きく依存する。すなわち、概念 A と概念 B に対し重要性が高く、かつ、その重みが近似するような場合、関連度が大きくなる。

4. 評価方法

この節では、関連度計算方式の有効性を検証するための評価用データ、並びに評価手法について述べる。

4.1 評価用データ

評価尺度として、4組1セットとなる概念表記群を準備する。これを (X, A, B, C) 評価用データ (Table1) と呼び、任意の基準概念表記 X に対し、同義・類義など最も関連が深いと考えられる概念表記 A 、概念表記 A ほどではないが関連があると思われる概念表記 B 、そして、まったく無関係である概念表記 C によって構成する。このような評価用データを人手によって1780組を用意した。

評価用データの各セットは、複数の人間によって作成され、そのデータを作成者以外の3人の人間に判定させ、全員一致で正しいとされたもののみを利用している。また、評価用データに採用する概念表記全ては、概念ベースで定義されている。

Table 1. Data for evaluation.

X	A	B	C
椅子	腰掛け	机	像
音楽	楽曲	音	電車
海	海洋	塩	車
...

4.2 評価方法

Table1 に示した, 概念 X と概念 A との関連度 $MR(X,A)$, 概念 X と概念 B との関連度 $MR(X,B)$, 概念 X と概念 C との関連度 $MR(X,C)$ が,

$$MR(X, A) - MR(X, B) > AveMR(X, C) \quad (7)$$

$$MR(X, B) - MR(X, C) > AveMR(X, C) \quad (8)$$

$$AveMR(X, C) = \frac{\sum_{i=1}^n MR(X_i, C_i)}{n} \quad (9)$$

の関係を満たすときを正解とする. n は評価セット数である. このとき, 評価用データ全体に対し, 正解したものの比率を順序正解率とする. この評価手法の特徴は, 概念 X と無関連である概念 C との関連度 $MR(X,C)$ は 0.0 になるのが理想であるが, 関連度計算方式の特性により, 基準概念と無関連概念の間に共通する属性が一つでも存在すれば, ある程度の数値が算出される. そこで, 無関連概念との間に算出された値を誤差と見なし, 各概念間の関連度から誤差以上の優位差を得ることができれば, 解釈は成功したと見なす. 以降, 関連度計算方式の評価はこの評価手法を基に行う.

5. 意味関連度計算方式の評価

5.1 評価結果

本節では, 意味関連度計算の評価と考察を行う. 意味関連度計算方式では, 計算対象となる概念の属性について, 重みの高い順に上位 t 個の属性を使用し, 関連度計算を行う. t の値を 1~50 まで変化させ, 順序正解率を調査した. 結果を Fig.2 に示す.

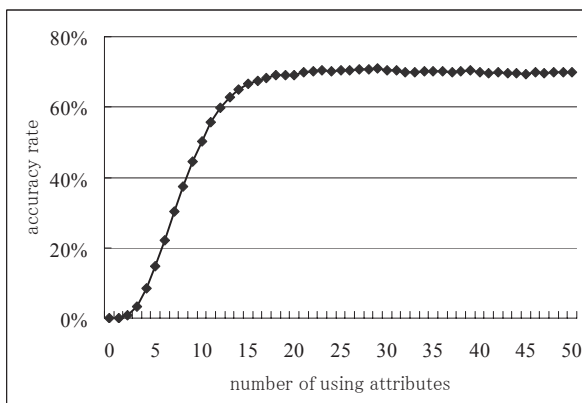


Fig. 2. Evaluation of degree of association.

意味関連度計算方式では, 使用属性数が 30 個以上の時, 順序正解率が最大となり 70.4%である. 以下, 意味関連度計算方式に関する考察を行う.

5.2 意味関連度計算方式の問題点

Fig.2 では, 重み順に上位 30 個以上の属性を関連度計算に使用するとき, 飽和値に近い順序正解率が得られることが示された. 上位 30 個の属性を利用した時 (30 個に満たない場合も含む), 基準概念 X に対し, 同義・類義などの高関連概念 A との関連度 $X-A$, 関係が認められる概念 B との関連度 $X-B$, 並びに無関連概念 C との関連度 $X-C$ の詳細を Table2 に示す.

Table 2. Details of $X-ABC$ evaluation.

	$X-A$	$X-B$	$X-C$
mean	0.2490	0.0474	0.0017
median	0.2150	0.0150	0.0010
minimum	0.0	0.0	0.0
maximum	1.0	0.7860	0.093
standard derivation	0.2102	0.0735	0.0047
dispersion	0.0442	0.0054	0.00002

意味関連度計算方式は比較対象となる概念にかかわらず, 計算対象となる両概念の属性を重み順に固定的に上位 30 個を上限として抽出し, 関連度計算を行う. しかし, 各概念を特徴づける属性は少ないもので 1 個, 多いもので 400 個ほどあり, 非常に多様である. したがって, 重み順に固定数の属性を利用する意味関連度計算方式では, 比較対象によっては重視されるべき属性が捨てられるという問題点がある (式 10).

$$A = \{a_1, a_2, a_3, \dots, a_{30}, \dots, a_L\}$$

$$B = \{b_1, b_2, b_3, \dots, b_{30}, \dots, a_1, \dots, b_M\} \quad (10)$$

(式 10) では, 概念 A と概念 B の関連度を算出するため, 双方の属性を重み順に上位 30 個を抽出し

た場合、 a_i という共通属性が関連度計算に使用されないことを示している。

6. 共通属性を考慮した動的意味関連度計算方式

6.1 計算方式の拡張

前節で述べたように、意味関連度計算方式では概念毎に属性が固定されているため、比較対象概念によっては正しく関連度を算出できない場合があった。そこで、意味関連度計算方式において、関連度計算に使用する属性を、概念を特徴づける属性の重み順に30個抽出し関連度計算を行うのではなく、計算対象となる概念の属性間に共通（完全一致：一致度=1.0）する属性が存在する場合、これらの属性を優先的に関連度計算に利用することを考える。具体的には関連度計算のアルゴリズムを以下のようにする。

1. 概念 A と概念 B の属性に共通する属性を各概念に定義されている多くの属性から抽出する。
2. 概念 A と概念 B の属性に共通する属性が s 個あったとすると、比較対象概念のそれぞれの残った属性から重み順に上位 $t-s$ 個抽出する。
3. 以下、3.1.2 と同様。

このように、重み順に属性を選出する前に共通属性を選出することによって、計算対象となる概念間の共通属性が強調されることとなる。概念ベースに定義される概念の属性は、電子化辞書や電子化新聞から機械的に構築しているため、概念を定義する属性には多義性が存在する。

たとえば、概念「車」と概念「馬」の関連度を算出するとき、それぞれの属性を重み順に上位30個を選出すると、概念「車」の属性は主として「乗り物」に関する属性が取得され、概念「馬」の属性は主として「動物」に関する属性が取得される。このとき、重み順に上位30個の属性間には共通する属性は一つも存在しない。しかし、概念「馬」には140個の属性が付与されており、重み順に上位30個に満たない重みの小さな属性に「乗り物」に関する属性であり、且つ概念「車」の属性と共通する属性が13個存在する(Table3)。したがって、重み順に属性を取得するのではなく、計算対象となる概念間

に共通する属性を優先的に選出することにより、概念が持つ多義性の中から比較対象を前提として人間の感覚に近い属性を抽出し関連度を算出することができると考える。

Table 3. Common attributes between concept *car* and concept *horse*.

common attribute	order of weight at <i>car</i>	order of weight at <i>horse</i>
荷車	26	50
車道	7	60
乗り物	49	107
車	46	106
運搬	52	82
利用	75	116
台	55	65
客	88	115
牛	57	38
長い	96	126
馬	65	43
数える	101	135
乗る	72	91
-	-	-

6.2 評価

意味関連度計算方式に加え、動的意味関連度計算方式の評価を行う。評価手法については4.2に準ずるものとする。また、意味関連度計算と同様、動的関連度計算に使用する属性数を1~50まで変化させ、順序正解率を調査した。結果をFig.3に示す。

共通属性を考慮した関連度計算方式においても、意味関連度計算方式と同様、使用属性数が30個のとき、順序正解率は最高となり、73.2%となった。また、意味関連度計算方式と比べ、2.8%順序正解率が向上した。

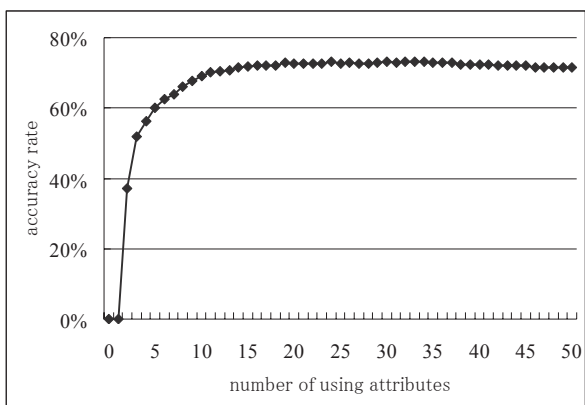


Fig. 3. Evaluation of degree of association considering common attributes.

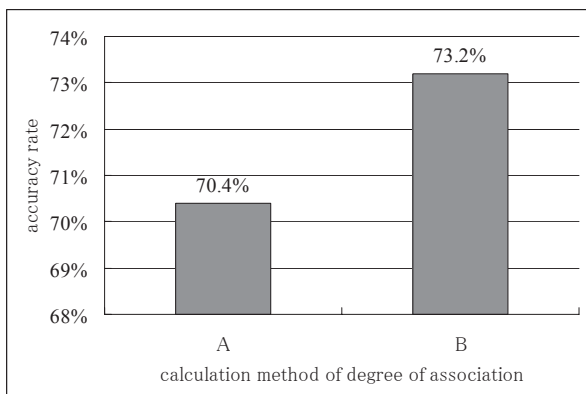


Fig. 4. Comparison with existing method.

Fig.4 において、A は意味関連度計算方式、B は動的意味関連度計算方式を示しており、それぞれ関連度計算に使用する属性数は 30 個である。Table4 に 5.2 と同様順序正解率の詳細を示す。

Table 4. Details of X-ABC evaluation.

	X-A	X-B	X-C
mean	0.3022	0.0653	0.0021
median	0.2456	0.0256	0.0008
minimum	0.0	0.0	0.0
maximum	1.0	1.0	0.1053
standard derivation	0.2572	0.1041	0.006
dispersion	0.0066	0.0108	0.00004

Table2, Table4 から、関連度計算の対象となる概

念に共通する属性を優先的に抽出することにより、高関連、中関連概念の関連度が上昇していることが分かる。また、前節で述べたように、無関連概念との関連度がわずかではあるが上昇しているのは、雑音属性との対応がとられるようになったためと考えられる。また、共通属性を利用することにより、属性間の対応が適切に定まった例として、先述の概念「車」と概念「馬」の全ての属性を重み順に示す。

「車」の属性 : 霊柩車 乳母車 輪状 車引き 新車 車中『車道』 車代 空車 三輪車 停車 轍 下車 乗車 車庫 車軸 自転車 大車輪 乗り上げる 貨車 牛車 人力車 馬車 車馬 駐車『荷車』 乗り合い 両輪 車寄せ 手車 (重み順に上位 30 個) 車輛 空回り 車輪 風車 心棒 積荷 電車 撞球 熱演 列車 車上 蛍雪 追い抜く 輪 自動車『車』 花車 香車『乗り物』 軸 回転『運搬』 乗せる 拾う『台』 推輓『牛』 機械 回す 巡り 渦巻き 降りる 焼き 枠『馬』 運ぶ 大正 貫く 物する 仕掛け 立て『乗る』 紋所 仕組み『利用』 巡る 果たす 道 逆さ 動く 欠く 中心 明治 捻る 通る 翳す 穴『客』 宛てる 移す 中央 停める 止まる 進む 助数詞『長い』 位置 出る 刷る 用いる『数える』 多い 為る

「馬」の属性 : サラブレッド 馬匹 竜馬 馬蹄 馬糞 騏 裸馬 愛馬 名馬 馬主 馬肉 駄馬 駿馬 騎手 馬乗り 奔馬 騎 馬耳東風 乗馬 馬術 悍馬 葛折り 馬上 風馬牛 牛馬 兎 鬣 木馬 騎馬 飼育 (重み順に上位 30 個) 競馬 桂馬 蹄 銀将 家畜 絆 轡『牛』 博労 下馬 鹿 乗り換える『馬』 竹馬 拍車 競走 食肉 聞き流す 草食『荷車』 奇 食用 脚立 足搔き 手綱 肥 乗り換え 王将 伸す『車道』 犬 双六 一匹 駒『台』 不払い 農耕 竜 駆ける 哺乳 取り立て 餞 絵馬 肥やす 念仏 取り替える 調子づく 四肢 尾 肉 将棋『運搬』 運送 毛色 房 走る 歩 勝ち 動物 駅『乗る』 者 同類 揺る 遅い 桜 雨 耕作 頭部 踏み台 支払う 動植物 代金 毛 白い『車』『乗り物』 革 屋 首 左右 取り立てる 耳 足『客』『利用』 下る 有する 開く 文 類い 早い 体 吐く 引く『長い』 同じ 善い 強い 在る 意 大きい 場合 使う『数える』 上

がり行く出来る表す人

『』内に関連度計算方式の拡張によって対応がとれた共通属性を示している。また、このように対応を決定することによって、意味関連度計算方式では $MR(\text{車,馬})=0.025$ だったものが、 $MR(\text{車,馬})=0.063$ となった。

7. 共通・類似属性を考慮した動的意味関連度計算方式

7.1 類似属性の利用

前節で述べたように計算対象となる概念に共通する属性を優先的に利用することにより、順序正解率は向上した。そこで、共通属性優先の考え方をさらに拡張し、類似する属性同士も優先的に対応づける手法を検討する。

本来、関連度計算方式では、概念の持つすべての属性間で一致度を算出し、一致度の総和が最大になる組み合わせを算出すべきであるが、計算量が膨大となるため、関連度のほぼ最大値が得られる重み順に上位 30 個の属性で近似計算していた。しかし、共通属性を優先的に利用する動的関連度計算方式による評価から、重みの低い属性にも有効に活用できる属性が存在することが分かる。そこで、以下のような属性の類似性を考慮した方式を考える。

関連度計算に t 個の属性を使用する場合、

1. まず、概念 A と概念 B の属性に共通する属性を選出する。
2. 概念 A と概念 B の属性に一致度が閾値を満たす属性があれば類似属性として選出する。
3. 概念 A と概念 B の属性に共通する属性が s 個、一致度が閾値を満たす属性が u 個とすると、残った属性から重み順に上位 $t-s-u$ 個抽出する。
4. 以下、3.1.2 と同様。

7.2 類似属性を決定する一致度の閾値

共通属性を優先的に選出する場合においては、表記的に完全に一致する属性を選出することで簡単に対処できた。しかし、類似属性を選出する場合は、一致度の閾値を定め、それ以上の一致度を持った属

性を類似属性として抽出する方法が考えられる。そこで、まず関連度計算に使用する属性数を 30 個に設定し、一致度の閾値と順序正解率の関係を実験により求める。Fig.5 に一致度の閾値を 0~1.0 まで 0.01 刻みで評価した順序正解率を示す。

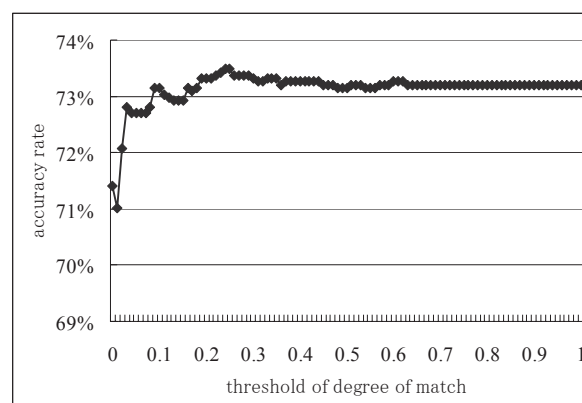


Fig. 5. Investigation for threshold of degree of match.

Fig.5 から一致度の閾値が 0.24 のとき、順序正解率が最大で 73.5%となる。この実験の結果より、類似属性を選出する際の一致度の閾値を 0.24 とする。

7.3 共通・類似属性を考慮した動的関連度計算方式の評価

次に、この方式で関連度計算に使用する属性数を決定するため、Fig.5 から求めた一致度の閾値である 0.24 を用いて、属性数を 1~50 まで変化させ、順序正解率を調査した。結果を Fig.6 に示す。

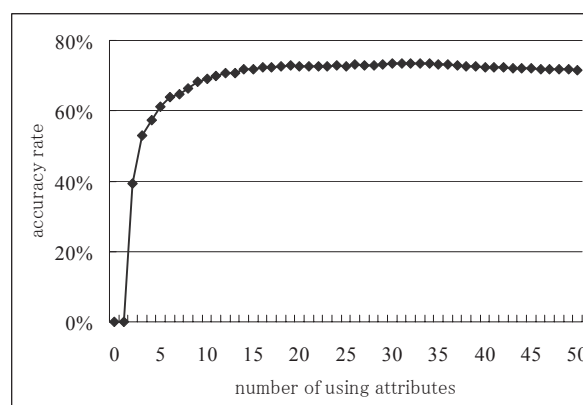


Fig. 6. Evaluation of degree of association considering common and similar attributes.

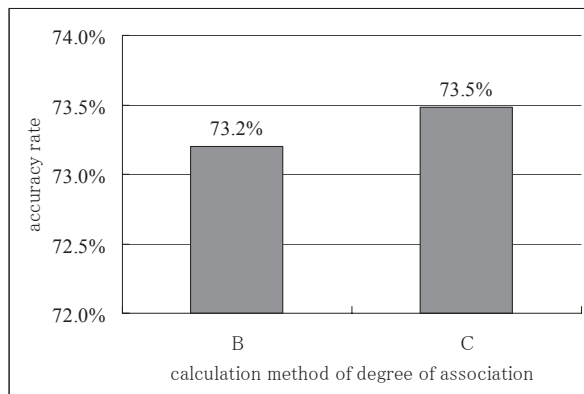


Fig. 7. Comparison of two suggested methods.

Fig.6 から、使用する属性数が 30 個の時、順序正解率が最高となることが分かる。したがって、この方式においても計算に用いる属性数は 30 個が適切であると言える。また、Fig.7 に示したように、共通属性のみを優先的に利用した動的関連度計算方式 (Fig.7 のグラフ B) に比べ 0.3%ではあるが、順序正解率が向上している (Fig.7 のグラフ C)。さらに、評価の詳細を以下に示す。

Table 5. Details of *X-ABC* evaluation.

	<i>X-A</i>	<i>X-B</i>	<i>X-C</i>
mean	0.3018	0.0648	0.0020
median	0.245	0.026	0.001
minimum	0.0	0.0	0.0
maximum	1.0	0.927	0.093
standard derivation	0.2568	0.1018	0.0054
dispersion	0.00659	0.0104	0.00003

目視による判定で明らかに人間の感覚では関係があるとされた概念について、関連度が算出できなかった例として「案内状」「招待」の属性を示す。

「案内状」の属性：案内書 案内状 貴書 入門 催す 手引き 会合 手紙 誘う 通信 予定 来る

「招待」の属性：御呼ばれ 請じる 招請 招致 招来 招聘 招待 招き 乞い 申し入れる 呼び寄せる 客人 無料 来す 食客 稀に 引き出物 客分 芸妓 亭主 招く 渡米 遊興 旅館 迎える 旅人 乗客

合図 木戸 見物 遊女 俳優 正式 旅行 見難い 得意 呼ぶ 寄寓 手真似 訪れる 旗 意向 泊まる 寄せる 客 看板 買う 用意 乗り物 側 乗る 物品 古く 手 自分 人

この例のように、人間の判断では明らかに関係のある概念であり、且つ、一見確からしい属性が付与されている概念であっても、完全に共通する属性が無く、また、類似の閾値を満たす属性も存在しない概念については正しく関連度を求めることができず ($MR(\text{案内状,招待})=0.0$)、このような概念の関連性を正しく評価するためには、概念に適切な属性の追加が必要となる。

7.4 考察

以上の検証結果により、関連度計算方式において、関連度計算に使用する属性を抽出する際、各属性に付与されている重み順に上位 30 個を抽出するよりも、評価時点で概念間に共通する属性と類似する属性を抽出し、残りを重み順に抽出するほうが順序正解率が高くなることが実験により確認できた。Fig.8 において、グラフ A は意味関連度計算方式、グラフ B は共通属性のみを考慮した動的意味関連度計算方式、グラフ C は共通・類似属性を考慮した動的意味関連度計算方式である。

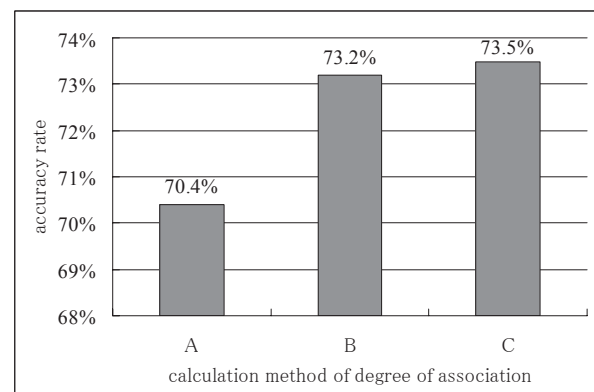


Fig. 8. Evaluation of degree of association.

共通属性のみを優先する動的意味関連度計算方式では従来の意味関連度計算方式と比較しても計算時間に大きな差は見られなかったが、類似属性

を使用する方式では、その抽出に計算時間が 30 倍近くかかり、順序正解率において 0.3% 程度の向上しか得られないことから、実用的には、共通属性のみを使用する動的意味関連度計算方式が適切であると考えられる。

また、順序正解率において約 3% の精度向上が見られたことについて補足する。Fig.9 におけるグラフ a は各概念の属性に対する重みを 1.0、すなわち、重みなしとした概念ベースの順序正解率、グラフ b はグラフ a の属性に対し、種々の重み付けの改良を施した概念ベースの順序正解率を示しており、グラフ c はグラフ b の重みをさらに精練し、本研究で用いた概念ベースの順序正解率を示しており、それぞれ意味関連度計算方式による評価を行ったものである。

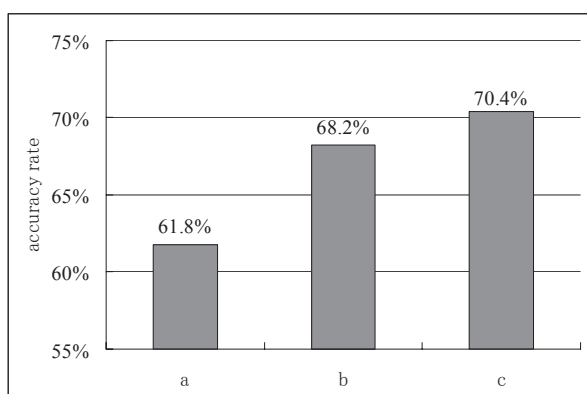


Fig. 9. Supplement concerned precision improvement.

これまで、概念ベースの様々な精練作業を通じ、6.4% の精度向上を図ることができ、さらに精練を施すことにより 2.2% の精度向上ができた。しかし、概念ベースの精練はほぼ限界に来ており、1% の精度向上を図るのは極めて困難であり、大量の作業を伴う状況となっている。したがって、関連度計算方式の拡張により、約 3% の精度向上が見られたことは、十分に評価できるものであると言える。

8. おわりに

人間とコンピュータの自然な会話を実現するにはコンピュータに連想機能を持たせることが必須となる。この連想機能により、会話において知らな

い語を知っている類義語に置き換え文の意味を理解したり、入力文に含まれる語から応答にふさわしい語を連想し、自然な応答文を作り出すことが可能となる。

語と語の関連の強さを定量的に評価するための手法として、ベクトル空間モデルでは目的が類義語の評価に限定される。また、従来の関連度計算方式では、計算に使用する属性が比較対象語に関係なく概念ベースに固定的に定義された属性が用いられていたため評価に柔軟性が欠けていた。これに対し、本稿では、語の意味を定義する属性を比較対象語が決まった時点で動的に選択する、すなわち、比較対象語に応じて語の意味を定義する属性を柔軟に変化させる関連度計算方式を提案した。また、実験により提案方式では、高関連/中間連/無関連の順序正解率を従来の関連度計算方式（属性固定）に比べ約 3% 向上できることが確認できた。なお、順序正解率における 3% の向上は、これまでに行った概念の精練（属性信頼度など）により得られた精度向上に比べても十分に評価出来るものである。

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における研究の一環として行った。

参考文献

- 1) 川島貴広, 石川勉, “言葉の意味の類似性判別に関するシソーラスと概念ベースの性能評価,” 人工知能学会論文誌, Vol. 20, No. 5, pp. 326-336, (2005).
- 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).
- 3) 渡部 広一, 奥村 紀之, 河岡 司, “概念の意味属性と共起情報を用いた関連度計算方式,” 自然言語処理, Vol. 13, No. 1, pp. 53-74, (2006).
- 4) 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-語間の論理関係を用いた属性拡張,” 自然言語処理, Vol. 11, No. 3, pp. 21-38, (2004).
- 5) 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).
- 6) 小島 一秀, 渡部 広一, 河岡 司, “常識判断のための

概念ベース構成法—概念ベースにおける概念属性の確からしさによる概念属性の重み決定法,”信学技報, 電子情報通信学会, AI2001-39, 39-46, (2001).

- 7) 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).
- 8) 奥村 紀之, 渡部 広一, 河岡 司, “電子化新聞を用いた概念ベースの拡張と属性重み付与方式,” 情報処理学会研究報告, 2005-NL-166, 55-62, 情報処理学会, (2005).