

ベクトル表現された概念に対する類似度計算法

Methods for Measuring Semantic Similarity between Vector-presented Words

グエン ベト ハー^{†,1}

石川 勉[†]

笠原 要[‡]

Nguyen Viet Ha

Tsutomu Ishikawa

Kaname Kasahara

[†] 拓殖大学工学部情報工学科

Department of Computer Science, Takushoku University

[‡] NTTコミュニケーション科学研究所

NTT Communication Science Laboratories

あらまし 概念と概念の類似性を判別するため、国語辞書を用いて概念に関する知識を獲得する手法を提案している。各概念は、複数の属性と重みで表現され、概念間の類似度は基本的には属性の共通な度合いに基づいて算出する。本論文では、シソーラスを利用して、属性間の関係を考慮した計算手法を提案する。また、観点を考慮した類似度について考察を行い、この類似度の条件を明らかにする。実験の結果、提案した手法が従来の手法より有効であることを確認した。

1 はじめに

最近、コンピュータ上で柔らかな処理を実現しようとする研究が注目されている。従来の情報処理技術はデータや処理法がきちんと整理された問題を対象にしてきた。しかし、実世界では、問題に関する全てのデータを集めることは困難である。従って、不完全なデータを許容する処理技術が、コンピュータが普及する現在、極めて重要である。

このような処理の実現を目指して、人間の言葉を工学的に表現し、言葉と言葉の意味的な類似性をコンピュータ上で判別する研究が進められてきた。例えば、自然言語処理では、単語を意味的に分類したシソーラスを利用して類似性判別を行う方法が提案されている [1]。この方法では、単語間の関係を状況や文脈から独立して静的に扱っている。

これに対し、我々は日常用いられる数多くの単語（以下、概念と呼ぶ）に対し、観点に応じた柔軟な類似性判別を行える方式について研究してきた [2, 3]。具体的には、国語辞書から容易に獲得可能な知識だけを用いて概念に関する知識ベース（以下、概念ベースと呼ぶ）を構築し、これをベースに判別手法を検討してきた。この概念ベースでは、各概念は、複数の属性と重みのペアで表現され、属性空間におけるベクトルであると見なすことができ、概念ベクトル

を変化することにより観点を考慮した動的な類似度の計算が容易となる。

概念を属性空間で表現する際には、各属性が互いに意味的に独立、すなわち直交していることが望ましい。これは、概念間の類似度が属性ベクトルの内積や距離等で容易に計算できるためである。これに対し、これまで、属性をシソーラス上のカテゴリに変換することで対処してきた [3]。しかし、これらのカテゴリが完全に意味的に独立していないため、十分な直交基底とはいえない。直交な意味空間を構築する方法としては、行列の直交変換や主成分分析が利用されている [4, 5]。例えば、宮原らは辞書から獲得したデータに対して、固有値分解を行い、直交な意味空間を構築している [4]。しかし、これらの方法では構築された空間の基底がデータに依存して変化するという問題がある。また、前述したベクトルの内積等で類似度を求める場合、直交変換された後の空間での類似度は変換前の空間での類似度と同じ値となる。本論文では、シソーラスの情報を利用して属性の非直交性を考慮した類似度計算を提案する。具体的には、属性と属性の関連度合いをシソーラスより求め、これを概念間の類似度計算に反映させる。

また、これまでの研究では、観点を用いたときの類似度の条件について十分な検討がされていない。本論文では、この類似度の必要条件を提示し、条件を満足する計算手法について考察を行う。

以下、2章では概念ベースの構成、3章では従来の

¹連絡先：拓殖大学工学部情報工学科

〒193-0985 東京都八王子市館町 815-1

E-mail: nguyen@cs.takushoku-u.ac.jp

類似度計算手法の問題点と非直交性を考慮する類似度計算手法、4章では観点を考慮した類似度の考察、5章では計算法の評価実験について述べる。

2 概念ベースの構成

概念ベースは国語辞書の語義文を用いて自動的に構築している。具体的には、辞書の見出しを概念とし、その語義文中の自立語を形態素解析で抽出してその概念の属性としている。また、属性の重みは語義文中での属性の出現頻度に基づいて設定している。このように構築された概念は以下のように表現される。

$$\text{概念 } g_i = \{(p_{i1}, q_{i1}), (p_{i2}, q_{i2}), \dots\}. \quad (1)$$

ここで、 p_{ij} は属性であり、 q_{ij} はその属性の重みである。

さらに、シソーラス [6] を利用して、属性をシソーラス上のカテゴリに変換する。カテゴリの総数を N とすると、概念は以下のようなベクトルとして表現される。

$$g_i = \{q_{i1}, q_{i2}, \dots, q_{iN}\}. \quad (2)$$

ここで、獲得されない属性（カテゴリ）に対してその重みを 0 とする。なお、重みの二乗和が 1 になるように ($\sum_{j=1}^N q_{ij}^2 = 1$) 正規化している。図 1 に概念ベースのイメージを示す。

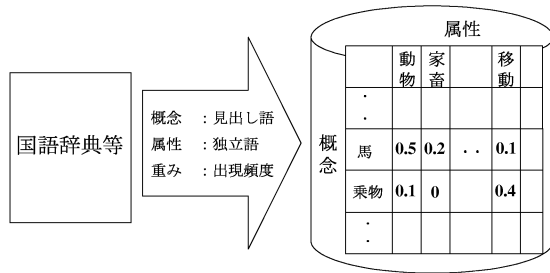


図 1: 概念ベースのイメージ

3 観点を考慮しない場合の類似性

3.1 類似度の条件と属性の直交性

概念間の類似性について考えたとき、人間は、何らかの方法で比較概念の共通な性質を見出し、類似性を判断すると思われる。従って、概念ベースを用いる場合、類似度は属性の共有の度合いに基づいて算出することが自然といえる。

この類似度は、文献 [3] に示したように、以下の条件を満たすことが必要といえる。

$$1) \ 0 \leq R(g_i, g_j) \leq 1.$$

$$2) \ R(g_i, g_j) = R(g_j, g_i).$$

$$3) \ R(g_i, g_i) = 1.$$

$$4) \ R(g_i, g_j) > R(g_i, g_k) \text{ のとき, } g_j \text{ は } g_k \text{ よりも } g_i \text{ に類似している.}$$

ここで、 $R(g_i, g_j)$ は概念 g_i, g_j 間の類似度を表す。

これらの条件を満たすように、これまでの研究では、各属性が直交と仮定して、類似度は概念ベクトルの内積として求めている。

$$R(g_i, g_j) = \sum_{k=1}^N q_{ik} q_{jk}. \quad (3)$$

しかし、各属性はシソーラスのカテゴリであり、これらは木構造で表される関連性をもち、意味的に直交しているとはいえない。従って、同一の属性の重みだけを類似度計算に考慮する上記の方法は適切な類似度を算出することができない場合がある。例えば、概念 g_a には属性“男”と属性“夫”の重み、概念 g_b には、属性“女”と属性“妻”の重みのみが正しい値を持つと仮定する。

$$g_a = \{(\text{“男”}, 0.7), (\text{“夫”}, 0.7)\}.$$

$$g_b = \{(\text{“女”}, 0.7), (\text{“妻”}, 0.7)\}.$$

これらの属性の関係を図 2 に示す。上記の手法（式 (3)）を適用すると、類似度が $R(g_a, g_b) = 0$ になってしまう。しかし、実際“男”と“女”の間や“夫”と“妻”の間にはある程度の関連性が存在するので、 $R(g_a, g_b) > 0$ となるべきである。従って、適切な類似度算出ためには異なる属性の関連性を考慮する必要がある。

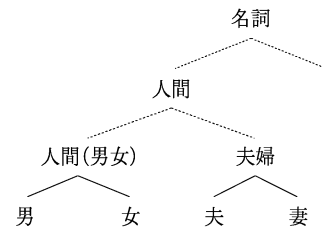


図 2: シソーラスの一部

3.2 上位属性を付加する手法

概念はある属性を持ったとき、その属性のシソーラス上の上位属性に対してある程度の関連性が存在する。これは、下位属性と上位属性との間には is-a 関係が存在するからである。もし、上位属性の重みが 0 となっている場合、下位属性の重みを利用して上位属性に対し重みを付加し、概念ベクトルを修正することができる。この上位属性の付加により属性空間の非直交性がある程度補完されると考える。

前節の問題について考えてみる．上記の手法を適用すると，上位属性“夫婦”や“男女”等に対し正の重みが付加され，類似度が $R(g_a, g_b) > 0$ となる．この場合，上位の属性に対し，どの程度の値（付加値と呼ぶ）を付加するか問題となるが，付加値は下位属性の重みと二つの属性の関係に依存すると考えられる．従って，ここでは，属性の情報量の比を利用して付加値を決定することとする．具体的には，ある属性 j を想定したとき，その上位の属性 k に対する付加値を以下のように求める．

$$q_{ik}^j = q_{ij} \left(\frac{I(k)}{I(j)} \right)^\alpha. \quad (4)$$

ここで， $I(k)$ は属性 k の情報量であり， α は付加の係数である．付加された後の属性の重みは，属性のものと重みと下位属性から算出された付加値の最大値とする．

$$\hat{q}_{ik} = \max \{q_{ik}, q_{ik}^{j_1}, q_{ik}^{j_2}, \dots\}. \quad (5)$$

ここで， $\hat{q}_{ik}$ は属性 k の付加された後の重み， $j_1, j_2, \dots$ はシソーラスにおける属性 k の下位属性を表す．

また，属性の情報量はその属性に属する下位属性の数で算出することとし，以下のように設定する．

$$I(k) = -\log_2 \left(\frac{n_k + 1}{N} \right). \quad (6)$$

ここで， n_k はシソーラスにおける属性 k の下位属性の数である．属性が上位の位置にあるほど下位属性の数が多くなり情報量が小さくなるので，式 (4) の付加値が小さくなる．

上位属性を付加した後，属性空間を完全に直交すると見なして，式 (3) を用いて概念間の類似度を計算する．なお，前節の条件 3) を満足するため，上位属性を付加した後，重みを再度正規化する．

3.3 属性間の類似度を利用する手法

属性の非直交性を補正するもう一つの手法のとして属性と属性との類似度を直接的に概念間の類似度計算に算入させることが考えられる．まず，各属性がある理想的な直交基底をもつ意味空間上でベクトルとして表現されるとする．この場合，各概念はこの意味空間上で属性ベクトルの和で表現されることとなる．

$$g_i = \sum_{j=1}^N q_{ij} c_j. \quad (7)$$

ここで， c_j は上記の空間における属性 j の意味ベクトルを表す．

このとき，概念間の類似度は，属性間の類似度（意味ベクトルの内積）を利用して以下のように求められる．

$$R^*(g_i, g_j) = \sum_{k=1}^N \sum_{l=1}^N q_{ik} q_{jl} c_k c_l. \quad (8)$$

ここで，実際は c_k, c_l を個別に定めることが望ましいが，これは困難なため，その積だけに着目する．具体的には， $c_k c_l$ を属性間の類似度とみなし，これを属性がシソーラス上のカテゴリーであることからシソーラス上の距離で近似して，以下のように定義する．

$$c_k c_l = 1 / (1 + d(k, l))^\beta. \quad (9)$$

ここで， $d(k, l)$ は属性 k と属性 l のシソーラス上の単純な距離であり， β は係数である．二つの属性は，シソーラス上の位置が離れるほど（異なる意味分類）その距離が長くなり，式 (9) で求める類似度が小さくなる．

また，3.1 節の条件 3) を満たすように， $R^*(g_i, g_i) = 1$ になるように，属性の重みを再度正規化する．なお，3.1 節の問題に対して，この手法を適用すると，“男”と“女”，“夫”と“妻”の間の類似性が考慮され， $R^*(g_a, g_b) > 0$ となる．

4 観点を考慮した場合の類似性

4.1 観点を考慮した類似性の特徴

概念と概念の類似度は静的に決定したものではなく，それをみる観点によって動的に変化するものである．観点を考慮したときの類似度とは，観点自体あるいはそれと関連する概念から見た両比較概念の似具合であると考えられる．この類似度について以下の場合を考慮する必要がある．

- a) 観点が両概念と関係する場合：例えば，“犬”と“猫”を“ペット”という観点で比較（この場合の類似度を R_a とする）．
- b) 観点が片方の概念と関係する場合：例えば，“犬”と“猫”を“お化け”という観点で比較（この場合の類似度を R_b とする）．
- c) 観点が両概念と無関係の場合：例えば，“犬”と“猫”を“国会”という観点で比較（この場合の類似度を R_c とする）．
- d) 無観点の場合の類似度：後述するように全ての属性が同じ重みをもつ仮想的な観点を想定し比較（この場合の類似度を R とする）．

これらの類似度の大小関係について考えてみる．

- 1) $R_a > R_b$ ：二つの概念に關係する観点で比較す

る場合の類似度が片方の概念のみに関係する観点の場合の類似度より大きいと考えるのは自然である。これは絶対条件であると考ええる。

- 2) $R_a > R_c$: R_c をどう定義すべきか問題であるが、少なくとも $R_c > R_a$ や $R_c = R_a$ とするのは不自然である。このため、これは絶対条件であると考ええる。
- 3) R_a と R の関係 : 無観点の場合、人間が両概念に関係する何らかの観点、例えば“ペット”又は“動物”等を想起して類似度を考える可能性があるため、その大小関係は不定と考えられる。
- 4) R_b と R_c の関係 : これも前述のように R_c をどう考えるかが問題となり、その大小関係は不定と考えられるが、少なくとも $R_b > R_c$ 又は $R_c > R_b$ を常に真とするのは不自然である。
- 5) $R > R_b$: 前述したように無観点の場合には何らかの関連する観点が想起される。このため、これは絶対条件であると考ええる。
- 6) $R > R_c$: 4) のように R_c をどう考えるかが問題となり、これは絶対条件とはいえないまでも、望ましい条件といえよう。

以上より、類似度計算手法としては絶対条件 1) , 2) , 5) を満足する必要がある。また、条件 4) , 6) を満足することが望ましい。

4.2 観点を考慮した計算モデル

一般的には、観点を考慮した類似度の計算手法としては、比較する概念を観点で変調させ、変調された概念の無観点の類似度を求めるという手法が考えられる。従って、概念を如何に変調するかが重要といえる。

この方式では一般的には、観点 g_v からみた概念 g_i , g_k 間の類似度 $R(g_i, g_k | g_v)$ は以下のように求める。

$$R(g_i, g_k | g_v) = R(g_i^v, g_k^v). \quad (10)$$

ここで、 g_i^v , g_k^v はそれぞれ、 g_i , g_k を g_v で変調した概念である。 g_i^v の重み q_{ij}^v は以下のように求める。

$$q_{ij}^v = q_{ij} M(q_{vj}). \quad (11)$$

ここで、 $M(q_{vj})$ は観点からみて属性 j が重要であるかを決定する変調関数である。また、3.1 節の条件 3) を満たすため、変調された概念に対し、属性の重みの正規化を行う。

これまで、 $M()$ として、観点の属性の重みがあるしきい値を超えれば概念の重みを定数倍する矩形関

数を用いている [3].

$$M(q_{vj}) = \begin{cases} m & : q_{vj} \geq \eta \\ 1 & : q_{vj} < \eta \end{cases}. \quad (12)$$

ここで、 η はしきい値であり、 m は変調倍率である。

この手法は、前節で定義した絶対条件 1) , 2) , 5) を満たしている。なお、理論的には常に $R = R_c$, $R_c > R_b$ となり、条件 4) , 6) を満たしていない。また、この手法では、重要な属性に対し一定な倍率で変調するため、観点中の属性の大小関係を反映することができない。ここでは、変調率を観点の属性の重みに依存する以下の変調関数について考える。

$$M(q_{vj}) = 1 + \gamma q_{vj}. \quad (13)$$

この変調関数も従来の変調関数と同様に、条件 1) , 2) , 5) を満たしているが、条件 4) , 6) を満たしていない。

条件 4) と条件 6) を満足するために、観点に関係ない属性を類似度計算に入れない変調関数を考える。

$$M(q_{vj}) = \begin{cases} 1 + \gamma q_{vj} & : q_{vj} > 0 \\ 0 & : q_{vj} = 0 \end{cases}. \quad (14)$$

この変調関数を用いると、理論的には R_b と R_c が 0 となり、条件 4) , 6) を満たしている。

なお、無観点の場合の類似性は、全ての観点で見たとときの類似性として考えることができる。この“全ての観点”のベクトルとしては、全ての属性が同じ重みを持つ白色的なベクトルと仮定できる。このベクトルで概念を変調すると、どの属性も同じ倍率で変調される。従って、変調された概念に対して正規化を行うと変調前の概念と同じ概念が得られる。よって、この方法で計算する無観点の類似度は、3 章で定義した観点を考慮しない類似度と同一なものとなる。

5 実験

5.1 観点を考慮しない類似度

5.1.1 評価手法

類似度の計算手法の特徴としては

- (1) 類似する概念との間の類似度は全く類似しない概念との間の類似度の差が大きい
- (2) 類似する概念との間の類似度は関連する概念（類似しないが何らかの関連性を持つ）との間の類似度より大きい

が必要といえる [7].

これまで提案してきた計算法が以上の性質を満たすかどうか評価する．具体的には，まず類語辞典を用いてサンプル概念 g_a , g_a に類似する概念 g_b ，関連する概念 g_c と無関係の概念 g_d を 200 組選択した．評価概念の一部を表 1 に示す．これらの概念を用いて，上記の (1) の評価指数 (F_1) として以下を設定する．

$$F_1 = (R_1 - R_3)/(\sigma_1 + \sigma_3). \quad (15)$$

ここで， R_1, R_3 はすべての組の $R(g_a, g_b)$, $R(g_a, g_d)$ の平均値であり， σ_1, σ_3 はその標準偏差である．

一方，(2) の評価指数 (F_2) としては，正しい判断 ($R(g_a, g_b) > R(g_a, g_c)$) の割合が大きいほどよいとし，以下を設定する．

$$F_2 = m/n. \quad (16)$$

ここで， m は正しい判断の数， n は全体の評価数 ($n = 200$) ．

これらを用い，類似度計算法の評価指数 (F) を以下のように設定する．

$$F = F_1 F_2. \quad (17)$$

表 1: 観点を考慮しない場合の評価データの例

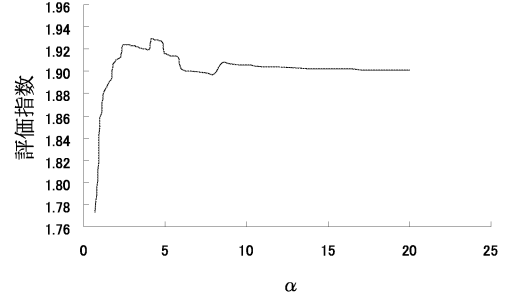
サンプル (g_a)	類似 (g_b)	関連 (g_c)	無関係 (g_d)
全力	総力	人力	鉦物
視野	視界	視線	戦死
悪臭	異臭	臭い	船
絶食	断食	試食	本
歓声	歓呼	叫ぶ	貯蓄
跳躍	飛躍	快速	唾
熟睡	熟眠	仮眠	番号
近道	早道	通行	学友
失踪	失跡	尾行	球
行為	行動	言動	橋

5.1.2 評価結果

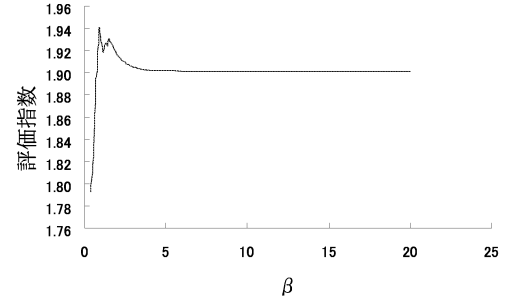
前節に述べた手法を用いて各類似度計算法を評価した．結果を図 3 に示す．式 (3) の従来の手法では，評価指数 F が 1.90 である．これに対し，上位属性を付与する手法では， $\alpha = 4.1$ のとき，評価指数 $F = 1.93$ が最大となっている．また，属性類似を利用する手法では $\beta = 0.9$ のとき，評価指数 $F = 1.94$ が最大となっている．なお，実験では，シソーラス上の距離が 4 より大きければ，二つの属性（カテゴリー）が全く類似しないと仮定し，式 (9) の属性間の類似度を 0 としている．

この結果から，本論文で提案する二つの手法がある程度優位であることが確認できた．

なお，提案した二つの手法は，従来の手法と比べて計算量が大幅増加する．二つの手法の評価指数がほぼ同じ最大値を持っているので，概念ベースの構成へのフィードバック（概念ベースの各概念の属性に予め上位属性を付加しておく）することにより計算量を削減できる上位属性を付与する手法が有効といえる．



a) 上位属性を付与する手法



b) 属性間類似性を利用する手法

図 3: 観点を考慮しない場合の評価

5.2 観点を考慮した類似度

5.2.1 評価手法

概念 g と観点 v_a において類似する概念 g_a ，同じく観点 v_b において類似する概念 g_b が与えられたとき，以下の大小関係が期待される．

- $R(g, g_a|v_a) > R(g, g_b|v_a)$.
- $R(g, g_a|v_b) < R(g, g_b|v_b)$.

このような二つの観点と三つの概念を評価データとして 50 組作成した．評価データの一部を表 2 に示す．評価指数の一つとしては，上記の大小関係を着目して，二つの観点で期待された大小関係が保たれる場合に 1 点，片方の観点で大小関係が反転する場合に 0 点，両方の観点とも大小関係が保たれない場合に -1 点とする．これを以下の式で求める．

$$E_1 = \frac{GG - BB}{50}. \quad (18)$$

ここで， GG は二つの観点で期待された大小関係が得られた場合の数， BB は二つの観点とも期待され

た大小関係が得られない評価の数を表す [3].

また, R_1 と R_2 をそれぞれ $R(g, g_a|v_a)$ と $R(g, g_b|v_a)$ の平均, σ_1 と σ_2 をそれらの標準偏差としたとき, 以下に設定する E_2 が大きい方がよいと考えられる.

$$E_2 = (R_1 - R_2)/(\sigma_1 + \sigma_2). \quad (19)$$

また, R_3 と R_4 をそれぞれ $R(g, g_a|v_b)$ と $R(g, g_b|v_b)$ の平均, σ_3 と σ_4 をそれらの標準偏差としたとき, 同様に以下の E_3 が大きい方がよい.

$$E_3 = (R_4 - R_3)/(\sigma_3 + \sigma_4). \quad (20)$$

これらを考慮して観点を用いた類似度計算法の総合評価指数 E を以下のように設定する.

$$E = E_1 \frac{E_2 + E_3}{2}. \quad (21)$$

この評価指数の設定の基本的な考え方は 5.1.1 節と同じである.

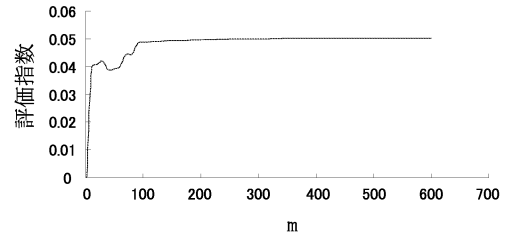
表 2: 観点を考慮した場合の評価データの例

サンプル (g)	概念 A (g_a)	観点 A (v_a)	概念 B (g_b)	観点 B (v_b)
馬	豚	動物	車	乗る
テレビ	ラジオ	電波	新聞	見
金	鉄	金属	ダイヤモンド	高価
飛行機	船	乗り物	鳥	飛ぶ
博物館	美術館	観光	図書館	調査
香水	口紅	化粧	ビール	液体
競馬	競輪	博打	闘牛	動物

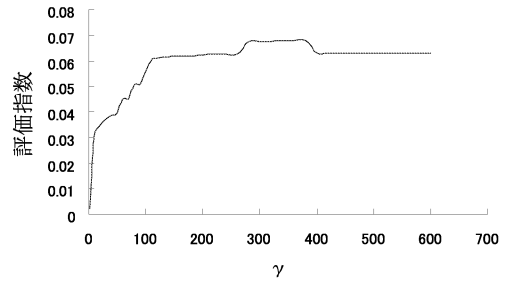
5.2.2 評価結果

従来の変調関数と新しく提案した二つの変調関数について前節の手法を用いて評価を行った. 図 4 に評価結果を示す. 同図から分かるように, 従来の関数 (式 (12)) と観点に含まれる属性のみを考慮する関数 (式 (14)) を用いる場合, 評価指数が変調率に比例して増加する傾向が見られる. 評価指数の上限はそれぞれ約 0.050 と 0.053 であった. これに対して, 観点の属性に比例した変調率 (式 13) を用いる場合, $\gamma = 380$ のとき, 最大の評価指数 $E = 0.068$ が得られた. 観点に含まれる属性のみを考慮する手法の評価指数が従来の手法とほとんど変わらない原因としては, 比較概念に関係ない観点に対する識別能力が評価指数に含まれていないことが考えられる.

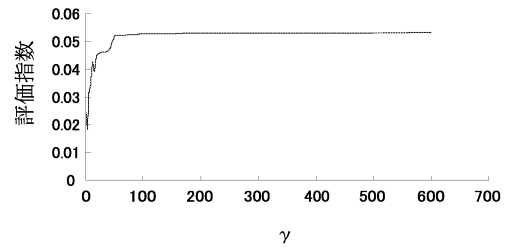
なお, 類似度計算としては従来の手法 (式 (3)) を用いている.



a) 従来の手法



b) 観点の属性に比例した変調率



c) 観点に含まれる属性のみを考慮する

図 4: 観点を考慮した場合の評価

6 むすび

本稿では, ベクトルで表現された概念 (単語) に対し, ベクトル空間の基底の意味的な関連性を考慮した概念間の類似度計算手法を提案した. 想定する概念ベースでは, 各概念は国語辞書の語義文をもとに機械的に構築され, 複数の属性と重みのペアで表現されている. これまでの研究では, 属性が互いに直交していると仮定し, 概念間の類似度は同一の属性の共有度合いで求めてきた. これに対し, ここではシソーラスを用いて異なる属性間の関連性を求め, これを考慮する概念間の類似度計算手法を提案した. また, 実験によりこの方法が従来法以上の類似性判別能力を有することを確認した.

また, 観点を考慮した概念間の類似度について検討し, この類似度の条件を明らかにするとともに, これらの条件を満たす二つの計算手法を提案した. これら計算方法についても実験により従来法以上の効果が得られることを確認した.

参考文献

- [1] J. Morris and G. Hirst : Lexical Cohesion Computed by Thesaural Relations as an Indicator of the Structure of Text, Computational Linguistics, Vol. 17, No. 1, pp. 21–48, 1991.
- [2] 笠原 要, 松沢和光, 石川 勉, 河岡 司 : 観点に基づく概念間の類似性判別, 情処学会論文誌 Vol. 35, No. 3, pp. 505–509, 1994.
- [3] 笠原 要, 松沢和光, 石川 勉 : 国語辞書を利用した日常語の類似性判別, 情処学会論文誌 Vol. 38, No. 7, pp. 1272–1283, 1997.
- [4] 宮原隆行, 清木 康, 北川高嗣 : 意味の数学モデルによる意味的連想検索の高速化アルゴリズムとその実現方式, 情処学会論文誌 Vol. 38, No. 7, pp. 1399–1411, 1997.
- [5] 小島秀樹, 伊藤 昭 : 文脈依存的に単語間の意味距離を計算する一手法, 情処学会論文誌 Vol. 38, No. 3, pp. 482–489, 1997.
- [6] 池原 悟, 宮崎正弘, 横尾昭男 : 日英機械翻訳のための意味解析辞書, 情処学会自然言語処理研資, Vol. 84–13, pp. 95–102, 1991.
- [7] 石川 勉, 井澤潤一郎, Nguyen V. Ha, 笠原 要 : 単語の意味に関する概念ベースの類似性判別能力からの最適構成, 人工知能学誌, Vol. 13, No. 3, pp. 470–479, 1998.