

連想のための大規模概念ベース構築における重み付け手法の一般化

北川 晋也[†] 渡部 広一[‡] 河岡 司[‡]

^{†‡}同志社大学大学院工学研究科 〒610-0394 京都府京田辺市多々羅都谷 1-3

E-mail: [†]dtg0717@mail4.doshisha.ac.jp, [‡]{hwatabe, tkawaoka}@mail.doshisha.ac.jp

あらまし 本研究では、コンピュータが人間のように常識的な判断を行うことにより、人間との円滑な会話を行うことを目的としている。この常識的な判断を連想メカニズムにより実現している。連想メカニズムは、ある単語(概念)に対して単語を特徴付ける語(属性)とその重みにより定義した概念ベースと、語と語の関連性を定量的に評価する関連度計算方式により構成される。現在の概念ベースの構築に用いられている重み付け手法は、目視評価を含むため、同一の概念・属性を用いても異なる概念ベースが構築され、再現性に乏しい。そこで、本研究では属性から得られる情報を用いて、再現性がある一般的な重み付け手法を提案する。

キーワード 概念ベース, 関連度, 重み付け

Generalization of Attributes Weighting Method in Construction of Large Scale Concept-Base for Association

Shinya KITAGAWA[†] Hirokazu WATABE[‡] and Tsukasa KAWAOKA[‡]

^{†‡}Graduate School of Engineering, Doshisha University

1-3 Miyakodani Tatara Kyotanabe-shi, Kyoto, 610-0394 Japan

E-mail: [†]dtg0717@mail4.doshisha.ac.jp, [‡]{hwatabe, tkawaoka}@mail.doshisha.ac.jp

Abstract An aim of this study is to perform smooth conversation between human and computer by association mechanism. The main components of this mechanism are a Concept-base and a calculation method of Degree of Association. Concept-base is a knowledge base which is defined by sets of attributes with weights as "the concept" of the word. The current attributes weighting method is poorly-reproducible because it includes the visual evaluation. For this reason, this study proposes for generalization of the reproducible attributes weighting method of a Concept-base.

Keyword Concept-Base, Degree of Association, Attributes Weighting

1. はじめに

近年、情報処理技術は急速に発展してきており、人間の生活になくてはならない技術となっている。本研究では、コンピュータ上で常識的な判断を可能とし、円滑な会話を行うことを目的としている。常識的な判断とは、例えば「贈り物」は「うれしい」といった、人間であれば常識と考える判断である。この常識的な判断を実現するのが常識判断メカニズム^[1]である。

このように、人間のような柔軟な判断を行うにはコンピュータ上で人間の持つ連想処理を実現する必要がある。その連想処理の実現のために、語の“概念”を、概念の意味・特徴を表す語(属性)とその重みの集合で定義した概念ベース^[2]や、概念間の関連の強さを定量的に評価するための関連度計算方式^[3]などで構成される連想メカニズムを利用している。このように、本研究における概念ベースの目的は「連想処理の実現」であり、ベクトル空間モデル^[4]などの概念をベクトルとみなして語と語の類似性を定量化するための概念ベー

スとは本質的に異なる。

概念ベースの構築は概念・属性の獲得と重み付けに大別される。現在の概念ベース構築手法^[5]は重み付けの段階で目視評価を含むため、再現性に乏しいという問題がある。そこで、目視評価を含まない再現性の高い概念ベースの重み付け手法、つまり概念・属性を獲得すれば一意に重みを決定可能な手法を目的とする。また、Webなどの情報源には雑音が多く含まれるため、提案手法の雑音に対する耐性を検証する。

2. 連想メカニズム

連想メカニズムは、主に概念ベースと、関連度計算方式から構成される。以下に各要素について述べる。

2.1. 概念ベース

概念ベースは、電子化辞書や電子化新聞、シソーラス等から機械的に構築された大規模知識ベースである。見出し語である概念 A を、概念 A を特徴付ける語であ

る属性 a_i と特徴付ける強さである重み w_i の対の集合として次式のように定義する。

$$\text{概念 } A = \{(a_1, w_1), \dots, (a_i, w_i), \dots, (a_n, w_n)\} \quad (1)$$

図 1 に概念ベースの一部を示す。任意の属性は、必ず概念ベース内に含まれる語で定義している。従って、概念の属性を 1 次属性とした場合、1 次属性に対してそれぞれの 1 次属性を導くことができる。この集合を 2 次属性と呼ぶ。同様に、概念は任意の次元までの属性 (n 次属性) を導出することが可能である。以降の節では、単に「属性」と表記する場合、各概念の 1 次属性を表すものとする。なお、概念自身は必ず属性として定義している。

語	属性(1次属性)
雪	(雪, 0.61), (白い, 0.30), (下る, 0.27), ...
白い	(雪, 0.16), (白地, 0.14), (色, 0.14), ...

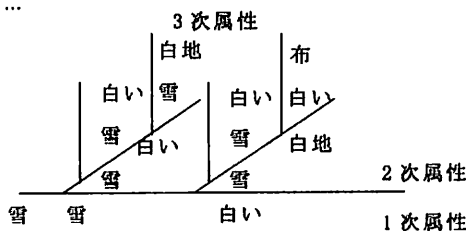


図 1 概念ベース(一部)

2.2. 関連度計算方式

関連度計算方式は、概念ベースに定義された語と語の関連の強さを定量化する手法である。これに対し、語と語の類似性評価手法として、シソーラスなどを用いて属性の意味的な圧縮を行った概念ベースを前提に各概念をベクトルと見なし、余弦を用いるベクトル空間モデルが広く利用されている。この方式は、類似性の評価には適しているが、「赤ちゃんと玩具」や「自動車と事故」といった語と語の関連性は、類似性という観点からは評価が困難であると考えられる。そのため、本論文に使用する概念ベースは、より柔軟に語と語の関連の強さを定量化するために関連度計算方式を前提としている。

以下、使用する属性の選択、概念間の一致度、並びに一致度に基づく関連度を求める関連度計算方式について述べる。なお、以降の説明では概念 A, B の属性を a_i, b_j 、対応する重みを u_i, v_j とし、それぞれ属性が L 個、 M 個あるとする。 ($L \leq M$)。

$$A = \{(a_1, u_1), \dots, (a_i, u_i), \dots, (a_L, u_L)\} \quad (2)$$

$$B = \{(b_1, v_1), \dots, (b_j, v_j), \dots, (b_M, v_M)\} \quad (3)$$

2.2.1. 使用する属性の選択

人間は二つの概念の関係を考えると、それらの似ている部分に着目していると考えられる。例えば、概

念「馬」に対して、概念「車」であれば「乗り物」という観点で、概念「兎」であれば「動物」という観点で関連性を見出ししていると考えられる。そこで、対象となる概念に合わせて一致度及び関連度の計算に使用する属性を変化させることにより、この仕組みを実現する。具体的には、対象となる二つの概念に共通する属性を優先して選出し、残りを重み上位より選出する。なお、この後、選ばれた属性は重みの和が 1 になるよう正規化を行う。また、過去の研究^[3]より属性は 30 個まで選出する。

2.2.2. 一致度

概念 A と概念 B の一致度 $DoM(A, B)$ (DoM : Degree of Match) を以下のように定義する。

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

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

ただし、 $a_i = b_j$ は属性同士が一致した場合を示している。すなわち、一致した属性の重みのうち、小さい方の重みの和が一致度となる。このとき一致度は 0~1 の実数値をとる。

2.2.3. 関連度

属性数の少ない概念 A の 1 次属性の並びを固定し、概念 B の各 1 次属性を、対応する概念 A の各 1 次属性との一致度の合計が最大になるよう並び替える。

$$B_x = \{(b_{x1}, v_{x1}), (b_{x2}, v_{x2}), \dots, (b_{xL}, v_{xL})\} \quad (6)$$

ここで、属性同士が完全一致する場合 ($a_i = b_{xn}$) は対応する属性の重み u_n, v_{xn} の大きさを重みの小さい方にそろえる。このとき、重みの大きい方はその値から小さい方の重みを引き、再度、他の属性と対応をとる。このように対応を決めると次式ようになる。

$$A = \{(a_1, u_1), \dots, (a_L, u_L), (a_1', u_1'), \dots\} \quad (7)$$

$$B_x = \{(b_{x1}, v_{x1}), \dots, (b_{xL}, v_{xL}), (b_{xL+1}, v_{xL+1}), \dots\} \quad (8)$$

a_L 以降に並ぶ概念 A の属性は元々概念 B の属性と完全一致していた属性で再度対応を取ったものである。 b_{xL+1} 以降の属性についても同様である。このとき対応の取れた属性の組み合わせが T 個の場合、概念 A と概念 B の関連度 $DoA(A, B)$ (DoA : Degree of Association) を次式で定義する。

$$DoA(A, B) = \sum_{s=1}^T \left(DoM(a_s, b_{xs}) \times \frac{u_s + v_{xs}}{2} \times \frac{\min(u_s, v_{xs})}{\max(u_s, v_{xs})} \right) \quad (9)$$

$$\max(\alpha, \beta) = \begin{cases} \alpha (\alpha \geq \beta) \\ \beta (\alpha < \beta) \end{cases} \quad (10)$$

関連度も一致度と同様に 0~1 の実数値を取る。

3. 関連度を用いた概念ベース評価手法

概念ベースの評価には、表 1 のように人手によって作成したデータ 1780 組を用いる。データは概念 X に対して、概念 A は高関連、概念 B は中関連、概念 C は無関連となるように作成する。

表 1 X-ABC 評価用データ (一部)

X	A	B	C
海	海洋	塩	車
ヨット	帆船	スポーツ	没年

ここで m をセット数とし、次式を満たしたとき正解とする。そして、正解したセットの割合を精度とする。

$$DoA(X,A) - DoA(X,B) > AveDoA(X,C') \quad (11)$$

$$DoA(X,B) - DoA(X,C) > AveDoA(X,C') \quad (12)$$

$$AveDoA(X,C') = \sum_{i=1}^m DoA(X_i, C_i) / m \quad (13)$$

4. 属性情報を用いた重み付け手法

本章では、既存の重み付け手法である属性信頼度^[5]と概念価値^[2]を用いた精練手法とその課題について述べ、提案手法である属性内頻度を用いた重み付け手法について述べる。

4.1. 属性信頼度を用いた重み付け手法

文書処理における重み付け手法として広く利用されている $tf \cdot idf$ ^[7] の考え方を概念ベースに適用し重み付けを行う。具体的には、概念 A の属性 B に対する重み $u(A,B)$ を、 tf に対して属性信頼度 $AR(A,B)$ (AR : Attribute reliability) を、 idf に対して概念価値 $CV_n(B)$ (CV : Concept Value) を用い、次式により定義する。

$$u(A,B) = AR(A,B) \times CV_n(B) \quad (14)$$

以下に属性信頼度と概念価値の算出方法について述べる。

4.1.1. 属性信頼度

属性信頼度とは、概念ベースにおける各概念に対する属性の確からしさを表した値である。概念ベースに定義されている概念と属性は、電子化辞書や電子化新聞などの情報を基に構築しているため、異なる情報から得られる信頼性だけの判断が困難である。そこで、シソーラス^[6]などの定性的な情報と、定量的な情報である関連度を用いて複数のランクに分類する。そして、ランクに分類された属性が所属するランクが、概念にとってどの程度確からしい属性を含んでいるかを、サンプル 100 概念を用いて評価する。そして、各ランクに含まれる属性が適切である割合を属性信頼度とする。

4.1.2. 概念価値

概念価値とは、概念ベース全体における各概念の価値を表す値である。 idf と同様に、概念ベースを仮想的な文書集合として捉えることで算出する。具体的には、

n 次属性集合内において、対象となる概念を属性として持つ概念数を算出する。概念価値を次式により定義する。

$$CV_n(s) = \log_2 \frac{V_{all}}{df_n(s)} \quad (15)$$

$CV_n(s)$ は n 次属性集合における概念 s の概念価値である。 V_{all} は概念ベースに定義されている概念数、 $df_n(s)$ は n 次属性集合内に概念 s を属性として持つ概念数である。

4.1.3. 属性信頼度を用いた重み付け手法の課題

属性信頼度を用いた手法では、構築するごとに人手によって目視評価を行う必要がある。そのため、構築に手間がかかり、概念ベースの拡張には適さない。そして、同じ概念・属性を使用しても目視評価の結果により異なる重みが付与されるため再現性に乏しい。また、固有名詞など、定性的な知識に登録されていない語に対する重み付けには適さないと考えられる。

4.2. 属性内頻度と属性内頻度割合を用いた重み付け手法

4.1.3 節で述べた属性信頼度を用いた重み付け手法の課題を解決するために、以下の条件を満たす重み付け手法を提案する。

- ・概念と属性は既存の手法により取得し、重みは付与されていないものとする
- ・属性信頼度を用いた重み付け手法とほぼ同精度の概念ベースを構築できる
- ・構築するごとに目視評価を行わない
- ・知識を使用せず固有名詞などにも対応する
- ・雑音の割合が多い属性群に対しても適切な重み付けを行うことができる

以上の条件を満たす手法として、関連度を用いる。しかし、関連度の算出には重みが必要である。そこで、関連度の算出に使用する重み(仮の重み)を n 次属性空間より定量的に算出できる属性内頻度と属性内頻度割合を用いて算出する。提案手法の流れを図 2 に示す。

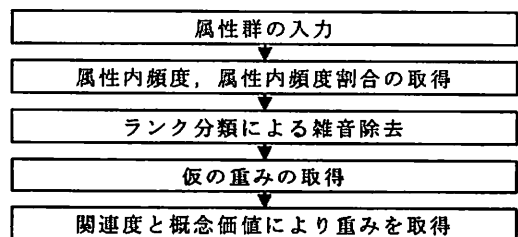


図 2 提案手法の流れ

各処理の詳細は以降の節で述べる。

4.2.1. 属性内頻度と属性内頻度割合

属性内頻度とは、 $tf \cdot idf$ 法における tf と同様の働き

をするものであり、ある概念の n 次属性での頻度を算出することにより求める。概念ベースに存在する属性は、主に国語辞書の語義文や新聞記事内で共起する語である。このため、属性は概念を説明する文書の自立語であると捉えることができる。例として、概念「個人情報」の1次属性と2次属性を表2に示す。1番上のセルに1次属性、2番目以降のセルにはその2次属性を表している。

表2 概念「個人情報」の1次属性と2次属性

1次属性	個人情報	個人	情報	識別
2次属性	個人情報	個人	情報	識別
	個人	人	事物	判断
	情報	費用	内容	色
	識別	擬人	様子	調査

概念「個人情報」の1次属性は「個人情報とは、個人を識別する情報」という語義文の自立語と考えることができる。同様に、2次属性は、1次属性を説明した文書である。つまり、概念に対する n 次属性はその概念を説明した文書の集合であるとみなすことができる。

そこで、概念 A における概念 B の n 次属性内頻度を概念 A の n 次属性における概念 B の出現頻度を定義する。また、概念によって属性数が異なることを考慮するために、 n 次属性内頻度を $n-1$ 次属性数で割った値を属性内頻度割合と定義する。1次属性には重複して出現する概念は存在しないため、 n 次属性内頻度の最大値は $n-1$ 次属性空間の属性数である。よって n 次属性内頻度割合は $0 \sim 1$ の実数値を取る。表3に表2の例をもとに取得した概念「個人情報」の2次属性内頻度と2次属性内頻度割合を示す。

表3 「個人情報」の属性の2次属性内頻度と2次属性内頻度割合

属性	個人情報	個人	情報	識別
2次属性内頻度	1	3	3	2
2次属性内頻度割合	0.2	0.6	0.6	0.4

雑音が多く含まれる概念では、関係のある属性であっても属性内頻度割合が低く算出される。これに対して、属性数が極端に少ない概念では、関係のある属性であっても属性内頻度が低く算出される。以上より、本論文では属性内頻度と属性内頻度割合の両方を用いた手法を提案し、その有効性を検証する。

4.2.2. ランク分類による雑音除去

属性内頻度は属性の有無のみで変化するため、属性同士の類似までみる関連度と比較して雑音に弱いと考えられる。そこで、雑音に対する耐性を向上させるため、属性内頻度と属性内頻度により複数のランクに分類する。具体的には、後述する $X-ABC$ 評価用データの概念 X に対する概念 C の属性内頻度及び属性内頻度割

合の傾向より、属性を α, β, γ の3つのランクに分類し、それぞれ $1, 0.5, 0$ をランクの値として付与する。このとき 0 に分類されたものは関連度の計算には使用しない。図3にランク分類の流れを示す。

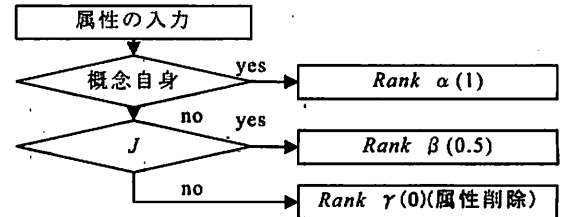


図3 ランク分類の流れ

J の判定は、次の条件のいずれかを満たす場合を真とする。

- ・概念 B が概念 A を属性として保持している
- ・属性内頻度が閾値以上
- ・属性内頻度割合が閾値以上

この処理の後で利用する関連度は複数の属性の一致度合いにより求められる。よって、少しの雑音が含まれていても影響は小さい。しかし、必要な属性が削除されてしまった場合、関連度は低く算出されてしまい影響は大きい。以上の理由により、 J の判定において雑音ではないと考えられる結果が1つでも出た場合はその属性を削除しない。

4.2.3. 仮の重みの取得

4.2.2節のランク分類により得られた値は、 $if \cdot idf$ の if にあたる部分だけ考慮されているため、 idf にあたる概念値を考慮する。このとき、精緻前の属性群は信頼性が低いため、1次概念値を用いる。よって、ランク分類により得られた値を $R(A,B)$ とした場合、概念 A の属性である概念 B の仮の重み $u'(A,B)$ を次式で定義する。

$$u'(A,B) = R(A,B) \times CV_1(B) \quad (16)$$

4.2.4. 関連度による重み付け

4.2.3節で得られた仮の重みを付与した概念ベースを用いて、関連度を算出し、概念 A に対する属性 B の重み $u(A,B)$ を次式により求める。

$$u(A,B) = DoA(A,B) \times CV_2(B) \quad (17)$$

つまり、関連度に概念値を乗じたものを重みとする。なお、ランク分類によりランクの値が 0 となった属性候補も関連度計算を行い、算出された重みが 0 にならなければ属性として採用する。

5. 検証実験

本章では、属性内頻度と属性内頻度割合によるランク分類の閾値の検証について述べた後、属性信頼度による重み付け手法との比較、提案手法の雑音に対する

耐性の検証について述べる。

5.1. 属性内頻度と属性内頻度割合の検証

属性内頻度と属性内頻度割合が概念と属性の関係の強さを表すために有効であるか、既存の概念ベースの属性群と人手で作成した *X-ABC* 評価用データ 1780 セットを用いて調査を行った。具体的には、概念 *X* の 2 次属性内に含まれる概念 *A* の数を概念 *X* に対する概念 *A* の 2 次属性内頻度、2 次属性内頻度を概念 *X* の 1 次属性数で割ったものを 2 次属性内頻度割合とした。同様に、概念 *X* に対する概念 *B*、概念 *C* の 2 次属性内頻度、2 次属性内頻度割合についても求めた。2 次属性内頻度と 2 次属性内頻度割合の累積頻度分布をそれぞれ、図 4、図 5 に示す。なお、重み付けの対象となる属性候補は、雑音であっても属性として定義されているため、概念 *X* の属性に定義されているとして、*X-ABC* 評価用データにおける 2 次属性内頻度、2 次属性内頻度割合を求めた。

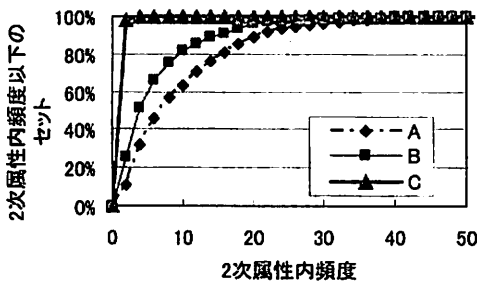


図 4 *X-ABC* 評価用データ 1780 セットにおける 2 次属性内頻度の累積頻度分布

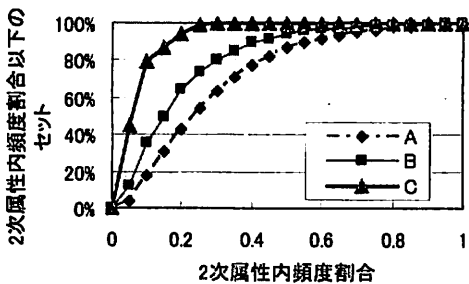


図 5 *X-ABC* 評価用データ 1780 セットにおける 2 次属性内頻度割合の累積頻度分布

まず、図 4 より、無関連である概念 *C* との属性内頻度はほとんど 2 以下に存在しているため、関連の有無を判断することはできると考えられる。しかし、高関連である概念 *A* と中程度の関連がある概念 *B* の分布は同じような傾向を持っているため、関連の強さを判断

することは難しいと考えられる。また、図 5 より、属性内頻度割合では概念 *C* と概念 *A*、概念 *B* を判断することも難しいと考えられる。これらの結果では、属性内頻度割合は不要と考えられるが、4.2.2 節で述べたように、概念の属性数が少ない場合、適切な属性であっても属性内頻度は小さくなるため、属性内頻度割合も用いる。

属性内頻度と属性内頻度割合の閾値については、*X-ABC* 評価用データの内、*X* に対する *C* の値が閾値以下のセットを一定割合含む値とする。例えば、90% の場合、*X-ABC* 評価用データの内、*XC* の値が閾値以下のセットが 90% 存在するように閾値を設定する。よって、図 4 より属性内頻度の閾値は 2、図 5 より属性内頻度割合の閾値は 0.167 となる。本論文では実験的に閾値以下に 95% 含む値とする。

5.2. 属性信頼度を用いた重み付け手法との比較

既存の概念ベースの属性群を用いて提案手法が有効であるかを検証する。まず、既存の概念ベースを CB1 とする。これは、4.1 節の属性信頼度を用いて重み付けを行った概念ベースである。次に、提案手法を用いて重み付けした概念ベースを CB2 とする。最後に、概念価値のみで重みを算出した概念ベースを CB3 とする。これらの概念ベースを 1780 セットの *X-ABC* 評価用データを用いて精度を求めた。結果を図 6 に示す。

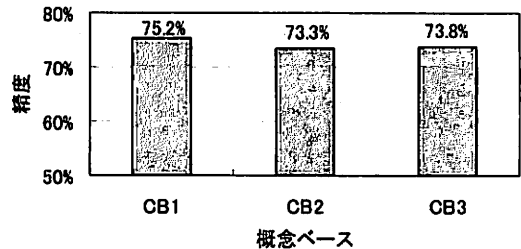


図 6 既存の概念ベースの属性群を用いた場合の各手法の精度

CB1、CB2、CB3 の精度はそれぞれ、75.2%、73.3%、73.8%であった。しかし、CB1 と CB2 の精度には 1.9% の差しかなく、ほぼ同等の精度が得られたと考えられる。CB1 と CB2 の精度に差が生じた原因として、以下の 2 点が考えられる。1 点は、関連度などの定量的な評価値に加え、シソーラスなどの定性的な知識を用いた点である。もう 1 点は、属性をランクに分類した後に属性を目視により評価し、ランクごとの精度をそのランクの属性信頼度としている点である。目視評価は人手による作業であるため、再現性に乏しく、労力を要するが、人の評価が反映されるため、定量的な値のみの提案手法よりも高い精度の概念ベースを構築する

ことができたと考えられる。

5.3. ランク分類による雑音除去

5.2 節の検証結果では、検証した 3 つの重み付け手法に明確な差が生じなかった。この原因として、使用した初期の属性群が、精練された属性群であることが挙げられる。そのため、意図的に雑音を付与した属性群を用いて各手法の雑音に対する耐性を検証する。

雑音を付与した属性群には、5.2 節で使用した属性群にランダムに一定個数属性を付与したものをを用いる。追加する属性数は 1 概念あたり 20 個、40 個、60 個の 3 種類を用いる。また、雑音として取得した語の中には、適切に概念を特徴付ける語が含まれる可能性があるが、約 12 万語ある概念から適切な概念を選択する可能性はきわめて低いと考えられるため、新たに追加した属性は全て雑音であるとしてよいと考えられる。

提案手法と比較する手法は概念価値のみを重みとする手法とする。属性信頼度を用いた手法は目視評価を必要とするため比較しない。雑音を付与した場合の概念ベースの精度を図 7 に、既存の属性と雑音として追加した属性の重みの平均を表 4 に示す。

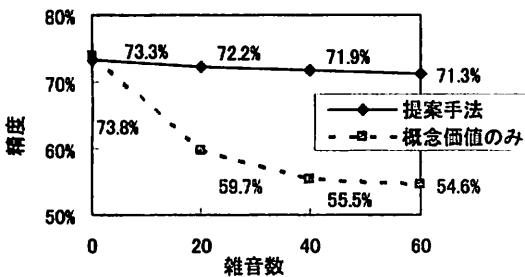


図 7 雑音を付与した場合の各手法の精度

表 4 既存の属性と雑音として追加した属性の重みの平均

1 概念あたりの雑音数		20	40	60
提案手法	既存の属性	0.065	0.056	0.048
	追加した属性	0.004	0.002	0.002
概念価値のみ	既存の属性	0.197	0.169	0.149
	追加した属性	0.265	0.193	0.147

図 7 より、いずれの手法も雑音を増加させることにより精度の低下が生じている。しかし、雑音を 60 個追加した場合の精度と、雑音を追加しない場合の精度を比較すると、概念価値のみの手法では、精度が約 19% 低下しているのに対し、提案手法では、約 2% の低下に抑えられている。

また、表 4 より、1 に概念価値のみの手法では、既存の属性と雑音として追加した属性の重みの平均がほぼ同じであるのに対し、提案手法では、既存の属性の重みの平均に対して雑音として追加した重みの平均は

1/10 以下になっている。このことより、提案する重み付け手法は雑音属性の重みを下位に移行することが可能であることがわかる。

以上のことから、概念価値のみの手法と比較して提案手法には雑音に対する耐性があるものと考えられる。

6. おわりに

本論文では、属性内頻度を用いた連想のための概念ベースの一般的な重み付け手法を提案した。そして、提案手法が、目視評価を必要とする属性信頼度による重み付け手法とほぼ同等の精度の概念ベースを構築できることを示した。また、提案手法が雑音に耐性があることを示した。この手法により、再現性があり、拡張にも柔軟に対応可能な概念ベースを構築できる。

より人間に近い判断を行うためには、人手による評価を考慮した概念ベース構築が理想である。それに対し、本論文では連想のための一般的な重み付け手法を確立することにより、膨大な知識ベースを持たない状況においても連想のための概念ベース構築が可能となることを示した。

謝辞

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ものである。

文献

- [1] 土屋誠司, 小島一秀, 渡部広一, 河岡司, “常識的判断システムにおける未知語処理方式”, 人工知能学会論文誌, vol.17, no.6, pp.667-675, Aug.2002.
- [2] 奥村紀之, 土屋誠司, 渡部広一, 河岡司, “概念間の関連度計算のための大規模概念ベースの構築”, 自然言語処理, vol.14, no.5, pp.41-64, Oct.2007.
- [3] 荒木孝允, 奥村紀之, 渡部広一, 河岡司, “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”, 同志社大学理工学研究報告, vol.48, no.3, pp.14-24, Oct.2007.
- [4] 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, July.1997.
- [5] 広瀬幹規, 渡部広一, 河岡司, “概念間ルールと属性としての出現頻度を考慮した概念ベースの自動精練手法”, 信学技報, Vol.101, No.711, pp.109-116, Mar.2002.
- [6] 池原悟, 官崎正弘, 白井論, 横尾昭男, 中岩浩巳, 小倉健太郎, 大山芳史, 林良彦(編), 日本語語彙体系, 岩波書店, 東京, 1997.
- [7] 徳永健伸, 情報検索と言語処理, 辻井潤一(編), 東京大学出版会, 東京, 1999.