E-001

同義語・類義語を用いて観点を拡張した観点付き関連度計算方式
Calculating Degree of Association Incorporating Viewpoint using Synonyms

松本 和也† 芋野 美紗子‡ 土屋 誠司‡ 渡部 広一‡
Kazuya Matsumoto Misako Imono Seiji Tsuchiya Hirokazu Watabe

1. はじめに

人間は，ある語から関連性のある語を連想する能力があり，この連想能力を日常の会話で役立てている．この連想能力をコンピュータに持たせることができれば，人間のように自然な会話ができると考えられる．そこで，ある語に対して，そこから連想できる他の語の集合を属性として持たせることで概念化した概念ベース[1]や，概念間の意味の近さを測る関連度計算方式[2]を用いてこの連想能力を実現している．そこで，この関連度計算方式に着目し，より人間の感覚に近い意味の近さを算出する手法について述べる．

人間が異なる概念間の関連性を評価する時，その概念の見方である観点を利用して，その関連性を評価していると考えられる．例えば，「飛行機と車」，「飛行機と鳥」の関連性を評価する時，観点「乗り物」を用いた場合には「飛行機と車」，観点「飛ぶ」を用いた場合には「飛行機と鳥」の関連が高くなると判断している．

そこで，概念間に観点として概念を指定し，3 つの概念を利用して関連度を計算する観点付き関連度計算方式[3]が提案されている．これにより，関連度を変化させ，より人間の感覚に近い関連度計算方式を実現している．

本稿では，観点とする概念の同義語・類義語を取得し，それらも観点として利用して関連度を計算する観点付き関連度計算方式を提案する．

2. 概念ベース

概念ベースは電子化された国語辞書や新聞記事などから自動的に構築された知識ベースである．ある語を概念と定義し，概念の意味特徴を表す属性と，属性の重要性を表す重みの対の集合により構成されている．ある概念 A は m 個の属性 a_i と重み w_i（>0）の対により定義される．

$$A=\{(a_i, w_i) \mid i=1\sim m\} \quad (1)$$

属性 a_i を概念 A の一次属性と呼ぶ．一次属性は概念ベースの中で概念として定義されている．つまり概念 A の属性 a_i を概念とみなし，更に属性を導くことができ，これを概念 A の二次属性と呼ぶ．このように，概念ベースにおいて概念は N 次までの属性の連鎖集合である．

3. 関連度計算方式

関連度計算方式は，2 つの概念間の関連の強さを定量的に評価する手法である．関連度は 0.0 から 1.0 までの実数値で表現され，概念間の関連が強いほど高い値を示す．関連度計算方式では，概念同士がもつ属性それぞれを意味が近いもの同士で対応付ける．

まず，属性の対応付けに用いる一致度について述べる．

概念 A，B の一次属性を a_i，b_j，重みを u_i，v_j とし，各概念が持つ属性の個数を L 個，M 個($L<M$)とすると，概念 A，B は以下のように表現される．

$$A = \{(a_i, u_i) \mid i=1\sim L\} \quad (2)$$

$$B = \{(b_j, v_j) \mid j=1\sim M\} \quad (3)$$

概念 A，B の一致度 $DoM(A,B)$ を以下の式で定義する．

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

$a_i=b_j$ は属性同士が表記的に一致した場合を示している．つまり，一致度とは概念 A と概念 B の両方が共通して持つ属性の内，小さい方の重みを足し合わせたものとなる．共通した属性は概念 A と概念 B でそれぞれ重みが付与されている．この重みの内，小さい方の重み分は概念 A と概念 B の両方の属性に有効であると考える．つまり一致度とは，両方の概念に有効な属性が持つ重みの和を示す数値である．

この一致度を利用して属性間の最も対応のよい組み合わせを決定し，概念 A と概念 B の関連度を計算する．

所持する属性数が少ない概念 A を基準とし，その一次属性の並びを固定する．その上で概念 B の一次属性を概念 A の各一次属性との一致度の和が最大となるように以下のように並べ替える．

$$B = ((b_{X1}, v_{X1}),\ (b_{X2}, v_{X2}),\ \cdots,\ (b_{XL}, v_{XL})) \quad (5)$$

概念 A，B の関連度 $DoA(A,B)$ は以下のように表現される．

$$DoA(A,B) = \sum_i DoM(a_i, b_{Xi}) \times \frac{(u_i + v_{Xi})}{2} \times \frac{\min(u_i, v_{Xi})}{\max(u_i, v_{Xi})} \quad (6)$$

概念 A と概念 B の関連度を計算する際，使用する属性の個数を変えることで，値は変化する．

4. 観点付き関連度計算方式

概念 A，観点 V の一次属性を a_i，v_j，重みを u_i，x_j とし，各概念が持つ属性の個数を L 個，M 個とする．このとき，概念 A と観点 V は，以下ように表現される．

$$A = \{(a_i, u_i) \mid i=1\sim L\} \quad (7)$$

$$V = \{(v_j, x_j) \mid j=1\sim M\} \quad (8)$$

概念 A の属性の重み u_i を以下のように変更する．

$$u_i = DoM(a_i, V) \quad (9)$$

各属性 a_i と観点 V との一致度を求め，それを属性 a_i の新たな重みとする．つまり，これまで一定であった概念の属性の重みを観点によって変更する．

5. 観点を拡張した観点付き関連度計算方式

観点付き関連度計算方式は，観点と概念の属性との一致度を概念の属性の新たな重みとすることで，観点を考慮した概念間の関連度が算出できる．しかし，観点と概念の属性が表記的に一致しないと一致度が小さい値や 0 となる属性が多数出現する．例えば，概念「飛行機」の属性「羽」と観点「飛ぶ」の一致度は 0 になり，観点を考慮して概念の属性の重みを変更できない可能性がある．しかし，「羽と飛ぶ」は関連があると考えられるのが一般的である．そ

†同志社大学大学院理工学研究科
Graduate School of Science and Engineering, Doshisha University
‡同志社大学理工学部
Faculty of Science and Engineering, Doshisha University

こで，観点と意味の近い同義語・類義語を取得し多角的に一致度の値を計算する．例えば，観点「飛ぶ」の同義語「飛行」を観点として属性「羽」の一致度を計算すると 0.08 となる．このように，観点を増やすことで観点を考慮した重みに変更が可能である．

5.1 同義語・類義語の獲得

観点の同義語・類義語は関係語辞書から取得する．関係語辞書は，岩波国語辞典[4]から見出し語とその同義語，類義語を取り出し，目視によって精錬を行って構築した辞書である．同義語の関係語辞書は 266 組，類義語の関係語辞書は 13064 組登録されている．

5.2 属性と複数の観点との一致度を重みとする手法

観点 V の同義語・類義語の 1 つを観点 W とする．概念 A, 観点 V, 観点 W のそれぞれの一次属性を a_i, v_j, w_k, 重みを u_i, x_j, y_k とし，各概念の持つ属性の個数を L 個，M 個，N 個とする．ここで，概念 A の属性の重み u_i を観点 V と観点 W を利用して u_{vi} と u_{wi} に変更する．u_{vi}, u_{wi} は概念 A の属性と観点 V, 観点 W との一致度である．

$$u_{vi} = DoM(a_i, V) \tag{10}$$

$$u_{wi} = DoM(a_i, W) \tag{11}$$

ここで，複数の観点から得られた重み u_{vi}, u_{wi} の内，表記的に一致する属性の最大値の重みを取得し，それを属性 a_i の新たな重みとする．表 1 の例では，概念「飛行機」の属性「翼」の重みは観点「飛ぶ」で変更した場合の 0.17 となる．最大値を用いることで，最も関連性が高い観点を利用した属性の重みを取得することができる．

表 1　複数の観点によって得られた属性の重みの例

概念	観点	（属性，重み）
飛行機	飛ぶ	(翼,0.17) (機翼,0.06) (羽,0.00) …
	飛行	(翼,0.11) (羽 0.08) (機翼,0.02) …

5.3 概念の属性の元の重みをかける手法

5.2 節と同様に，概念 A, 観点 V, 観点 W と定義し，重み u_i を観点 V と観点 W を利用して u_{vi} と u_{wi} に変更する．ここで，u_{vi} は概念 A の属性と観点 V との一致度に属性の元の重みをかけた値であり，u_{wi} は概念 A の属性と観点 W との一致度に属性の元の重みをかけた値である．

$$u_{vi} = DoM(a_i, V) \times u_i \tag{12}$$

$$u_{wi} = DoM(a_i, W) \times u_i \tag{13}$$

元の重みを考慮して得られた u_{vi}, u_{wi} の重みの最大値を取得し，それを属性 a_i の新たな重みとする．

6. 評価方法

概念 A と関連性の高い 2 つの概念 B と C, および概念 A と B に関連し，概念 C に関連のない観点 X, 概念 A と C に関連し，概念 B に関連のない観点 Y を用いる．この概念 A, B, C, X, Y により構成された評価セットを 200 組用意した．評価セットの例を表 2 に示す．

表 2　評価セットの例

A	B	C	X	Y
飛行機	自動車	鳥	乗り物	飛ぶ

以下の 2 つの式を満たす評価データを正解とする．$DoA(A, B \mid X)$ は観点 X を用いた概念 A と B の関連度である．

$$DoA(A, B \mid X) > DoA(A, C \mid X) \tag{14}$$

$$DoA(A, B \mid Y) < DoA(A, C \mid Y) \tag{15}$$

概念 A「飛行機」と概念 B「自動車」の両方に関連のある観点である観点 X「乗り物」を用いた場合の関連度が概念 A「飛行機」と観点 X「乗り物」では間違った観点である概念 C「鳥」との関連度よりも大きくなることで，観点を用いた効果があると判断できる．

7. 評価結果

一致度を重みとする手法が，使用属性数 60 個の時に最大精度 59.5%となり，既存手法に比べて最大精度は 8.5%向上した．各提案手法の精度を表 3 に示す．

表 3　各提案手法の精度

	使用属性数		
	50 個	60 個	70 個
既存手法	46.5%	51.0%	48.5%
一致度を重みとする手法	46.5%	59.5%	50.5%
元の重みをかける手法	41.0%	46.5%	42.5%

8. 考察

複数の観点を利用することで観点と概念の属性との表記が一致する割合を高めることができ，属性数が少ない概念に対しても適切に関連度が算出できたと考えられる．

評価セット 200 組の中で観点 X, 観点 Y の両方から同義語・類義語を取得できたのは 68 組あった．その 68 組で評価を行った際，既存手法では最大精度 52.9%に対して，提案手法では，最大精度 64.3%という結果を得ることができた．これは，複数の観点を利用することで 1 つの観点だけでは一致度が小さい値であった属性の個数を減らすことができたためだと考えられる．表 4 に観点「学校」と概念「先生」の属性との一致度が 0 の属性の個数を示す．概念「先生」の属性の個数は 136 個である．

表 4　各手法での重みが 0 の属性の数

	一致度が 0 の属性の数
既存手法	43 個
提案手法	24 個

9. おわりに

本稿では複数の観点を利用した観点付き関連度計算方式を構築した．複数の観点を利用することで観点と概念の属性との表記が一致する割合を高めることができ，一致度が小さい値であった属性に適切な値を付与できた．これにより，既存手法に比べて最大精度を 8.5%向上した．

謝辞

本研究の一部は，科学研究費補助金（若手研究（B）24700215）の補助を受けて行った．

参考文献

[1] 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).

[2] 井筒大志，渡部広一，河岡司，“概念ベースを用いた連想機能実現のための関連度計算方式”，情報科学技術フォーラム FIT2002，pp.159-160，(2002).

[3] Hirokazu Watabe, Misako Imono, Eriko Yoshimura and Seiji Tsuchiya“Calculating Degree of Association Incorporating Viewpoint Using a Concept-Base”， Proc. of ICAI2012 (WorldComp2012) CSREA Press， Vol. Ⅰ， pp.191-197，(2012)

[4] 西尾実，岩淵悦太郎，水谷静夫，“岩波国語辞典第五版”，岩波書店，(1994).