

観点を利用した関連度計算方式の拡張 The Expansion of the Degree of Association using the Viewpoint

古田 哲朗
Tetsuro Furuta

芋野 美紗子
Misako Imono

土屋 誠司
Seiji Tsuchiya

渡部 広一
Hirokazu Watabe

1. はじめに

近年、情報処理システムは目覚ましい発展を遂げているが、こういった発展の多くは機能面、性能面での高度化であり、操作方法は複雑化を辿る一方である。そこで、人間同士が会話をするように情報処理システムを扱う事が出来れば、簡単に操作が可能となり非常に便利となる。そのようなシステムを開発するためには、コンピュータがより多くの単語に対応出来なければならない。そのためには、各単語間の関連性を理解する能力が不可欠である。これを実現するために、単語を概念とした大規模知識ベースの概念ベース^[1]、その概念間の関連性を定量的に表現できる関連度計算方式^[2]を用いる。関連度計算方式によって得られる値を関連度と呼び、2つの概念間の関連度は常に一定の値である。

しかし単語間の関連性は常に一定ではなく、会話の流れ等により変化すると考える。そこで新たに観点を与えることで、その観点に基づいて関連度を変化させる観点付き関連度計算方式を提案する。その結果、より人間に近い常識判断を行うことが出来るようになり、人間との円滑なコミュニケーションの手助けになると考える。

2. 概念ベース

概念ベースとは、電子化された国語辞書や新聞記事などから、ある一定のルールに従って機械的に構築された知識ベースのことである。概念に対して、その特徴を表す属性、および属性の重要さを示す重みの対を複数付与することで構成されている。

ある概念 A は m 個の属性 a_i と重み w_i (>0) の対によって次のように表現される。

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

また、属性として用いられる語は必ず概念ベースの中で概念として定義されている。概念が持つ属性を一次属性と呼び、一次属性が持つ属性を二次属性と呼ぶ。概念ベースは N 次の属性の連鎖集合の構造である。

3. 関連度計算方式

関連度計算方式とは、概念ベースに定義されている二つの概念間の関連の強さを定量的に表現する手法である。関連度は $0.0 \sim 1.0$ の実数値で表される。概念間の関連度を計算する際には、各概念の属性間にどれくらい一致する属性があるかを示した一致度を用いる。以下にその一致度の求め方とそれを用いた重み比率付き関連度計算方式について述べる。

「同志社大学大学院工学研究科
Graduate School of Engineering, Doshisha University
同志社大学理工学
Faculty of Science and Technology, Doshisha University

概念 A, B の一次属性を a_i, b_j 、重みを u_i, v_j とし、各概念の持つ属性の個数を L 個、 M 個 ($L < M$) とする。この時概念 A, B の一致度 $DoM(A, B)$ を次式のように定義する。

$$DoM(A, B) = \sum_{a_i=b_j} \min(u_i, v_j) \quad (2)$$

また、概念 A の一次属性の並びを固定し、概念 B の各一次属性を対応する概念 A の各一次属性との一致度の合計が大きい属性の順に並べ替える。

$$B_x = ((b_{x1}, v_{x1}), (b_{x2}, v_{x2}), \dots, (b_{xL}, v_{xL})) \quad (3)$$

この時、概念 A, B の重み比率付き関連度 $DoA(A, B)$ は、

$$DoA(A, B) = \sum_{i=1}^L DoM(a_i, b_{xi}) \times (u_i + v_{xi}) / 2 \times (\min(u_i, v_{xi}) / \max(u_i, v_{xi})) \quad (4)$$

と表される。

4. 観点付き関連度計算方式

人間は異なる2つの概念間の関連性を考える時、何らかの観点に従ってその関連性を判断していると考えられる。例えば、「飛行機と自動車」、「飛行機と鳥」の関連度について考えた時、観点「乗り物」を用いた場合には「飛行機と自動車」、観点「飛ぶ」を用いた場合には「飛行機と鳥」の方が関連度が高くなると考える。

観点付き関連度計算方式では、観点を用いることにより関連度を変化させる。それにより、より人間の感覚に近い関連度計算方式を実現する。

4.1 観点を用いた重み変更手法

概念の持つ各属性の重要さを示す重みの値は常に一定である。しかし観点を用いた場合、観点の影響を考慮し各属性の重要さは変化させる必要があると考えられる。そこで、観点に基づいて各属性の重みを変化させる手法を提案する。重みを変化させることにより関連度を変化させる事が可能になる。属性の重みを変更するために観点との一致度を用いる。概念の属性が観点と同じ属性を持つ場合、その属性は重要であると考えられるため重みを高くする。

観点との一致度を重みとする場合、その重みは観点の持つ属性数に大きく影響される。観点と関係が強いと考えられる属性でも、観点の持つ属性数が少ないと重みが付かないという場合がある。そこで、観点の持つ属性数が少ない場合に対応するために観点の二次属性まで使用する。その方法として観点との一致度を重みとするのではなく、観点の各属性との一致度の総和を重みとする。これにより観点が持つ属性が少ない場合にも対応できる。

概念 A , 観点 V の一次属性を a_i, v_k , 重みを u_i, w_k とし, 属性がそれぞれ L 個, M 個 ($L < M$) とする. 使用属性数を N とし, 観点を V とした時の概念 A の属性 a_i の新しい重み u_i' を一致度を用いて以下のように定義する.

$$u_i' = \sum_{k=1}^N DoM(a_i, v_k) \quad (5)$$

このように, 観点の二次属性まで展開することで, 属性数の少ない観点に対応した計算が可能になると考える.

4.2 各概念と観点の関連度に着目した手法

観点付き関連度計算を行う際に, 観点と概念の関連性は非常に重要であると考えられる. 観点と概念の関連度は, 関係のない観点をを用いた場合には低くなり, 関連性のある観点をを用いた場合には高くなる. そこで, 「概念と観点の関連度」を用いて, 概念の属性の重みを変える手法の補正を行う. 以下にその計算手法を述べる.

概念 A , 概念 B , 観点 V とし, 観点によって重みを変えることにより得られる関連度を $DoA(A, B|V)$, 「概念と観点の関連度」によって $DoA(A, B|V)$ を補正した関連度を $DoA'(A, B|V)$ とし, 以下のように定義する.

$$DoA'(A, B|V) = \sqrt[n]{DoA(A, B|V)^n \times DoA(A, V) \times DoA(B, V)} \quad (6)$$

重みを変更する観点付き関連度計算方式に各概念と観点の関連度を掛け合わせて計算を行う. 「概念と観点の関連度」を掛け合わせて補正を行うことにより, 特に関係のない観点をを用いたときの関連度を下げる効果があると見込める. この補正の強さは n の値を変化させることで調整することが可能である. n の値を大きくすると補正効果は弱まり, 小さくすると補正効果は強くなる.

5. 評価方法

アンケートによって作成した 134 組の評価セットを用いて精度評価を行う. 表 1 に評価セットの例を示す.

表 1 評価セットの例

A	B	C	X	Y
飛行機	鳥	車	飛ぶ	乗り物
マウス	モニター	馬	パソコン	動物
針	時間	糸	時計	裁縫
ー	ー	ー	ー	ー

概念 A とその概念に関連の高い概念 B, C , 概念 A, B 間に関連のある観点 X , 概念 A, C 間に関連のある観点 Y とし, 以下の 2 つの式のどちらも満たす評価データを正解とする.

$$DoA(A, B|X) < DoA(A, C|X) \quad (7)$$

$$DoA(A, B|Y) > DoA(A, C|Y) \quad (8)$$

概念 A, B 間の関連度, 概念 A, C 間の関連度の値の大小関係が観点を変えることにより入れ替わる時に, 観点をを用いた効果があったと判断することが出来る.

6. 精度評価

観点付き関連度計算方式の精度を図 1 に示す. 使用属性数 39 個の時に最大精度 67.2% となった.

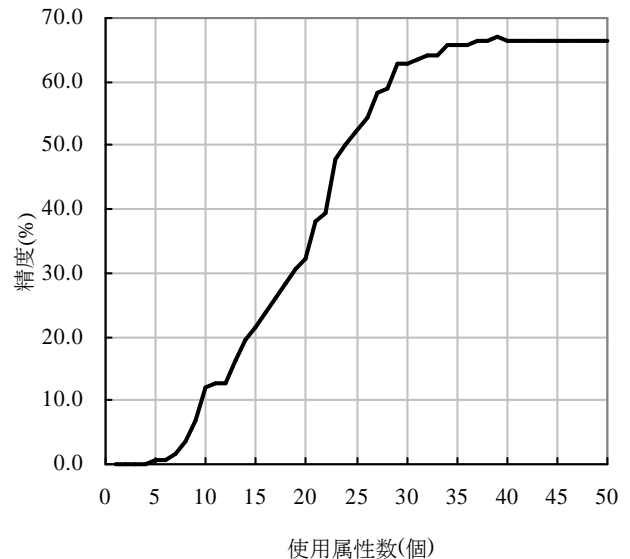


図 1 観点付き関連度計算方式の精度

7. 考察

観点をを用いた重み変更をする際に, 観点の持つ属性数が少ないと重みがかかず, 関連度が得られないという問題が生じると考えられた. しかし, 観点の二次属性まで展開し, また使用属性数を増やすことで適切な新しい重みを与えることが出来たと考える.

式(6)の n の値は 2^5 の時, 最も精度向上が得られた. n の値は小さくするにつれて「概念と観点の関連度」によって強く補正される. $n=2^5$ の場合に得られる関連度の補正効果は非常に強く, $n=0$ とほとんど同じ値になる. $n=0$ とは「各概念と観点の関連度を掛け合わせた値」である. このことから「概念と観点の関連度」は観点付き関連度計算方式にとって非常に重要だということが分かった.

8. おわりに

本研究では, より人間の感覚に近い常識判断ができるシステムの実現のために, 既存の関連度計算方式では常に一定であった関連度を変化させる観点付き関連度計算方式を提案した. 「観点を利用した新たな重み付け」と「観点と概念の関連度」を利用することにより最大精度 67.2% が得られた.

謝辞

本研究の一部は, 科学研究費補助金 (若手研究 (B) 21700241) の補助を受けて行った.

参考文献

- [1] 芋野美紗子, 吉村枝里子, 土屋誠司, 渡部広一, お重み配分に着目した概念ベースの精練, 情報科学技術フォーラム FIT2009, F-062 (2009).
- [2] 山村伊織, 吉村枝里子, 土屋誠司, 渡部広一, お属性数に差のある概念間の関連度計算, 情報科学技術フォーラム FIT2009, E-030 (2009).