

アナロジー推論中における知識類似度の計算方法*

1P-2

中川 賢一郎, 乾 伸雄, 小谷 善行
(東京農工大学 工学部 電子情報工学科)

1 はじめに

フレーム表現を用いたアナロジー推論 [1] を行うときに必ず通るステップとして、目的知識に類似した知識フレームの知識ベースからの検索がある。しかし一般に、同じ意味の知識でも表現が少し変わってしまうと類似のものを見分けるのが困難になる。

本稿では、知識類似度をノードの類似度とパスの類似度によって算出する方法を示し、類似知識問題を人間に解かせる実験との比較を紹介する。

2 知識類似度

本稿では知識を、「二つのノードが一つのパスによってつながった単位」と定義する。そして知識の類似度 R をパス類似度 R_P と二つのノード類似度 R_N を求めることによって導出する (図1)。

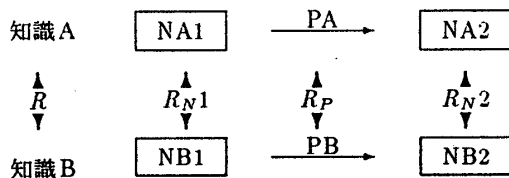


図1 類似度の計算

R は下の式で求めることにする。

$$R = R_P \times \frac{R_{N1} + R_{N2}}{2}$$

パス類似度 R_P は “is_a” と “alike” の類似度が 0.8 といったようにパスの種類に応じて定義される。ノード類似度 R_N を求めるためには比喩の解析が必要となる。詳しくは3章で述べる。

3 比喩関係の類似度算出

比喩が顕現性を用いて解析できる場合は R_{Ken} を R_N とし、それ以外の場合は R_{Te} を R_N とする。

3.1 顕現性を用いた類似度 R_{Ken} の算出

顕現性の導出には、[3] の顕現性の式を利用する。

属性値の確率値 (属性値の注目度) を P_{ij} とし、属性名の重要度 (属性の注目度) を D_i とする。ある属性 i における属性値の数を M_i とおくと顕現性 S_i は次の式で表す。

$$S_i = D_i \times (1.0 - H_i)$$

このとき H_i は、

$$H_i = \begin{cases} \frac{\sum_{j=1}^{M_i} P_{ij} \log_2 \frac{1}{P_{ij}}}{\log_2 M_i} & (M_i \geq 2) \\ 0 & (M_i = 1) \end{cases}$$

となる。ここでの P_{ij} と D_i の値は調査実験により決めた数値である。

例として「雲」の顕現性を求める。雲の「色」としての D_1 を 0.45、「色が白」としての P_{11} を 0.78、「色が黒」としての P_{12} を 0.13、「色が灰」としての P_{13} を 0.08 とすると、

$$H_1 = \frac{0.280 + 0.383 + 0.292}{1.585} = 0.603$$

$$S_1 = 0.45 \times (1.0 - 0.603) = 0.18$$

となり、「雲」における「色が白」の顕現性は 0.18 である。

概念 A と概念 B が等しい属性値 Z を持っていたとする。概念 A と概念 B の属性値 Z における顕現性をそれぞれ K_A と K_B とおくと、その属性値に対する一致度 $C(Z)$ は、

$$C(Z) = K_A \times K_B$$

の式で表す。結局、概念 A と概念 B の顕現性を用いた類似度 R_{Ken} は、概念 A と概念 B の持つ同じ属性値全ての $C(Z)$ を求め、その最大値とする。

*A Method of Calculating Similarity in Analogical Reasoning,
Kenichirou NAKAGAWA, Nobuo INUI, Yoshiyuki KOTANI
Tokyo University of Agric. and Tech., Dept. of Computer Science

3.2 適合度を用いた類似度 R_{Te} の算出

「車がガソリンを飲む」といった活喩などの比喩解析には、顕現性を用いた解析では対応しにくい。そこで「適合度」を用いて解析する方法を示す。「適合度」はある空間にある単語がどれだけ自然に入り込めるかを表した数値である。

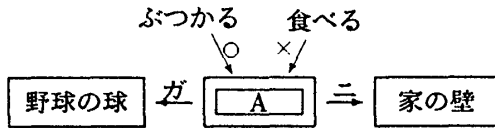


図2 空間Aのイメージ

図2のような空間Aと動詞Bがあり、動詞Bの空間Aでの適合度をTとする。このときに、Aにつながるノード N_1, N_2, \dots, N_n (図2の場合では「野球の球」や「家の壁」とBの用いられる事例を持つノードが知識ネットワーク中で S_1, S_2, \dots, S_n の距離にあるとすると、

$$T = \frac{\sum_{i=1}^n 0.8^{S_i}}{A \text{ が持つノード数}} - 0.1 \times \frac{\text{対応しないノード数}}{B \text{ が持つノード数}}$$

の式で適合度を求める。0.8と0.1は類似度減点用の定数であり、筆者が経験的に定めた。

動詞1と動詞2の適合度を用いた類似度 R_{Te} は、動詞1の入るべき空間における動詞2の適合度と考えることができる。

4 知識類似度の算出

4.1 実験方法

問題知識と四つの知識を与え、その四つの類似度を算出するという方法で行った。また、人間にも同じ問題を与え、一番似ていると思われる知識を選んでもらった。問題数は30問で知識ベースからランダムに生成した。

4.2 例題

問題：「太郎が飲む」

候補知識 1. 「球を打つ」 2. 「犬が走る」
③. 「猿が食べる」 4. 「太郎は人間」

表3 知識の類似度の算出例

順位	知識	類似度
1	猿が食べる	0.608
2	犬が走る	0.35
3	太郎は人間	0.05
4	球を打つ	0.031

4.3 結果

実際に問題30問を解いた結果を表4に示す。

表4 システムの出力した順位に対する人間の選んだ答えの分布

システムが出力した順位	人間の選んだ答えの数	比率
1	12	0.4
2	11	0.37
3	3	0.1
4	4	0.13

人間の選んだ答えは本来1位に集まるべきなのだが、結果にばらつきが出てしまっているのがわかる。これは“似た知識を選ぶ”という曖昧な目的しか被験者に与えなかったことと、実験で用いた知識ベース(知識数で300)が小さすぎたことが結果に影響を与えているのではないだろうか。

5 おわりに

二つの知識の類似度を求める際に、ノード類似度とパス類似度を求めることによって算出する方法を示した。

この方法により、比喩などを含んだ表現方法の違う知識でも柔軟性をもって類似度を計算することができる。

参考文献

- [1] 伊藤ほか：「フレーム表現で実現したアナロジー推論の一手法」, 第42回全国大会講演論文集(3), pp.3 - 212, 213 (1990)
- [2] 有馬ほか：「アナロジーの数理的研究」, 情報処理, vol 34, No 5, pp.547 - 557 (1993)
- [3] 内海ほか：「文脈の中の比喩-関連性による隠喩理解モデル」, 自然言語処理, 101 - 13, pp.97 - 104 (1994)