

1 F-6 アバウト推論における多観点概念ベース

笠原 要 松澤 和光 石川 勉

NTT 情報通信網研究所

1 はじめに

知識が不完全であっても概略的に近似解を求められる推論方式として、「アバウト推論」の研究を進めている [1]。その特徴の1つは、推論システムに常識的な知識を内蔵して、不完全な知識を補充/代替することにある。本稿では、そのために必要な概念知識源である「多観点概念ベース」の実現方式を提案する。

2 観点に基づく概念間関係

多観点概念ベースとは、知識を構成する基本単位である「概念」を内蔵し、概念間の関係(類似, 被覆等)を判別するシステムである。その際に問題となるのは、概念間の関係が一樣ではなく、用いられる場面や対象とする問題等の「背景となる知識」によって動的に関係が変化することである。

例えば、概念『馬』に対して、概念『自動車』と『豚』のどちらが類似しているかを判別することを考える(図1)。

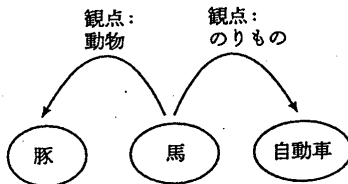


図1 観点による類似関係の変化

もし動物に関する問題を扱っている場合、『馬』は『豚』と似ているが、乗物に関する問題であれば『自動車』の方が似ていると考えられる。このような概念間の関係を考える際の背景知識を1つの概念として捉え、「観点」と呼ぶことにする。(属性継承及び比喩関係についての同様な考え方として「視点」の研究がある [2]。)多観点概念ベースでは、アバウト推論に

おいて生成されるこのような観点に基づいて、多様な概念間の関係を判別する必要がある。

3 観点変調方式

本稿では、概念間の関係の1つである類似性を取り上げ、観点に基づく類似性判別方式として観点変調方式を提案する。この方式の特徴は、概念間の類似判別を行なう前に、観点に基づいて概念の内容を変調させる点にある。これは、人間が概念を用いて思考する時に、概念に含まれる内容すべてを用いるのではなく、観点によって概念の一部分だけを切りとって利用していることをモデル化したものである。

3.1 概念の表現

多観点概念ベースに含まれる概念 G は、概念の特徴を表す属性 p_i と、その属性 p_i が概念 G においてどれだけ重要であるかを表す重要度 q_i ($0 \leq q_i \leq 1$) の対の集合により表現する。

$$G = \{(p_1, q_1), (p_2, q_2), \dots, (p_n, q_n)\} \\ = \{(p_i, q_i)\} \quad (i = 1, \dots, n) \quad (1)$$

概念の表現法として他にも種々考えられるが、例えば概念間の関係を固定的に決定したネットワーク構造として蓄積する方法では、多様な観点における概念間の関係の記述が困難となり、大規模化に対応できない恐れがある。従ってここでは、各概念を独立した構造で概念ベース内に保持する形式とした。また、具体的な属性、重要度は、辞書などの文書情報より完全自動で獲得することを想定している。この場合自動的な手段により還元的な属性を獲得するのは不可能であると考え、文書から得られた単語をそのまま属性とみなすことにする。

3.2 概念の規格化

類似性の判別は属性同士の比較によって行なうが、各属性は互いに独立であるとは限らないので、属性の意味を考慮して比較を行なう必要がある。このため、属性を分類するための分類体系を定義する。分類 c_j は、同様な意味を持つものとして分類された属性の集合をあらわす。この分類体系には、既存のシソーラス等を用いることができる。

$$c_j = \{p_{j1}, p_{j2}, \dots, p_{jl}\} \quad (2)$$

$$(j = 1, \dots, l)$$

この分類 c_j の集合が分類体系 C である。

$$C = \{c_1, c_2, \dots, c_k\} \quad (3)$$

次に、分類体系 C を用いて概念の規格化表現 $\hat{G}$ を以下の通り定義する。

$$\hat{G} = \{(c_j, Q_j)\} \quad (j = 1, \dots, l) \quad (4)$$

$$Q_j = \sum_{p_i \in c_j} q_i / \sum_{i=1}^n q_i$$

なお、属性間の分類を概略から詳細まで段階的に変化させ得る木構造の分類体系を用いると、概念間の関係を判別する時の精度を変化させることが可能となる。

3.3 観点による概念の変調

類似性を判別する際に種々の観点考えられるが、ここでは、多観点概念ベースに含まれる任意の概念が観点となり得ると考える。観点の規格化表現 $\hat{K} = \{(c_j, Q'_j)\}$ に基づいて、対応する分類ごとに規格化された概念 $\hat{G} = \{(c_j, Q_j)\}$ 中の重要度を変調させることにより、変調概念 G^K を得る。

$$G^K = \{(c_j, Q_j^K)\} \quad (Q_j^K = Q_j \cdot M(Q'_j)) \quad (5)$$

式(5)における M は変調の度合を決定する変調関数であり、例えば図2のような閾値関数が考えられる。この関数は、観点中の一定値以上の重要度を持つ分類と同じ概念の分類の重要度のみを残す作用がある。

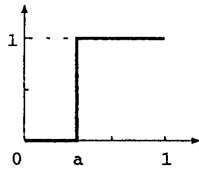


図2 変調関数の一例

3.4 類似判別

2つの概念 G_A, G_B 間の観点 K に基づく類似判別 $R_C(G_A, G_B, K)$ は、概念の規格化と観点による変調を行なった変調概念 $G_A^K = \{(c_j, Q_{A_j}^K)\}$, $G_B^K = \{(c_j, Q_{B_j}^K)\}$ を元にして以下のように定義する。

$$R(G_A, G_B, K) = R'(G_A^K, G_B^K) \quad (6)$$

R' は、観点が介在しない場合の単純な類似判別を表す。ここでは、 R' として、

$$R'(G_A, G_B) = \sum_{j=1}^l \sqrt{Q_{A_j}^K \cdot Q_{B_j}^K} \quad (7)$$

を考えた。 R' は、以下の要請を満たす関数の一つである。

$$0 \leq R' \leq 1$$

$$R'(G_A, G_A) \equiv 1$$

$$R'(G_A, G_B) \leq R'(G_C, G_B) \text{ ならば、概念 } G_B \text{ に対して } G_A \text{ よりも } G_C \text{ の方が類似している。}$$

4 実験結果

例として、概念『馬』に対する概念『豚』と概念『自動車』の類似度 R を、異なる観点『動物』と『のりもの』について算出した。ここで、概念の属性は国語辞典の説明文を形態素解析することにより獲得した。分類体系として、3つの階層 (C_1, C_2, C_3) をもったシソーラスである角川類語辞典を用いた。これにより3つの判定精度をもつ類似度が得られる。なお観点による変調では、最も概略な精度で規格化した観点をを用いた。その結果を表1と表2に示す。

分類精度	G=『豚』	G=『自動車』
概略 (C_1)	0.38	0
中程度 (C_2)	0.38	0
詳細 (C_3)	0.15	0

表1 $R(G, \text{『馬』}, \text{『動物』})$

分類精度	G=『豚』	G=『自動車』
概略 (C_1)	0.16	0.44
中程度 (C_2)	0.13	0.29
詳細 (C_3)	0	0

表2 $R(G, \text{『馬』}, \text{『のりもの』})$

観点『動物』においては、どのような分類精度においても、『馬』に対して『自動車』よりも『豚』が類似していることが示される。一方『のりもの』を観点とした場合、精密な判定では両者の類似の度合に違いがないが、判別を概略的に行なうにつれて『自動車』の類似度が高くなっている。

5 おわりに

多観点概念ベースの実現法として、観点に基づいて概念間の類似を複数の精度で判別する方法を提案したが、さらに類似以外の概念間の関係(被覆など)についても定式化を行なう予定である。

参考文献

- [1] 松澤他。“アバウト推論：柔らかな知識処理の提案”，本大会予稿集，(1992)
- [2] 田中他。“比喩を含む言語理解における視点の役割”，情報処理学会自然言語処理研究会，NL 73-7, pp.51-58, (1989)