

量的な判断常識を備えた人工知能

2N-2

— 概念関係の獲得 —

堀口 賞一 小濱 千恵 飯田 敏幸

NTT情報通信網研究所

1 はじめに

人間のような柔軟な理解や判断能力を計算機に持たせることを目的として、常識知識を用いて量的な判断を行なう人工知能の研究を行なっている [1]. 量的な判断とは量に関する曖昧な表現を意味理解して具体化 / 数値化することである。意味理解を実現するには、単語が表す概念を計算機上で実現する必要がある。計算機上での概念の実現は、『ある概念に対し、それに関係する概念を関係づけること』と捉えることができる。単語間の関係をテキストから自動獲得する研究 [2] の成果等を利用し、概念関係をシソーラスに基づきテキストから自動的に獲得することを検討している [3].

シソーラス上の意味を考慮して分類された単語の集まりを概念とし、単語間の関係を各単語を含む概念の間の関係に変換する。このとき、多義の単語があるために、テキストから得られた単語をどの概念に変換するかが問題となる。また、単語の出現頻度の偏りが大きいために、全ての概念に対し概念関係を抽出するには膨大なテキストが必要であるという問題がある。

本稿では、少ない単語間の関係から多くの概念関係を得ると共に、単語の多義性を解消する手法について述べる。また、本手法を量的な判断に必要な概念関係の1つである、判断対象を表す概念(対象概念)とその対象の性質を数値に対応づける尺度を表す概念(量属性)との概念関係(対量関係)の獲得に適用した結果について述べる。

2 概念関係の獲得

多義のない単語間の関係から得られる少ない概念関係を用いて、前述した2つの問題を解決する手法について具体的に述べる。

2.1 概念関係の導出(挟み打ち手法)

以前、シソーラスに基づき上位概念の持つ概念関係を下位へ継承することにより、多くの概念関係を導出することを試みた [3]. 概念 α と概念 x との概念関係がある場

合に $C(\alpha, x)$ と書くと、概念関係を継承するとは、例えば、 $C(\$具体物, \$重量)$ を“具体物”の下位の“金属”に対して、 $C(\$金属, \$重量)$ とすることである。しかし、継承した結果、多くの適切でない概念関係が導出されるという問題があった。例えば、図1で示す構成のシソーラスを考える。 $C(\$具体物, \$重量)$, $C(\$具体物, \$熱伝導率)$ の2つの概念関係を下位へ継承することにより、 $C(\$道路, \$重量)$, $C(\$人間, \$熱伝導率)$ などの適切でない関係が導出される。そこで、概念関係をシソーラスに基づき部分的に継承させることを考える。

上位下位関係にある2つの概念は多くの性質を共有していることに着目する。概念 α の下位関係を辿って得られる概念の集合を $K(\alpha)$ とする。 $\beta \in K(\alpha)$, $r \in K(\beta)$ のとき、

$$C(\alpha, x), C(r, x) \text{ ならば } C(\beta, x) \quad (1)$$

という仮説に基づき概念関係を部分的に継承させる。例えば、同じシソーラスを考えると、 $C(\$具体物, \$熱伝導率)$ でかつ、 $C(\$鉄, \$熱伝導率)$ であるとき、両者に挟まれる“金属”だけに対して $C(\$金属, \$熱伝導率)$ が導出される。

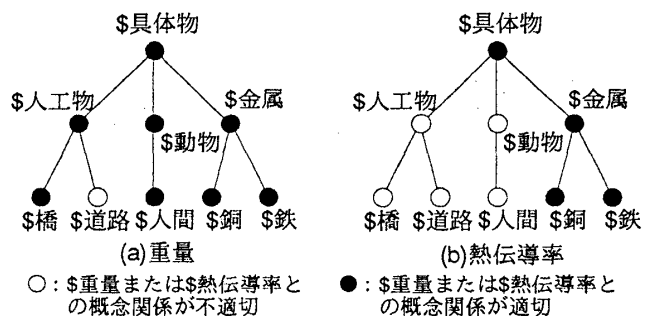


図1: 概念関係の部分的継承

2.2 多義性の解消

シソーラス上の概念同士の距離を、ある概念から別の概念までに辿らなければならない上位下位関係の数と定義する。例えば、図2において、 $\$bicycle$ と $\$自動車$ との距離 $D(\$bicycle, \$自動車)$ は2である。

多義のある単語から複数の概念関係候補 $C'(a_i, x)$ ($i=1, 2, \dots, n$) が考えられるとすると,

$$C(\beta, x), D(a_{i_0}, \beta) = \min_{i \in I} D(a_i, \beta)$$

ならば, $C(a_{i_0}, x)$ (2)

但し, $i_0 \in I = \{1, 2, \dots, n\}$

という仮説に基づき多義性を解消する. 例えば, 「軽い自転車」中の単語“自転車”は \$bicycle と \$cycling の2つの概念に含まれる. この時, 概念関係 $C(\$自動車, \$重量)$ があれば, $D(\$bicycle, \$自動車)$ が最小値であるので, 概念関係 $C(\$bicycle, \$重量)$ となる.

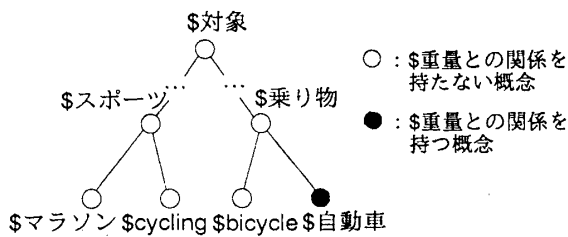


図 2: 多義性の解消

3 対量関係の獲得

3.1 獲得対象の対量関係

獲得すべき対量関係は, 大多数の人が判断対象と結び付けて容易に想起できる量属性との関係とする. 例えば“人間”に対しては, “身長”, “体重”などであり, “熱伝導率”, “価格”など容易に想起されないものは含まない.

このように, 容易に想起できる対量関係は, 辞書の説明文や例文, 新聞記事などに出現し易いと考え, それらから獲得を行う. この対量関係を持つことで, どのような尺度で判断対象の性質を具体的な数値へ対応づけるべきかが分かるため, “高い人”という入力に対して“価格”ではなく“身長”が高い人であることが分る. 更にこの対量関係を利用すると提案型の知識処理が考えられる. 1 応用例として DB 設計を考えると, “人”に関するテーブルには“身長”, “体重”などの項目を設けるべきであるという提案をすることなどが考えられる.

3.2 評価

重量および熱伝導率との単語間の関係 50 組ずつを辞書の例文や説明文から抽出し, 機械翻訳用に開発されたシソーラス [4] をベースとした約 160,000 のカテゴリとなるシソーラスへ各々の関係を対応づけることで対量関係を獲得した. 表 1 に, 多義のない単語間の関係から得られる対量関係を (1) 継承しない場合, (2) 下位へ継承した場合, (3) 抜き打ち手法を適用した場合の関係数および正当率を示す. (2) における正当率は, 重量との対

量関係では比較的高いが, 熱伝導率との対量関係では低い. これにより, 量属性には継承範囲の広いものと狭いものとの 2 タイプがあり, (2) の手法が, 継承範囲の広い量属性には有効であるが, 継承範囲の狭い量属性には適していないことが分る. しかし (3) では, 重量と熱伝導率とのどちらの関係も適切な対量関係を導出できることが分る.

表 1: 対量関係数および正当率

	重量		熱伝導率	
	関係数	正当率	関係数	正当率
(1)	18	100.0	18	100.0
(2)	4,632	83.4	4,419	54.3
(3)	48	100.0	45	100.0

また, 単語間の関係各 50 組のうち 36 組ずつが多義性のある単語であった. 表 2 に, この 36 組ずつの単語間の関係に対し, (1)(2)(3) の対量関係を用いて多義性解消を行なったときの多義性の解消率を示す. 表 1 と表 2 から, 正当率の高い対量関係を予め多く蓄積することにより, 正しく多義性を解消できることが分る.

表 2: 多義性の解消率

	重量	熱伝導率
(1)	53.1	50.0
(2)	84.4	68.8
(3)	84.4	78.1

4 おわりに

本稿では, 単語間の関係から概念関係を得る為に, 概念関係導出手法と単語の多義性解消手法を提案した. 前者では少しの概念関係から適切な概念関係を導出できること, 後者では正当率の高い概念関係を予め多く蓄積するほど正しく多義性を解消できることを示した.

参考文献

- [1] 小濱他: 常識人工知能とそのデータベース処理への適用, 電子情報通信学会技術研究報告, AI93-22, 1993.
- [2] 長尾: 自然言語で表現された内容を知識ベースに変換し, 知識ベースを構築する研究, システム技術開発調査研究, 2-R-2(9), 1991.
- [3] 堀口他: 量的な判断常識を備えた人工知能 — 量的な属性の自動獲得手法 —, 情報処理学会第 46 回全国大会, 9D-4, 1993.
- [4] 池原他: ALT-J/E における翻訳辞書の構成と意味解析型翻訳機能, 1991 年電子情報通信学会春季全国大会, SD-5-2, 1991.