

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 —

9 D - 4

堀口 賞一 小濱 千恵 島田 茂夫 飯田 敏幸  
NTT 情報通信網研究所

1 はじめに

人間の持つ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、量的な判断常識を備えた人工知能の研究を行っている<sup>[1]</sup>。量的な判断常識は、(1)対象を表す概念(対象概念)、(2)量の尺度を表す概念(量属性)、(3)比較の程度を表す概念(比較概念)、(4)比較の程度を修飾する概念などから構成される<sup>[2]</sup>。

量的な判断常識の構築において、対象概念および対象概念と量属性との関係(対量関係)は大規模となるため、対象概念として既存ソーラスを用い、テキストから抽出された対量関係を対象概念と量属性間に張ることを試みた。しかし、テキスト中に含まれる対量関係が非常に少ないために、大規模な対量関係構築の見通しが得られなかった。そこで、既存ソーラスを利用し、抽出された対量関係を継承することにより対量関係の導出を考える。しかし、既存ソーラスは必ずしも量の観点から階層化されていないので、量の観点から妥当な階層だけを使った対象概念の階層化により対量関係を継承させる手法を提案する。

本手法は、1)対量関係の抽出、2)対象概念の階層化、3)対量関係の整理の手順からなる。また、本手法を辞書の例文に適用した結果を示す。

2 対量関係とは

対量関係とは『対象概念である'山'が、量属性として'標高'を持つ』という知識であり、「標高3000メートルの山」などのテキストから抽出できる。このように抽出された対量関係と予め設定されている比較概念と量属性の関係を利用することにより、例えば、「高い山」から『高いのは'山'の'標高'である』ことを推論(補填)できる。この量属性の補填は、量の観点からの対象理解の1つである。

3 対量関係を表す語の組の抽出

対象概念を表す語(対象概念語)と量属性を表す語(量属性語)の組をテキストから抽出する。但し、各々の量属性はそれ自身を表す語の集合とし、量属性全体はそれらの量属性を階層化したものとして、予め人手により構築しておく。

抽出の手順は、以下の通りである。1)形態素解析により、テキストを文節および形態素に分割する。2)1文節中に自立語が1つしかない場合は、それが量属性に登録されている語ならば量属性語とし、登録されていない語で名詞であれば対象概念語とする。また1文節中に自立語が複数ある場合は、それらを複合語とみなし、末尾語が量属性に登録されている語ならば量属性語とし、登録されていない語で名詞であれば対象概念語とする。3)構文解析を行なう。4)抽出された対象概念語および量属性語が、係り先、係り受けの関係であるならば、対量関係を表す語の組として抽出する。具体例を図1に示す。

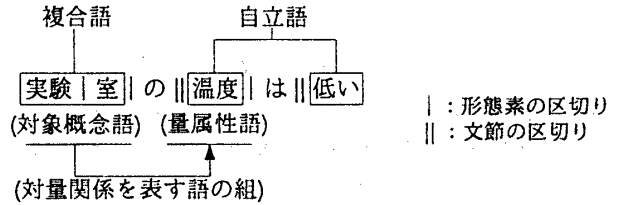


図1: 語の組の抽出の具体例

4 量の観点からの対象概念の階層化

人間には、知識を組み合わせることで別の知識を導出する能力がある。例えば、『'車'に'排気量'がある』という知識を得ることで、『'車'の下位の'バイク'にも'排気量'がある』という知識が導出できる。これは、予め人間が対象の上位下位関係を知っているためである。

計算機上では、対象の上位下位関係を示すものとしてソーラスが開発されてきた。しかし、既存ソーラスの階層を単純に対象概念の階層として利用すると、『'車'の下位の'自転車'にも'排気量'がある』といった誤った継承をする危険性がある。

そこで、以下で述べる量の観点からの対象概念の階層化では、対量関係の継承する範囲を限定することで、誤った継承を少なくする。

4.1 量属性の階層の深さに基づく対象概念のまとめ上げ

量属性全体は、図2に示すような階層構造をなす。最上位の階層を level0、level0 から n だけ下位の階層を leveln とする。

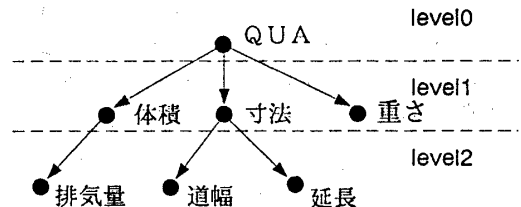


図2: 量属性の階層構造

ある量属性  $r_{ni} \in R$  ( $R$  は量属性の全体集合) に着目しまとめ上げた対象概念の集合  $G_{r_{ni}}$  を、次のように定義する。

$$G_{r_{ni}} = \{t \in T | r_{you}(t, r_{mj}) \text{ for } \forall r_{mj} \in D(r_{ni})\} \quad (1)$$

但し、 $T$  は対象概念の全体集合、 $r_{ni}$  は leveln の量属性 ( $i = 1, 2, \dots$ ) または levelm ( $m < n$ ) の下位概念を持たない量属性、 $D(r_{ni})$  は  $r_{ni}$  の下位関係を辿って得られる概念の集合 ( $r_{ni}$  自身を含む)、および  $r_{you}(t, r_{mj})$  は対象概念  $t$  を表す語  $w_i$  と量属性  $r_{mj}$  を表す語  $w_{r_{mj}}$  との組が抽出されたことを示す関係(対量関係)である。例えば、図2において、level0の  $G_{QUA}$  は、 $D(QUA) = R$  の要素と対量関係にある対象概念の集合である。また、level1の  $G_{寸法}$  は、 $D(寸法) = \{ '寸法', '道幅', '延長' \}$  の要素と対量関係にある対象概念の集合である。

leveln での対象概念のまとめ上げとは、 $G_{r_{ni}} (i = 1, 2, \dots)$  を求めることである。

4.2 既存ソーラスを利用した階層化

既存ソーラスを利用して  $G_{r_{ni}}$  を階層化すると、 $G_{r_{ni}}$  に含まれる2つの対象概念が、既存ソーラスにおいて上位下位関係にある時、その上位下位関係を2つの対象概念間に生成することである。また、level  $n$  での階層化とは  $G_{r_{ni}} (i = 1, 2, \dots)$  を階層化することである。

$G_{r_{ni}}$  を階層化した時に、対象概念  $t \in G_{r_{ni}}$  に継承される量属性の集合  $k_{r_{ni}}(t)$  は、次式で表される。

$$k_{r_{ni}}(t) = \{r_{mj} \in R | r_{you}(t_i, r_{mj}) \text{ for } \forall t_i \in A(t) \cap G_{r_{ni}}\} \quad (2)$$

但し、 $t_i$  は対象概念、 $A(t)$  はソーラスにおいて対象概念  $t$  の上位関係を辿って得られる概念の集合 ( $t$  自身を含まない) である。また、level  $n$  での対象概念  $t$  に継承される量属性は、 $\cup_i k_{r_{ni}}(t)$  である。

例えば、あるテキストから  $r_{you}$ (建造物, 重さ),  $r_{you}$ (道路, 道幅),  $r_{you}$ (地下道, 延長) が抽出され、ソーラスでは'建造物'の下位概念が'道路'、'道路'の下位概念が'地下道'である時、level 0 では  $G_{QUA} = \{ '建造物', '道路', '地下道' \}$  に対して階層化するので、'建造物'が'道路'や'地下道'の上位概念のままであり、知識利用時に  $k_{QUA}$ (道路) = { '重さ' },  $k_{QUA}$ (地下道) = { '重さ', '道幅' } という誤った継承が起きてしまう。一方、level 1 では  $G_{重さ} = \{ '建造物' \}$ ,  $G_{道幅} = \{ '道路', '地下道' \}$  の各々を階層化するため、'建造物'と'道路'や'地下道'との間には上位下位関係が生成されず、'道路'と'地下道'の間のみ上位下位関係が生成され、 $k_{道幅}$ (地下道) = { '道幅' } という正しい継承だけが起きる。level 2 では  $G_{重さ} = \{ '建造物' \}$ ,  $G_{道幅} = \{ '道路' \}$ ,  $G_{延長} = \{ '地下道' \}$  となり、上位下位関係が生成されないため、継承は起きない。

4.3 知識量と知識の正しさ

上位の level で階層化するほど、知識利用時に継承される対量関係数は増えるが、誤った継承をする危険性は高くなる。このように、知識量と知識の正しさはトレードオフの関係にある。そこで、level ごと知識量および知識の正しさを調べる必要がある。level  $n$  での知識量  $K_n$  を、知識利用時に継承される対量関係と抽出された対量関係との和とし、次式で表す。

$$K_n = \sum_i \sum_{t \in G_{r_{ni}}} (|k_{r_{ni}}(t)| + P(t)) \quad (3)$$

但し、 $| \cdot |$  は集合の要素数、 $P(t)$  は、対象概念  $t$  に対して抽出された対量関係数である。また、知識の正しさを適合率とし、抽出された対量関係および知識利用時に継承される対量関係の中で、人間が正しいと判断する対量関係の割合で表す。level ごと知識量および知識の正しさを調べることで、適切な level を求め、その level で階層化することで、対量関係を継承させる。

5 継承に基づく対量関係の整理

以下の手順で、継承を利用して対量関係を整理する。

- (1) 対象概念から見て、より下位の対量関係を優先する。
  - (2) 冗長な対量関係を削除する。
  - (3) 上位の対象概念を生成し、共通な対量関係をまとめる。
  - (4) 対象概念の全下位概念に共通な対量関係を持ち上げる。
- (1) は量的な判断常識として詳細な知識を必要とするため、(2),(3) は知識の内容を変えずに対量関係を削減するため、(4) は知識に一般性を持たせるために行なう。特に(4)ではテキストから抽出されなかった対量関係が生成されることになる。

6 評価

自然言語処理用に開発されたソーラス<sup>[3]</sup>を既存ソーラスとし、本手法を用いて、辞書の例文から量的な属性の自動獲得を行なった。その結果得られた level と知識量、知識の正しさおよび対量関係の削減率との関係をそれぞれ図3、図4、図5に示す。尚、対量関係の削減率とは、テキストから抽出された対量関係数に対する、5章で述べた手順により削減された対量関係数の割合である。

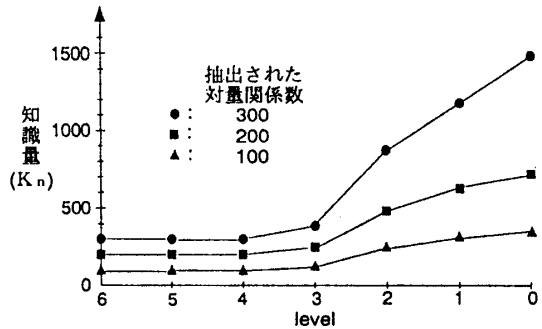


図3: level と知識量との関係

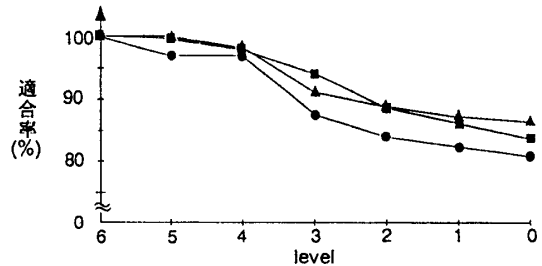


図4: level と知識の正しさの関係

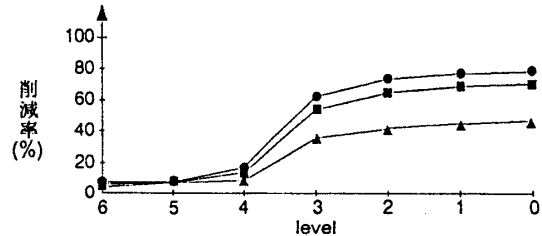


図5: level と対量関係の削減率との関係

7 おわりに

量的な属性の自動獲得において、対量関係をテキストから抽出し、既存ソーラスを利用して量の観点から対象概念を階層化した後、その階層に基づき対量関係を整理する手法を提案した。本手法により、知識利用時の誤った継承を抑えることができる。また、本手法を辞書の例文に適用した場合の level と知識量、知識の正しさおよび対量関係の削減率との関係を示した。今後は、データ量を増やし適切な level を求めると共に、獲得した対量関係の正しさを量的な判断常識人工知能の中で評価する予定である。

参考文献

- [1] 飯田他: 量的な判断常識を備えた人工知能 — 知識と能力 —, 情報処理学会第 45 回全国大会, 1H-11, 1992.
- [2] 小濱他: 量的な判断常識を備えた人工知能 — 知識表現モデル —, 情報処理学会第 43 回全国大会, 5E-10, 1991.
- [3] 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.