

常識スケールを用いた未知語に対する量的判断メカニズム

1 M-7

菊山善久 渡部広一 河岡司

同志社大学大学院 工学研究科

1 はじめに

コンピュータが人間の意図を理解し、より柔軟な判断を行うための基本要素技術として、言葉の論理関係から物の大小、長短、軽重などを適宜に判断できる量的判断メカニズムについて研究を進めている。本稿では、入力されたいかなる語の対に対して適切な比較判断を可能とするメカニズムの構成について述べる。

具体的には、人間が持つ量に関する常識的なスケールを知識ベースとして設定し、それを用いた量的判断メカニズムの構成方法、さらに未知語に対しては、電子化国語辞書から作成した概念ベース（4万語）と電子化事典などを段階的に活用することで、判断を可能とする手法を提案する。

2 量語スケールの設定と量的知識ベース

数万語以上もあるとみられる物を表現する語の全てについて種々の量的属性を定義するのは不可能である。そこで我々は「量語スケール」という考え方を導入する。これは人間が常識的に持っているであろう量に関する物を代表語として約1400語定義し、それらの語を約50の適切な量値レベル（目盛り）に振り分けるという手法で構成されている。以下に常識スケールの設定にあたり考慮した事項を記す。

1. ミクロの世界から宇宙のレベルまで広い領域をカバーする。
2. 人間に身近な物の量値範囲では、よりきめ細かくする。
3. 日常語のカバー率を高くする。
4. 目盛りは人間の感覚に近い型で与える。
5. 同じ目盛りには関連度の低い語を配置する。

以上のようなことを考慮に入れ、作成した量語スケールの一部を図1に示す。

大きさの量語スケールとして選出した代表語に関し種々の量的属性をデータベース化した量的知識ベースを構築した。現在、1400語の代表語に対して、大きさ、材質、形状などの量的属性を、さまざまな量的判断に対応できるように形で定義してある。

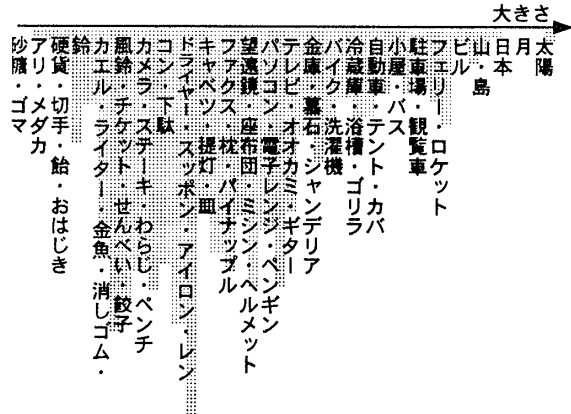


図1 大きさの量語スケール（一部）

3 量的判断メカニズム

現時点での量の判断モデルとしては、二つの比較語と量語（大きい、重い、高い etc）を入力とし、判断結果として程度語（かなり、少し etc）と判断した結果の語を得るという形式になっている。以下に量の判断モデルの例を示す。例：入力（自動車、飛行機/大きい）

↓
出力（かなり*飛行機）

このモデルに従い、前述の量語知識ベースを活用することで、入力された質問の意味を理解し、状況に応じた適切な処理を施し柔軟な判断を行うのが、量的判断メカニズムの役割である。以下に量的判断メカニズムの構成を示す。

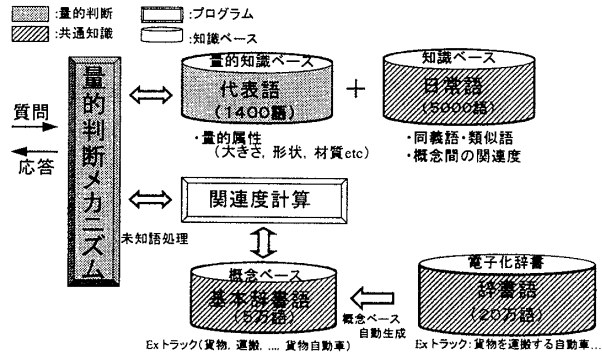


図2 量的判断メカニズムの構成

用語の定義
 知識ベース：日常語として5000語定義されている。この5000語には代表語1400語のすべてが含まれている。さらに、5000語同士の相互関連度と類似語同義語などが入っている。
 概念ベース：基本辞書語として5万語定義されている。各基本辞書語に対して、それに関連するような語が複数定義されている。
 例：トラック（乗物、運搬、…、貨物自動車）

Quantitative Judgement Mechanism to unknown objects by using Scale of Common Knowledge
 Yoshihisa Kikuyama, Hirokazu Watabe and Tsukasa Kawaoka
 Graduate School of Engineering, Doshisha University

4 未知語処理

いかなる物に対しても比較判断を可能とするために、代表語以外の語（未知語）が比較対象語として入力された場合でも、独自に作成した概念ベースと電子化辞書を段階的に活用することで判断を可能とする処理方法について説明する。基本的な考え方としては、未知語と最も関連の高い代表語をさまざまな手法を用いて選び出し、その量的属性を使用することで判断を可能にするということである。未知語処理の手順を以下に記す（図2参照）。

0. 比較語が代表語に存在しない時、その比較語を未知語とする。
1. 知識ベースの同義語・類似語を検索し、未知語と代表語を結びつける語が存在しないか調べる。存在すれば、その代表語の量的属性を使用し、判断を行う。存在しなければ、次の処理へ進む。
2. 知識ベースの日常語の中に未知語と一致する語が存在しないか検索する。存在すれば、相互関連度マトリックスより最も類似の代表語が即座に決定し、その量的属性を使用することで、未知語の判断が可能となる。存在しなければ、次の処理へ進む。
3. 概念ベースの基本辞書語の中に未知語と一致する語が存在しないか検索する。存在すれば関連度計算(文献[2])を行い、最も類似の代表語を決定する。今回の実験で使用した手法を以下の図3に示す。

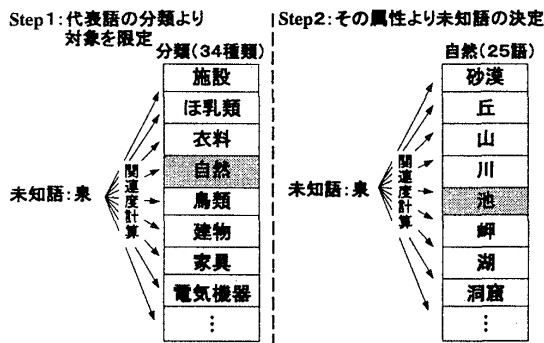


図3：関連度計算による未知語の決定

関連度計算による未知語の決定方法についてはさまざまな手法が考えられ、今回取り上げたのはその中の一つであり、これが最適な手法であるとは考えていない。最適な手法の検討についてはこれから行う予定である。

5 実験結果と考察

代表語の選択及び未知語処理の評価を行うためにある一定の条件のもとで選出した語に対してどの程度の割合で認識が可能になるのか実験を行った。以下にその実験結果と考察を述べる。抽出した語は、量に関する物と限定し、無作為に抽出した。なお今回の実験では、図2の電子化辞書は使用していない。

表1 実験結果①

抽出方法	A	B	C
*和英三辞典 (247語)	83%	8(1)%	9%
国語辞典 (100語)	53%	17(2)%	30%

*一人歩き自由自在の一部である和英三辞典の中から抽出した

A: 代表語と完全一致した割合

B: 未知語処理を行うことで、認識できた語の割合 (そのうち非常識的な判断結果となった割合)

C: 概念ベースに未知語が存在せず、関連度計算を行えなかった割合

表1より和英三辞典のAでは、83%の認識率を得た。これより代表語の選定がうまく行えたと考えている。国語辞典のAの結果では、53%と低い値であるが、未知語処理を行うことで70%まで認識率を上げることができた。Cの結果については概念ベースが精練されていくことにより減少していくと考えられる。次に、より具体的な未知語処理の評価を行うために、以下のような実験を行った。乗り物に関する未知語を入力し、それが代表語の中に存在する乗り物を表す語 (飛行機, 気球, 車, 船, 電車, 汽車, 自動車, 単車, 戦車, 人力車, 潜水艦, 三輪車, 自転車, 馬車, 筏) の中のどの語と最も関連度が高いと判断するか確かめた。その結果を以下に示す。

未知語	候補1	関連度	候補2	関連度	候補3	関連度
列車	汽車	453	電車	402	自動車	210
機関車	電車	405	汽車	388	自動車	308
航空機	飛行機	381	気球	233	自動車	136
飛行船	気球	373	飛行機	311	電車	107
貨車	馬車	371	車	324	汽車	284
渡船	馬車	247	船	183	人力車	137

図4 実験結果② (最高値: 900)

この結果より、貨車、渡船を除けば、関連度の高い上位2つの語 (候補1, 候補2) のいずれかを取れば、そこそこの人間の感覚に近い形で未知語処理が行われると考えられる。今後は、未知語処理の精度をさらに向上させる手法について検討していく予定である。

6 おわりに

未知語処理の導入により、代表語以外の語についても判断が可能となり、より人間的な量的判断能力をコンピュータに付加することができたと考えている。今後は、さらに高度で深い知識に基づく評価尺度の検討及び様々な質問形式への対応についても研究を進めていく。

参考文献

- [1] 辻野他：比較判断のための常識スケールの構築と量的判断メカニズム, 情報処理学会第56回全国大会 1M-2 1998
- [2] 入江他：知的メカニズムのための概念間の類似度定量化方式, 情報処理学会第58回全国大会 1M-6 1999