

量的な判断常識を備えた人工知能

2N-1

— 質問文理解 —

牧野 俊朗 太田 昌克 飯田 敏幸

NTT 情報通信網研究所

1 はじめに

従来のエキスパートシステムの枠を超えて、より柔軟な判断や処理を行なうことができるシステムには、人間が持つ常識的な知識とそれを扱う推論系が必要である。我々は常識のうちの量に関する部分に絞って、常識知識ベースの構築とその知識をうまく扱える推論系の確立を目指して、「量的な判断常識を備えた人工知能」の研究を行なっている[1]。「量的な判断常識を備えた人工知能」は、量に関するあいまいな表現を含む指示文や質問文などの入力文を理解し、その指示に従った処理や質問に対する回答のための処理を実行するために必要とする情報を提供する。我々はこの「量的判断常識を備えた人工知能」のアプリケーションの1つとして知的DB処理[2]を実現しているが、本稿ではその入力文となる数値DBに対する質問文の理解の手法について述べる。

2 質問文理解とは

質問文を理解するとは、相手が何を尋ねているかを把握し、その質問に対して回答を出すことができるということである。そこで、コンピュータが質問を理解するには、人間の入力した質問文から質問に回答するために行なう処理に必要な情報を抽出し、コンピュータが利用可能な形式に変換することが必要である。例えば、質問文「重い車は何ですか」は、話題が車に関することであり(対象概念)、車の持つ量的な特性のうちの車体重量(量属性)が重い(比較概念)ものを探している(質問タイプ)ということを意味している。我々は、他の様々な質問文についても考察し、以下の情報を抽出することが質問文の理解の上で必要なことだと考えた。

1. 質問タイプ: 質問の形態 (例: which 型、yes/no 型)
2. 対象概念: 判断対象を表す概念 (例: 車、山)
3. 量属性: 量の尺度に関する属性 (例: 重量、価格)
4. 比較概念: 量の値の程度、値の変化の傾向を表現する概念 (例: 重い、安い、急騰している)

5. 修飾概念: 比較概念を修飾する概念 (例: 非常に)
6. 時間概念: 時間を表す概念 (例: 最近)
7. 時間修飾概念: 時制概念を修飾する概念 (例: ごく)
8. 比較対象: 比較の対象となるもの (例: カローラ)
9. 比較値: 比較の基準となる値 (例: 1000kg)
10. 比較種類: 比較の種類 (例: より、以上、以下)
11. 接続語: 上記のいずれかの項目が複数併記されていた場合の接続条件 (例: かつ、または)

例えば、質問文「重くて非常に安い車は何ですか」から抽出された上記の各種情報は図1に示すようなテーブル形式に変換される。以後この cst テーブルと rel テーブルからなる形式を cst(concept set) 形式と呼ぶ。

量的な判断常識を備えた人工知能は、この各 cst テーブル毎に「重い」などのあいまいな語の定量化などの量的判断を行ない、その結果を rel テーブルの情報で結合してアプリケーション部との間のインタフェースに適合する形式に変換しアプリケーション部に渡す。アプリケーション部はそれをもとに適当なデータベース操作言語を作成し、データベースの検索を行なう。

rel	接続語	かつ
	接続語で結合されるcstのID	cst 1
	接続語で結合されるcstのID	cst 2

cst	ID	cst 1	ID	cst 2
	対象概念	車	対象概念	車
	量属性	車体重量	量属性	価格
	比較概念	重い	比較概念	安い
	修飾概念		修飾概念	非常に
	時間概念		時間概念	
	時間修飾概念		時間修飾概念	
	比較対象		比較対象	
	比較値		比較値	
	比較種類		比較種類	

図1 「重くて非常に安い車」の cst 形式表現

3 質問文理解の手順

3.1 全体の処理の流れ

質問文は形態素解析により単語分けされる。次に、入力語のうち語尾(～は、～か等)やキーワード(どちらが、どれが等)から、質問タイプを決定する。次に、人工知能が持つ知識ベースを利用して、各語が上記のどの概念

に属するかを判断しラベルづけを行なう。このラベルをもとに cst 形式への変換を行なう。続いて、入力文で省略された情報の補完を行なう。

3.2 入力語処理

入力語処理では、単語分けされた各語に対して概念知識ベース、登録語辞書、不要語辞書、キーワード辞書を用いてラベルづけを行なう。概念知識ベースに記述してある概念語には、知識ベースに従って、対象概念、量属性、比較概念、修飾概念、時制概念、時間修飾概念のラベルをつける。接続語や比較内容を示す語に関してはキーワード辞書に従い、その意味のラベルをつける。数値は、比較値としてのラベルをつける。ユーザが定義した登録語の辞書中の語は、相当する概念語、さらに対象に依存する場合はその対象概念の情報を付加し、概念知識ベースを調べてその所属する概念のラベルをつける。文中に出現する対象語が、登録語を限定する対象概念かその下位概念に属さない場合は、未知語としてラベルをつける。不要語辞書に記述されている助詞、助動詞などの不要語は不要語であることを示すラベルをつける。それ以外の語に関しては、未知語であることを示すラベルをつける。

3.3 cst 形式への変換処理

入力語処理で付与されたラベルをもとに、cst 形式のテーブルの各項を埋める。この際、接続語の部分で区切り、それぞれを1枚の cst とし、rel テーブルの rel_w に接続語の意味を、rel_cst1, rel_cst2 にその接続語の前後に対応する cst テーブルの id を記述する。未知語がある場合は、cst テーブルに一時的に未知語の項目をつくり、そこに記述する。不要語はこの段階で捨てられる。

3.4 補完処理

通常の文章ではしばしば語の省略が行なわれる。例えば「重くて長い車の中で非常に安いのは」と言う文は、「重量が重い車かつ長さが長い車の中で非常に価格の安い車は」と言うことを意味している。このアンダーラインの部分省略された語である。人間は文法的な性質や各語の意味について知っているので省略があっても意味がわかるが、現状のコンピュータにはそれだけの知識がないので省略されたままだと意味がわからない。そこで、省略されている語を補完が必要がある。補完処理には、前 cst テーブルからの対象概念の複製、cst テーブル間の語の複製処理、量属性の補完^[3]がある。前 cst テーブルからの対象概念の複製とは、質問が続けて行なわれて対

象概念が省略された場合に前の質問文の対象概念をその質問文の対象概念とすることである。例えば、「重い車は何ですか」に続いて「安いのは」という質問文が入力された場合、前者の質問文の対象概念である「車」を後者の質問文の対象概念とする。cst テーブル間の語の複製処理とは、接続語を含む質問文中の語の省略を補うものであり、rel によって接続された cst テーブル間で必要な情報を複製することによって実現される。この処理はあらかじめシステムに与えられたルールに従って行なわれる。量属性の補完とは、対象概念と比較概念より量属性を推定することを言う。この処理は概念知識ベース中や対象世界知識に記述されている対象概念と量属性の関係、比較概念と量属性の関係を用いて行なわれる。

3.5 未知語処理

未知語の処理は、上記の補完処理の段階で対象概念が決定済みの場合と未決定の場合にわけて行なう。対象概念が決定済みの場合は対象世界知識を調べる。未知語のうち対象世界特有の語は、システムが一般的な語に置き換えたり、固有名詞として判定し、比較対象と見なしたりする。対象特有の語でないものは未知語としてユーザに質問する。対象概念が未決定の場合は、まずユーザに質問し、対象概念が決まったあとは同様に処理を行なう。

4 おわりに

理解とは、「コンピュータが処理を行なうための情報を抽出し利用できる形式に変換することである」という立場に立って量に関する質問文の理解について述べた。まだ、話題及び、文の形式に制限はあるものの、文理解の研究の第一歩となると考えている。今後はまず、様々な質問文が適切な形式に変換されるかを調べ、処理手順やそこで用いられるルールの精緻化を行ない、さらに指示文などの他の文形式や他の話題の文の理解の研究を行なっていきたい。

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 - 知的 DB 検索への適用 -, 情報処理学会第 45 回全国大会, 5H-08, 1992.
- [3] 島田他: 量的な判断常識を備えた人工知能 - 推論方式 -, 情報処理学会第 43 回全国大会, 5E-07, 1991.