E-034

連想処理を用いた複数文からの質問応答方式
A Method of Question Answering from Plural Sentences with Association Processing

長束 謙吾† 吉村 枝里子† 渡部 広一† 河岡 司†
Kengo Nagatsuka Eriko Yoshimura Hirokazu Watabe Tsukasa Kawaoka

1. はじめに

近年，情報社会においてコンピュータは大きな役割を担っており，日々進歩を続けている．そのためには，コンピュータに人間との常識的な質問応答をさせる必要がある．その手段として現在，様々なタイプの質問応答システムが提案されているが，それらの殆どは記号検索処理を基本としている．

それらに対して本稿では，意味処理に基づく質問応答システムの開発を目的としている．例として「理科」に関する文書情報知識を対象とし，これにプロダクションシステムの考え方を導入し，連想処理を用いて関連する複数の知識文から回答を得る質問応答システムを提案する．提案手法では，連想処理を実現するために，概念ベースや関連度計算から成る連想システムを用いる．

2. 連想システム[1]

2.1 概念ベース

概念ベースは，電子化された複数の辞書から抽出した概念表記や属性によって機械的に構築され，約 12 万語の概念を蓄えた大規模知識ベースである．

概念はある語 A をその語と関連の強いと考えられる語（属性）a_i と重み $w_i(>0)$の対の集合として以下に定義する．

$$A=\{(a_1,w_1),(a_2,w_2),\cdots,(a_n,w_n)\}$$

ここで，属性 a_i を概念 A の一次属性と呼ぶ．また，属性 a_i も概念ベースに登録されている 1 つの概念である．従って，a_i からも同様に属性を導くことができる．a_i の属性 a_{ij} を概念 A の二次属性と呼ぶ．

2.2 関連度計算方式

関連度計算方式は，概念ベースに定義された語と語の関連の強さを，同義性，類似性のみに関わらず定量化する手法である．

関連度は，0 以上 1 以下の連続的な実数で表され，概念同士の関連が大きいほど関連度は高くなる．この関連度を求める計算は，それぞれの概念を二次属性まで展開し，その重みを利用した計算によって最適な一次属性の組み合わせを求め，それらが一致する属性の重みを評価することで算出する．関連度計算の例を表 1 に示す．

表 1 関連度計算の例

基準概念	対象概念	関連度
飛行機	航空機	0.418
	自動車	0.057
	花	0.003

† 同志社大学大学院工学研究科
Graduate School of Engineering , Doshisha University

3. 理科に関する文書情報知識

3.1 理科知識ベース

理科知識ベースは質問文に対して，回答となる文章を返すために作成した知識ベースである．理科知識ベースには，web や参考書から知識となる文章を抽出し，表 2 のように知識文として 1622 文格納した．この格納された知識文を質問文に対する回答文として用いる．また，知識文を形態素解析した自立語群も格納されており，この自立語群は関連度計算や知識文検索に使われる．

表 2 理科知識ベースの例

知識文	自立語群
花粉が付く雌蕊の先を柱頭という	花粉，付く，雌蕊，先，柱頭，いう
花粉を作る雄蕊の先にある袋を葯という	花粉，作る，雄蕊，先，ある，袋，葯，いう

3.2 理科概念ベース

既存の概念ベースは「家」や「学校」といった一般的な概念を登録しており，「ベガ」や「アルタイル」といった特別な理科に関する概念が無いため，正確な関連度計算を行うことができない．そこで別途，理科概念ベースを構築する必要がある．理科概念ベース構築方法は，理科知識ベースの知識文を用いて既存の概念ベースに理科の概念を追加する．理科知識ベースにある知識文から理科概念ベースに概念を追加する一例を簡潔に図 1 に示す．この際に，新概念の動的追加を可能とする概念ベースの自動拡張手法[2]を用いた．

知識文

マツやススキのように風で運ばれる花粉がある

⇩

理科概念ベース

概念	属性
マツ	ススキ/ 1.32，運ぶ/ 0.98，花粉/ 0.89 …

図 1 理科概念ベースへの追加例

これにより，今まで取れなかった「マグニチュード」と「地震」といった既存の概念ベースでは未定義だった概念同士の関連度も算出できる．

4. プロダクションシステム

一般的なプロダクションシステムとは，知識ベースを基にして必要な知識を判断し，まとめることにより問題解決や推論を行うシステムである． 推論の方法としては，既知事実から新しい結論や事実を得るプロセスである前向き推論を用いることとする．前向き推論でのプロダクションシステムの流れを図 2 に示す．プロダクションシステムを用いれば，問題解決に必要な知識をまとめて，汎用的な問題解決に対応させることができると考えられている．

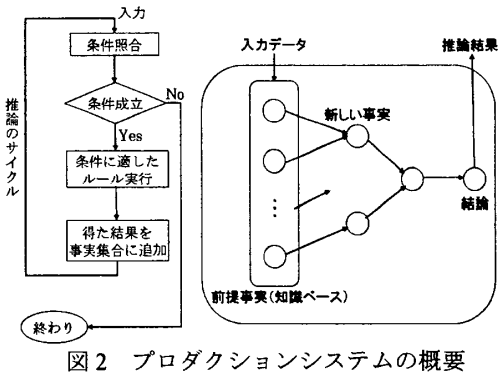

図 2　プロダクションシステムの概要

プロダクションシステムを含め，従来の質問応答システムは，データベースやプロダクションルール（if 条件 then 事実）のような整理された知識表現の知識ベースを検索することにより回答を得ているが，これまでのシステムはいずれも記号検索を基本としているため多様な表現の質問に対応することができない．そこで提案手法では，（if 条件 then 事実）のように整理された知識表現ではなく，教科書等で表現される説明文のような自然文を用いており，また概念ベースや関連度計算方式から成る連想システムを用いた提案手法によって，未知の語を既知の語に置き換える．

5．提案手法

5.1　提案手法の構築

前節での一般的なプロダクションシステムの考え方を導入し，連想処理を用いた手法の構築を行った．基本的な処理の流れを以下に示す．本稿での研究対象としては，典型的な三段論法(A⇒B⇒C)で解くことのできるような問題に限定している．

例として質問文「鉄に塩酸を加えると何が発生するか」を提案手法で考える．このとき，理科知識ベースに「鉄は金属である」，「金属に酸性の水溶液を加えると水素が発生する」という知識が存在するとする．

[提案手法の流れ]

1) 質問文を基に理科知識ベース全体から候補 50 件抽出
   具体的には質問文と関連が強い知識文 50 件を関連度計算より抽出する．
2) 候補 50 件から(A ならば B)を決定
   ここでは，「鉄は金属である」という事実が(A ならば B)として決定される．この際，決定基準としては質問文と最も関連の強かった知識文を用いる．
3) 決定した(A ならば B)から事実 B を決定
   決定した知識文「鉄は金属である」の自立語のうち，質問文の自立語でないもの「金属」を事実 B として推定し，質問文の自立語である「鉄」と置き換える．
4) 新しく置き換えた質問文の自立語を基に(B ならば C)を候補 50 件から導出
   提案手法では新しく置き換えた質問文の自立語群を用いて，候補 50 件のそれぞれの知識文の自立語群との連想処理を行い，関連度順に並べた．そして，関連の強かった「金属に酸性の水溶液を加えると水素が発生する」を回答文として出力する．

5.2　提案手法の検証と考察

前節の流れに従って提案手法の検証を行った．尚，評価した質問文として，典型的な三段論法で解くことができるもの 46 文を用意し，評価基準として最終的に出された上位 5 件の知識文の中に回答となる語が含まれていれば正解とした．評価結果を表 3 に示す．

表 3　提案手法での評価結果

	上位 1 件	上位 5 件	上位 10 件	上位 15 件
精度	13.0%	43.5%	52.2%	56.5%

表 3 より，関連度計算を用いた手法の精度が上位 15 件でも 56.5%である．改善点としては 2 点考えられる．

理由 1：上位 1 件から事実 B を推定している

考察した結果，事実 B を推定する時点で上位 1 件に事実 B が含まれているのは 30.4%，上位 5 件中では 82.6%の精度が得られていることが分かった．よって上位 5 件から事実 B を推定するアルゴリズムへの改良が望まれる．

理由 2：自立語群に意味の持たないものも含んでいる

本稿での自立語群は知識文または質問文を形態素解析したものであり，重要でない語も含まれていることから，関連度に信頼性が生まれない．そこで，重要な語のみを取得できる知識フレーム[3]の使用または形態素解析の調整を行う必要がある．これにより，重要な自立語のみを用いた関連度計算が可能となる．

また，前節で，例として取り上げた質問文「金属に塩酸を加えると何が発生するか」に対して，提案手法では回答文として「金属に酸性の水溶液を加えると水素が発生する」が出力された．従来の記号検索では概念の持つ意味を考慮せず，「塩酸」という表記でのみ検索するため，「金属に塩酸を加えると水素が発生する」のような知識文が知識ベース内に無いと回答できないが，提案手法では，「酸性」と「塩酸」の意味的な近さを連想処理によって考慮しているので，人間の持つ柔軟な連想が可能となっている．よって，失敗もあったが，概念の持つ意味を拡張できているといえる．

6．おわりに

本稿では，（if 条件 then 事実）のような決められた知識表現ではなく，自然文を用いて複数知識文を用いて回答できる手法を提案した．提案手法では一般的なプロダクションシステムの考え方を基本とし，概念ベースと関連度計算を用いた連想処理によって概念の意味を拡張させている．これより，多様な表現に対応できる質問応答システムを提案できたと考えられる．

参考文献

[1] 渡部広一，河岡司：“常識的判断のための概念間の関連度評価モデル”，自然言語処理，Vol.8，No.2，pp.39-54，2001

[2] 後藤敏貴，渡部広一，河岡司：“新概念の動的追加を可能とする概念ベースの自動拡張手法”，信学技報，NLC2006-74，pp.7-12，2006

[3] 中本一志，渡部広一，河岡司：“web 情報文からの教養知識の自動学習方式”，信学技報，NLC2007-93，pp.33-37，2007