

自然言語処理技術を用いた中学理科教授学習システム

小西 優輔[†] 黒岩 眞吾^{††} 任 福継^{††}

本稿では、自然言語処理技術を用いた中学理科教授学習システムの構築手法を提案し、現在構築中の実験システムの概要を述べる。本システムは、Super Function (SF) を用いて学習者入力文解析、理科問題文自動生成を行なう。従来の教授学習システムでは、予め用意された提示パターンしか出さない場合が多く、また解答方式は選択式のため学習者主体の学習環境とはいえない。本システムでは解答方式は学習者の自由文入力を可能とし、誤答に対してはその間違いによる現実世界での現象を提示する。またシステムのインタフェースには音声認識、音声合成を用いる。これにより学習者のタイピング技術に左右されず、キーボードになじめない学習者もシステムを容易に使用することが期待できる。

キーワード：教授学習システム，自然言語処理，教育，学習支援，理科

A Junior High School Science Education Support System Using Natural Language Processing Techniques

Yusuke Konishi[†], Shingo Kuroiwa^{††} and Ren Fuji^{††}

In this paper, a new method for constructing a junior high school science education support system using natural language processing technique is presented, and the outline of an experimental system based on the proposed method is described. In the system, the input sentence is analyzed and the science question sentences are generated automatically by using Super Function. The system can give more interesting questions and answers providing students pleasant learning environment. Moreover, the speech recognition and the voice synthesis are used for the system interface.

KeyWords: *Learning system, Natural language processing, Education, Educational support, Science*

1 はじめに

現在の教育に目を向けてみると、若者の理科系離れが社会的問題となっている。これは、日本の入学試験のシステムが原因で生まれたと考えられている。入学試験のシステムにより学習の形態が、知識詰め込み型、暗記中心型となり、自分で考えるという機会が少なくなった。理科系科目の特徴は、自分で物事を考え試行錯誤を繰り返しながら答えを導き出すことである。しかし、暗記中心となることで理科系科目に対する面白さ、物事に対する興味が薄れていき若者の理科系離れは進んでいる。理想の学習形態は、個人個人に専属の教育者が付いて教育を行なうことであるが、このことが不可能であることは明白である。

このような背景から、コンピュータを用いた学習ソフトや教育支援システムへの期待が高まっている。そこで、我々は自然言語処理技術を用いた中学理科教授学習システムを提案する。以下2章では提案システムの概要について述べ、3章では自然言語解析、自然言語生成、興味喚起モジュールについて述べる。

[†] 徳島大学大学院工学研究科 知能情報工学専攻, Department of Faculty Engineering, Tokushima University

^{††} 徳島大学工学部 知能情報工学科, Department of Faculty Engineering, Tokushima University

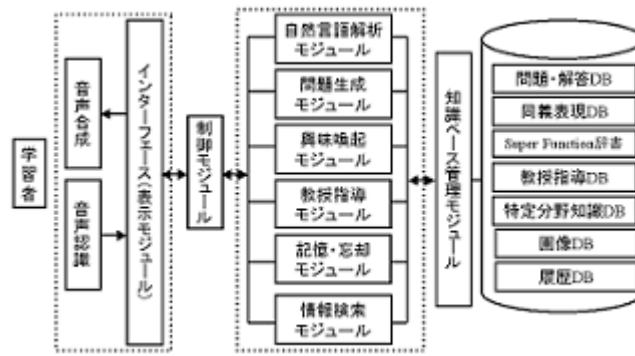


図1 システム構成図

2 提案システムの概要

本節では、提案する自然言語処理技術を用いた中学理科教授学習システムの概略について説明する。概略を図1に示す。本システムは、図1に示す各モジュール、知識ベース、辞書から構成される。本システムでは演習問題の解答方法、ヘルプ機能は選択式ではなく学習者の自由文入力とする。入力文は自然言語解析モジュールにより解析される。自然言語生成モジュールでは、理科問題文(演習問題)を問題・解答DBから自動生成する。学習者が入力した解答が誤答の場合、興味喚起モジュールにより学習者の誤答に対する現実世界での現象を提示する。教授指導モジュールでは教授指導、確認テスト、演習問題を行なう記憶・忘却モジュールでは、学習者の履歴を管理する。システムのインターフェースには音声認識、音声合成を用いる。

3 自然言語処理技術を用いた中学理科教授学習システム

3.1 自然言語解析モジュール

自然言語解析モジュールでは、入力された文章の解析を行なう。入力される文章は主に、理科演習問題の解答、理科問題文、HELPでの学習者からの質問である。入力文に対し、まず日本語形態素解析システム茶筌を用いて形態素解析を行なう。

3.1.1 語尾の取り扱い

SFを用いた問題文解析を実装するにあたって、まず日本語特有の、表現の多様性という問題点が挙げられる。SFを用いた手法では、各問題文に対応した辞書が必要なため、辞書の数が爆発的に増えてしまう。この問題を避けるために、特定の語尾を除去する。除去の対象となる語尾は、中学理科問題集から収集した1分野問題文1148文から解析に問題とならない語尾に決定した。

3.1.2 複合語の処理

形態素解析の結果、複合名詞等は、水酸化 / ナトリウムというように2つの名詞となる。これでは解析の際問題となる。よって本研究では以下の場合、複合語と決定し1つの要素として扱う。

1. 複合名詞 : サ変接続名詞と後に続く一般名詞は結合し複合名詞とする。

例: 水酸化 / ナトリウム ⇒ 水酸化ナトリウム

2. 形式名詞 : 形式名詞(こと、とき、etc)は自立語として扱わない。

例: 加えた / とき ⇒ 加えたとき

3. 補助動詞 : 補助動詞(いる、みる、etc)は自立語として扱わない。

例: 決まって / いる ⇒ 決まっている

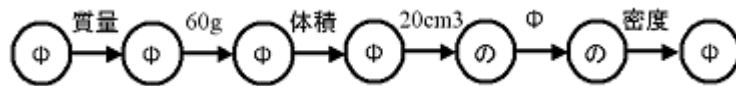


図2 方向グラフ

密度
名詞-一般
って
助詞-格助詞-連語
何
名詞-代名詞-一般
です
助動詞
か
助詞-副助詞/並立助詞/終助詞

図3 例1形態素解析結果

私	の
名詞-代名詞-一般	名詞-非自立-一般
は	か
助詞-係助詞	助詞-副助詞/並立助詞/終助詞
密度	全然
名詞-一般	副詞-助詞類接続
が	わから
助詞-格助詞-一般	助詞-自立
何	ない
名詞-代名詞-一般	助動詞
な	
助動詞	

図4 例2形態素解析結果

4. サ変接続名詞+サ変動詞

例：反射（サ変接続名詞） / する（サ変動詞） ⇒ 反射する

5. 数字+単位：数字と単位はまとめて一つの複合語とする。また「何グラム」のように「何+単位」も一つの複合語とする。

例：時速 / 100 / km ⇒ 時速 100km

3. 1. 3 Super Function(SF)による理科問題文解析

本研究では、原言語を学習者が入力した文章とする。目標言語は、原言語に情報が付加されたものである。原言語に付加された情報は、問題文の解答または公式、問題文の種類、ノード数、変数である。原言語のSFと目標言語のSFが一致すると、辞書より解答を出力する。

SFによる理科問題文解析のプロセスを以下に示す。

- (1) 形態素解析
- (2) 語尾、複合語の処理
- (3) SF, 名詞, 未知語に分類
- (4) ノードテーブルが一致するSFを検索
- (5) 辞書から必要な情報を取り出し、解答

まず学習者の入力文を形態素解析し、3.1.1, 3.1.2で述べた語尾と複合語の処理を行なう。その結果を基に各形態素を定数と変数に分類する。ここでは複合語は一つの形態素と仮定する。例として「質量60g体積20cm³の物質の密度をもとめよ」という問題文を(3)まで処理した結果の方向グラフを図2に示す。例文の「物質」は問題を解く際に必要ないため変数Φとする。(4)では作成したSFと一致するSFを辞書より検索し、一致すれば(5)の処理に移る。(5)では辞書より必要な情報を取り出す。この例では辞書より密度を求める公式と必要な変数の情報が取り出される。その情報を基に密度を計算し解答する。単位は標準単位とする。

3. 1. 4 HELP機能での学習者質問文解析

本研究では、学習者は演習問題を行なっている途中または教授学習中に、HELP機能を用いてわからない事をシステムに質問することができる。例えば「密度」が何なのかわからない場合、「密度ってなに?」というように自由文によりシステムに質問を行なう。以下、学習者からの質問文解析手法について述べる。

まず入力された文章に対し形態素解析を行なう。その結果を基に3.1.2で述べた複合語の処理を行なう。その際、複合名詞は一般名詞と品詞付けする。次に、後方検索により一般名詞を検索する。

検索の際に非自立の名詞と代名詞は一致させない。一致した名詞が質問の核となる名詞とする。単語辞書より一致した名詞を検索しその名詞に対応する説明文を表示する。

例として、例1：「密度ってなにですか」、例2：「私は密度が何なのか全然わからない」という2つの質問文を用いる。形態素解析の結果を図3、図4に示す。上記の方法で検索を行なうと2文とも「密度」が質問の核となる名詞になる。よってシステムでは密度の説明を表示する。

3. 2 自然言語生成モジュール

自然言語生成モジュールでは理科問題文の生成を行なう。問題文の生成には、スロット法を用いる。スロット法とは、文の骨組みを作っておいて、その中のいくつかの単語（スロット）を適切に決めることによって文を作るという方法である。文の骨組みには、自然言語解析で使用したSF辞書中のノードテーブルを使用し、スロットはエッジテーブルを使用する。例として「質量60g体積20cm³の物質の密度をもとめよ」という問題文を用いる。文の骨組みにはノードテーブルを用いるので、

$$\langle s_1 \rangle \times \langle s_2 \rangle \times \langle s_3 \rangle \times \langle s_4 \rangle \text{ の } \langle s_5 \rangle \text{ の } \langle s_6 \rangle \text{ を求めよ}$$

$\langle s_i \rangle$: スロット

となる。次にエッジテーブルよりスロットを決定する。最後に数値をランダムに決定し、語尾を付ける。ここで数値を決定する際、ランダムに数値を決定するが、単位及び温度、物質によって数値の範囲を予め決めておく。これは、食塩を水に溶かす場合などに溶解度を考慮しなければ正しい理科問題として成立しないためである。

3. 3 興味喚起モジュール

学習者に飽きることなく繰り返し学習させるためには、学習者の興味を引き出す事が重要である。そのために、誤答の場合でも間違いと表示し、解説を提示するだけでなく、その間違いに対しどのような事が起きるかということを示す。例えば、「重力加速度はいくらですか」という問題があるとする。この問題の正答は「9.8」である。これに対し学習者が、「-9.8」といった解答を行なえば「それではあなたが空に飛ばされるよ」と表示し、実際に動物が空に飛ばされるアニメーションを提示する。これには、学習者の誤答の程度、種類を判定する、前向き推論エンジンを用いる。前向き推論エンジンは判定結果としてのための、キーワードを返す。システムはそのキーワードに応じ、学習者の誤答に合ったアニメーション及びテキストを提示する。

4 おわりに

現在同システムは開発途上であり、現段階では教授学習システムにおけるSFの有効性を検証できるまでには至らなかった。中学理科を対象領域とする場合の問題点として、SFはコーパスからの半自動抽出は可能であるが、中学理科のコーパスが現在存在しないことが挙げられる。よって、今後中学理科のコーパスを作成する必要がある。

謝辞

本研究は科学研究費補助金基盤研究B (No. 14380166) によって実施された。

参考文献

- Fuji Ren, "Super-function based machine translation," Communications of COLIPS, pp.83-100, 1999
- 小西優輔 正田忠 黒岩眞吾 任福継, "自然言語処理技術を用いた中学理科教授学習システム", 電子情報通信学会信学技報, NLC2003-10, (2003)
- Yusuke Konishi, Shingo Kuroiwa, Fuji Ren, "A Junior High School Science Education Support System Using Natural Language Processing Techniques", Third International Conference on Information, Nov. 29 - Dec. 2, 2004.