

基礎的な力学問題解決における正規化処理法

青柳 祥平† 奥村 紀之†

†長野工業高等専門学校 電子情報工学科

1 はじめに

本研究の目的は、自然言語文章で記述された基礎的な力学問題を、コンピュータにより解決することである。本稿では特に、問題文の文法的な解析後の解答に必要な語句を特定するための正規化処理法を提案する。

2 システム概要

この節では、基礎的な力学問題解決システムにおける正規化処理について述べる。基礎的な力学問題とは、高等学校程度の力学の知識等を踏まえ、式(公式)の組み合わせを適切な順に解決していくことによって、解を数値として一意に定めることが可能な問題文とする。例として、以下のような問題が該当する。

Ex.) 長さ 1.5[m] の単振り子の周期は何 [s] か。

初めにシステム全体の流れを紹介し、2.1, 2.2 節では構文解析と正規化処理について説明する。更に各節では、構文解析などの項目を構成する具体的な処理について扱うものとする。

本システムの基本的な処理の流れを図 1 に示す。

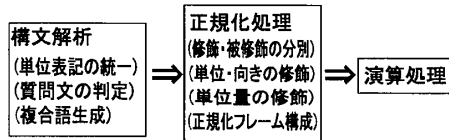


図 1: システムの概略

2.1 構文解析

本研究における構文解析には形態素解析器 MeCab[1]を用いるものとする。以下に、構文解析のステップとして、2.1.1, 2.1.2, 2.1.3 を示す。

2.1.1 単位表記の統一

本システムでは、コンピュータを用いて自然言語文章中から単位をデータとして抽出する処理を行うために単位変換データベース(基本的な日本語記述単位と英数記述単位の 25 種類の組み合わせ)を作成し、日本語単位を英数単位に変換する処理を行う。これは、日本語と英数で同一の意味を持つ単位が混在することを

避けるためである。ただし、'°' や 'Ω' などの半角表記できないものはそのままの扱いとする。

2.1.2 質問文の判定

問題を正確に解決するに当たり、正規化処理を施し必要なデータを取得する上で、問題文中から解答とする部分を把握することが重要である。例として、以下の問題の場合、解答として導くものは「重さ」であるため、問題文中の「重さ」を解答部分として認識しなければならない。

Ex.) 100[kg] の人の重さはいくらか。

そこで、問題集より抽出した 259 題の問題文中で解答を訊ねている文章の文法傾向を調査した。全ての文に共通して、解答を訊ねている文章は以下の傾向が確認できたため、本システムではテンプレートとのマッチングにより質問文を判定するものとしている。

質問文の形式: 「～か。」 「～よ。」 「～なさい。」

2.1.3 複合語の生成

形態素解析を実行するに当たり、MeCab の辞書に登録されていない語句(未知語)は誤って分解されることがある。また、力学には「相対速度」などの複合語が多く存在する。そこで、一般名詞もしくは力学用語(C1) および数値+単位(C2) の 2 種類の複合語に連結条件(MeCab による解析結果の繋ぎ方)を定め、前述の 259 題に対し評価を行った。

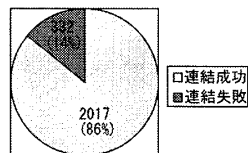


図 2: C1 の連結結果

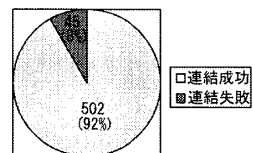


図 3: C2 の連結結果

評価結果を表したものが図 2, 3 である。どちらも評価値は 86%, 92% と高く、連結条件は定義通りであれば問題ないと考えられる。

2.2 正規化処理

既存の数値問題解決システム [2] では、限定された被修飾語とその語が持つ単位関係をデータベースに登録しておき、問題文中に両方が出現した場合に一組のデータとして正規化するという手法などが提案されている。

一方、本研究では、被修飾語のもつ単位の種類や被修飾語自体の量が限定されておらず力学分野全体に及

A Normalization Processing Method in Fundamental Dynamics Problem Solving

† Shohei AOYAGI

† Noriyuki OKUMURA (noriyuki.okumura@ei.nagano-nct.ac.jp)

Nagano National College of Technology (†)

ぶため、既存の手法は利用できない。そこで、以下に示す正規化手法を提案する。

2.2.1 提案する正規化手法

2.1.3 の複合語処理によって、「一般名詞もしくは力学用語」、「数値+単位」に分類されたデータを利用する。一般名詞もしくは力学用語は被修飾語に該当するものである (厳密には、力学用語は単位の修飾を受けた後、一般名詞を修飾する)。また、数値を含む単位は力学用語を修飾する。このため、前項の処理が終了した時点で、複合語は修飾語と被修飾語に分類される (図 4)。

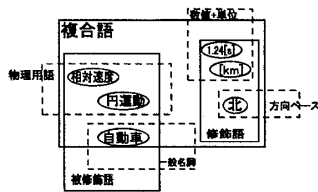


図 4: 修飾語・被修飾語

2.2.2 方角ベース

方角ベースとは基礎力学問題に出現し得る方角を表す語句 14 種類を登録したもの (表 1) であり、複合語生成の後、このベースと一部一致した語句は向きを表すデータとする。

表 1: 方角ベース (一部抜粋)

鉛直上	鉛直下	北	東	右向き	逆向き
-----	-----	---	---	-----	-----

2.2.3 単位・用語ベース

単位・用語ベースとは力学問題に必要な単位 (SI 単位系) とその単位をもつ用語 (単位量) の組み合わせ 71 種類から構成されるものである (表 2)。

表 2: 単位・用語ベース (一部抜粋)

SI 基本単位	SI 単位	単位量 1	単位量 2
s	-	時間	-
m/s	-	速さ	速度
$m^2 \cdot kg \cdot s^{-2}$	$N \cdot m$	エネルギー	仕事

2.2.4 代表語ベース

代表語ベースとは基礎力学問題に頻出する運動対象 55 種類を登録したものである (表 3)。

2.2.5 修飾語・被修飾語

修飾語は、自分に最も近い距離にある単位量を探し、その単位量により近い位置に他の修飾語が存在しなければ修飾する。修飾できないときは次第に距離を伸ばしていき、被修飾語が見つからない場合には、単位・用語ベースを逆参照し単位量を決定する。修飾された単位量は代表語ベースに登録されている語句 (被修飾語) を修飾する。これにより、品詞の係り受け関係を

表 3: 代表語ベース (一部抜粋)

物体	球	人	ボール	ばね	石
----	---	---	-----	----	---

把握し、次項に示すような正規化フレーム (図 5) を取得する。

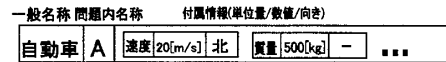


図 5: 正規化フレーム

2.2.6 評価結果および考察

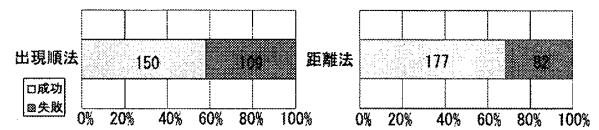


図 6: 本システムの正規化結果

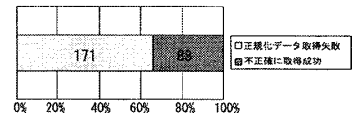


図 7: 既存システムの正規化結果

2 種類の評価方法を用いて本システムを評価 (図 6) し、更に [2] のシステムに対して同様の問題を入力した場合の正規化結果 (図 7) を調査した。図 6 に示す出現順法とは、単位量を、その単位量の左側かつ一番近い距離にある語句 (代表語ベースに登録されているもの) を修飾する方法であり、同様に距離法とは最も近い距離にある語句を修飾する方法である。図 6 より、距離法の方が出現順法よりも高い精度で正規化に成功することが確認できる。また、図 7 より既存システムに問題文を与えた場合、正規化データがまったく得られないことが多く、得られた場合にも全てにおいて不正確であることが分かる。

この結果より提案したシステムを利用することで既存システムでは対応できなかった力学問題問題を正規化できることが分かった。

3 おわりに

本研究では、既存のシステムでは対応できないような問題を解決するための正規化手法を提案した。今後は、正規化後の演算処理について具体的に検証を行う。

参考文献

[1] MeCab(<http://mecab.sourceforge.net/>)
 [2] 稲葉栄美子, 渡部広一, 河岡 司 "常識知識を用いた算数問題の意味理解", 情報処理学会研究報告. 自然言語処理研究会報告 IPSJ SIG Notes 2004(1) pp.147-154