

高等学校物理問題の自動解答探索システムにおける問題文解析手法の検討

A study of text analysis of high school physics workbook including technical terms.

佐藤 柚希† 志子田 有光‡
Yuzuki Sato Arimitsu Shikoda

1. はじめに

本研究では、高等学校の物理や電気回路学などの問題文に対し、それが求めている解答を解答・解説文の中から探索するシステムを構築することを考える。高等学校物理の問題文から解答を検索する際、はじめに問題に対する解答・解説の中から対応するものを選択する。テキスト分類は情報検索、パターン認識、機械学習などのさまざまな研究成果を用いて実現されている[1]。このような場合、以下の二つの課題が存在する。(1) **問題文の理解**: 高校物理の問題文は、専門的な用語が多く含まれている。さらに力学の分野では、物体の位置やその重さや速度などの用語に含まれる情報が存在し表層的な言語の一致だけでは、処理できない。(2) **問題文の不完全性**: 問題文の中の表層的な情報のほかに考えられる情報として単位があり、その組み合わせでほかの単位を持つ用語が解答に含まれる。専門用語の単位の情報を用いて解答を推定や物理の問題は専門的な用語が多く含まれており、このような解析を行うために、問題文を理解する非文法的な言語処理を行う必要がある[2]。

本稿では、「(1)問題文の理解」の前段階として、物理や電気回路学の問題文から専門用語とそれを指し示す記号、物理量を示す単位および物理量などの関係を解析する一手法について検討する。

2. 物理用語を含む文に対する単位タグの付与

高等学校物理の問題文の解析を行う際、文中に存在する単語の中には、単位を持つものが存在する。解答を探索する際登場する単位から分野を限定し、出現する単位の組み合わせによって構成される解答を探索できるため用語に単位のタグをつけることが必要である。

2.1 人手による単位の付与

高等学校の物理に登場する単位のうち最も基本となるものを表 1 に示す。

表 1 SI 基本単位

単位記号	名称	読み方
m	長さ	メートル
kg	質量	キログラム
s	時間	秒
K	熱力学温度	ケルビン
mol	物質質量	モル
cd	高度	カンデラ

問題文に登場する用語に対して名称と単位が直接存在する場合に対応するタグをつける。さらに教科書には、

この SI 単位の組み合わせによって存在する。これの単位も追加した 17 種類の単位タグを付与した。

2.2 推定による単位の付与

出現する用語の中には、その単語が示す単位が同じである。例えば、質量と重さ、速度と速さなどが考えられる。そこで、単位タグを推定し、付与をする手法を検討した。

2.2.1 推定によるタグ付け手法

単位の推定手順を図 1 に示す。問題文に形態素解析を行い名詞のみを抽出し、問題文を整理する。問題文に登場する単語に対して SI 単位とその組み合わせによって登場する単位 17 種類の名称と一致する場合は単位を決定する。一致しない場合、単位の名称と出現した用語の分散表現を取得し、類似度を測りしきい値以上で最大のものを単位とする。

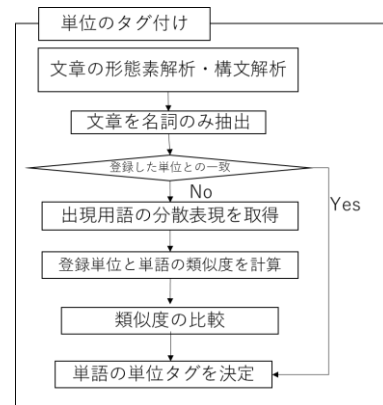


図 1 タグ付けの手順



図 2 単位タグの付与の例

2.3 分散表現

形態素解析ツール MeCab を使い、問題文の名詞のみを取得し、高等学校物理の教科書を学習コーパスとしたモデルによる分散表現を取得する。

2.3.1 Word2vec を用いた分散表現

ニューラルネットワーク言語モデルによる代表的な言語モデルである word2vec[3]による分散表現を高等学校物理の教科書を学習コーパスにして分散表現モデルで作成し、物理用語の分散表現を取得した。そのうち取得した分散表現の一例を表 2 に示す。

表 2 物理用語と最も似ている単語の例

時間		速度		摩擦	
類似単語	類似度	類似単語	類似度	類似単語	類似度
VA	0.91	時刻	0.91	摩擦力	0.98
2V	0.89	ゴール	0.89	係数	0.97
相対速度	0.86	走者	0.89	綱引き	0.95
走行	0.86	変位	0.87	垂直抗力	0.92
直線	0.85	流水	0.87	さまたげ	0.9

Word2vec を用いて得た分散表現を使い、cos 類似度を計算した。問題に出現する単語を順番に w_1, w_2, \dots, w_n 番目まであるとき、 w_n 番目の単語の分散表現を b_n と表し、17 種類の単位名称の分散表現の集合 $S\{s_1, s_2, \dots, s_m\}$ とする。この時、問題に出現するすべての単語に対して cos 値を計算し最大を求める式を以下に示す。

$$\text{Sim}(b_n, S_m) = \max_{1 \leq j \leq m} (\cos(b_n, s_j)) \quad (1)$$

3. 実験結果

市販の物理問題集内の問題に対して形態素解析を行い名詞のみを抽出し、単位タグを付与した。

その結果の一部を表 3 と図 3.4 に示す

表 3 問題文に対する単位タグ付与の割合の一例

問題名	出現単語数	単位名称一致	単位名称の一致割合	推定可能単語	単位の付与割合
8 平均の速さ	24	4	0.17	17	0.71
170 ばねと物体の運動	126	9	0.07	113	0.9
41 斜めの衝突	27	2	0.07	25	0.93
124 棒の重心の決定	38	3	0.08	31	0.82
78 糸で連結された物体の運動	27	3	0.11	25	0.93
414 気柱の固有振動	86	1	0.01	47	0.55

問題:x-tグラフ

図は、x軸上を運動している物体の位置x [m] と時刻t [s] との関係を示したx-tグラフである。物体の速さはいくらか。

図:None /x:面積 /軸:変位 /上:None /運動:None /物体:力 /位置:None /x:面積 /m:長さ /時刻:変位 /t:時間 /s:秒 /関係:None /x:面積 /t:時間 /グラフ:None /物体:力 /速さ:速さ /いくら:None /

図 3 問題文の単位タグ付与の例 1

問題:鉛直投げ上げ

鉛直投げ上げ地面から、速さ19.6m/sで鉛直上向きに小球を投げ上げた。重力加速度の大きさを9.8m/s²とする。(1) 地上14.7mの点を小球が通り過ぎるのは何s後か。(2) 小球が最高点に達するまでの時間は何sか。(3) 最高点の高さは何mか。(4) 小球が再び地面に落ちてくるまでの時間と、そのときの速度をそれぞれ求めよ。

地面:None /速さ:速さ /19.6m:長さ /s:速さ /鉛直:None /上向き:加速度 /小球:None /重力加速度:質量 /9.8m/s²:加速度 /1:None /地上:None /14.7m:長さ /点:None /小球:None /何:速さ /s:秒 /2:None /小球:None /最高:None /点:None /時間:時間 /何:速さ /s:速さ /3:None /最高:None /点:None /高さ:None /何:速さ /m:長さ /4:None /小球:None /地面:None /時間:時間 /とき:仕事 /速度:速度

図 4 問題文の単位タグ付与の例 2

3.1 結果と考察

図 3 から単位が見つからない図やグラフといった単語に対して単位をつけず、単位であるmやsに関しても単位をつけられている。しかし変数に関しては、t は時間と付与しているが x は面積である。この場合の x は物体の位置であるため長さ(m:メートル)の単位が正しい。

結果として単語と単位の認識があらかじめ用意した単位と名称の一致をとると問題文に出現する割合は低いことから word2vec による分散表現を用いた単位の推定を行うことで単位を推定することが可能である。しかし現在は、単位の無い単語について単位を付与してしまう。

大きな問題として専門用語の形態素解析で固有名詞の分割ができない場合や問題に出現する変数を単位ととらえてしまうことも今後の課題である。

またその他の分散表現の手法や類似度を求める手法に関しても検討したい。

謝辞

本研究は、JSPS 科研費 JP19K02979 の助成を受けたものです

参考文献

- [1] 岸田和明:文書クラスタリングの技法:文献レビュー, Library and Information Science, No.49, pp.33~75, (2003),.
- [2] 吉田達平, 松崎拓也, 佐藤理史, “大学入試化学の計算問題の自動解答システム”, 2015 年度人工知能学会全国大会, 2015
- [3] Tomas Mikolov, Kai Chen, Greg Corrado, Jeffrey Dean, “Efficient Estimation of Word Representations in Vector Space”, ICLR, 2013,