

三段階表示による記述式解答群の主要な内容の把握支援 - 類似表現の集約による改善 -

伊藤 慎治*1・高瀬 治彦*1・川中 普晴*1・鶴岡 信治*2
Email: 416m202@m.mie-u.ac.jp

*1: 三重大学大学院工学研究科電気電子工学専攻

*2: 三重大学

◎Key Words 演習支援システム, 自然言語処理, 記述式小テスト

1. はじめに

講師は、自身の講義において、学習者ごとの理解状況に応じた授業改善を行うことを試みている。しかし、さまざまな状況により特に大学などの高等教育機関ではそれが困難な場合が増えている。文部科学省の調査⁽¹⁾によると、大学入試において、推薦入試やAO入試などが盛んに採用されている。その結果、同じ大学の学習者でも、さまざまな学習能力・特性を持った者が混在するようになった。この状況は、多人数で行う講義を、講師から学生へと一方向のみで行うことを困難にしている。また、小林らは、大学などの学習者が多人数で行われる講義において、学習者は講義に集中せず、反応しない傾向があるとしている⁽²⁾。

そのため、学習者の様子から理解状況を把握し、授業改善を実現するには、講師からの能動的な行動が必要である。西森の大学の数学基礎教育に関する調査⁽³⁾によると、授業改善の方法として、小テストを多くの講師が利用していることを述べている。井上らは、小テストは学習者に授業参加を促す一因になることを示唆している⁽⁴⁾。このように、小テストは授業改善の方法として有用である。その反面、小テストにもとづいた授業改善を素早く行うことは、講師に多大な負担を強いる。

小テストにおける講師の負担を軽減させるため、これまでに、さまざまな支援の方法が検討されてきた。例えば、Moodle⁽⁵⁾などのeラーニングシステムでは、小テストを出題し、学習者の解答を回収し、解答形式によっては自動で採点できる。また、小テストの結果を自動で分析・集計することで、講師が学生の理解状況を把握しやすいように支援できる。多肢選択式・穴埋め式の小テストでは、eラーニングシステムによる支援が容易である。その反面、学生は推測で解答できるため、学生の状況の把握という観点からは好ましくない。また有効な選択肢の作成など、出題時の講師の負担が大きくなる。そのため、講義中に、学生の状況を捉えたいと思った瞬間に、これらの小テストを実施することは難しい。自由記述式の小テストは、多肢選択式とは反対の特徴を持つ。記述式の解答では、自分自身で文章を作成するため、解答文中に理解状況に関する情報が多く含まれる。加えて、講師は問題を準備するだけでなく、講義中にすぐ実施できる。また村山は、記述式小テストを実施すると学生自身が理解を深めるように学習することを示している⁽⁶⁾。他方で、講師は理解状況を把握するために一人ひとりの解答文を読む必要があり、その負担は重い。

これらをふまえ、本研究では、1, 2文の短答記述の小テストを実施した際、すばやく学生の主要な理解状況を把握できるように支援する計算機システムを構築することを目的として研究してきた。大庭は、三段階表示システム⁽⁷⁾を提案した。これは、学生が計算機を用いて入力した解答群を収集・解析し、講師に提示することで、解答群の主要な内容を素早く把握できるように支援する。

本稿では、この三段階表示システムを改良する。これまでの三段階表示システムでは、類似した表現の解答が別々に表示されるために、講師の負担が依然として軽減しきれない問題点が見られた。そのため、自由記述が原因で生じる学生ごとの類似表現の揺らぎを集約することで、類似した表現の解答が分散して表示されないようにする。

2. 三段階表示システム

この章では、大庭により提案された講師が三段階表示システムの概略を示す。これは、キーワード表示、フレーズ表示、全文表示の三種のインターフェイスを段階的に使用することで、解答群の概略をとらえやすくする。

2.1 各インターフェイスの概要

第一のキーワード表示インターフェイスは、解答群から自動抽出した重要そうな語を表示する。これらは語の重要度⁽⁸⁾に従い色分けされ、また使用頻度が高いほど上方に表示される(図1)。講師が注目した語を選択すると、次のインターフェイスに遷移する。

第二のフレーズ表示インターフェイスは、文節ごとの三列の表形式となっている(図2)。中央の列はキーワードを含む文節を示し、左右の列の文節は中央の文節との修飾関係をもつ。文節の隣には文節の解答群中での使用回数⁽⁹⁾が示されている。また、使用回数に応じてセルの配色は濃くなり、上位に表示される。注目した文節を選択することで、第三の全文表示インターフェイスに遷移する。この画面では、先に指定された文節を含む解答全文を確認することができる(図3)。

2.2 システムの問題点

三段階表示システムにより、講師が解答群から注目すべき記述を探し出せることを評価実験により確認した⁽⁹⁾。

その過程で、キーワード表示インターフェイス、フレーズ表示インターフェイスにおいて、類似した意味を持つ

キーワード・フレーズが別々に表示される問題点が指摘された。例えば、図1のキーワード表示インターフェイスでは、類似した意味を持つ語が分かれて表示されている。例えば、「表現」「表わさ」「表さ」といった語は全て類似した語であり、「表わす」ことを意味するが、別の単語として表示されてしまっている。これにより、講師は選択した単語と類似した意味を持つ語に対して複数のインターフェイスを行き来する必要がある、閲覧の負担が軽減できない。図2のフレーズ表示インターフェイスにおいても、類似した意味をもつ文節が別々に表示されている。例えば、被修飾節(図中の一番右の列)の「命令する」「命令して、」の「する」「して、」という表現は、修飾節の「CPUに」という文節との関係に限定して考えると、共通の「する」という意味であるといえる。このような表現の揺れが分けて計数されるため、実質的な解答数をとらえ難い。

これらの問題を解消するため、類義語・表現のゆらぎを考慮して処理する方法を検討する。いずれの問題も、これまでの三段階表示システムでは、解答の文字列に基づいて処理していることに起因する。キーワード表示インターフェイスでは、語同士の意味の近さを考慮していない。また、フレーズ表示インターフェイスでも、字面が異なる文節は別の文節として計数している。



図1 キーワード表示インターフェイス



図2 フレーズ表示インターフェイス

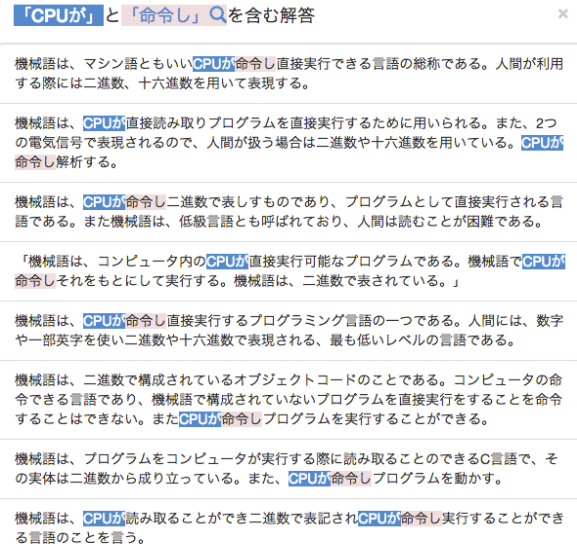


図3 全文表示インターフェイス

3. 類似表現集約のための提案手法

この章では、2章で示された問題を解消するため、類義語の集約方法について検討する。

3.1 キーワード表示における類義語の集約方法

類義語を判定するための簡単な方法は、類義語辞書を使用することであろう。しかし、類義語を集約する際に、講師ごとの考えの違いや、小テストの問題の内容の違いによっては、類義語だと判断する基準は変化する可能性を考慮しなければならない。そのため、似ているか否かのみを判断をする類義語辞書では、十分でないと考えた。

そこで、Word2vec⁽¹⁰⁾に着目する。これは、さまざまな単語をその意味にもとづいてベクトル化する手法である。その結果、ベクトルの類似度を判定することで、数値として類似度を得ることができると考えた。

Word2vecは、ニューラルネットワークの一種であり、文書群を学習することで、その文書群に含まれる各単語の並び順に基づき、各単語をベクトルとして表現する。その結果、得られたベクトルにより「パリ」-「フランス」+「日本」=「東京」というような意味に基づく演算が可能になる。本稿では、単純に着目した二つの単語に対応するそれぞれのベクトルのコサイン類似度(ベクトル同士の角度の差の余弦)により、二つの単語の類似度を求める(類似度は0.0以上1以下となる)。また、Word2vecにより、多量の一般的な文書群(コーパス)をあらかじめ学習しておき、その結果をもとに解答群から抽出された各単語をベクトル化して類似度を評価する。

3.2 類義判定のためのWord2vecの評価

この節では、Word2vecが本システムにおける類義語判定のために適切かどうかを検討するため、語同士の類似度を算出した際、その結果が人の感覚による類似度とどの程度差異があるのかをアンケートにより確認する。

このアンケートでは、ある文中で調査語を設定し、同様に、1組の基準語とそれを使った文(の一部)、5組の対照語とそれを使った文を設定する。これら計7組の単語と文を1セットで被験者に提示し、それぞれの語の文中で

の使われ方から、調査語と他の語の類似度を判断してもらおう。そのうえで、基準語と比べ類似している語(◎)、基準語と同程度に類似している語(○)、および、そうでない語(×)に、対照語を分類してもらった。

同じ語群に対して、Word2vec でも調査語と他の語の類似度を求め、アンケートの場合と同様に対照語を分類した。Word2vec に事前に学習させるコーパスとしては Wikipedia の日本語版の全文のテキストを使用した。処理にあたり、動詞は全て基本形に修正し、「機械語」などの複数形態素が組み合わさった名詞は日本語係り受け解析器 Cabocha⁽¹⁾により得られる文節内で形態素を連結し、1つの単語として処理を行った。

被験者による分類結果と、Word2vec による分類結果を比べ、Word2vec による類似度判定の妥当性を評価する。

評価実験は、大学生 5 人を被験者として行った。調査語としては、3 語の動詞「調べる」「進める」「用いる」と、3 語の名詞「必要」「構文」「認識」を用いた。これらの語は実際の講義中に出題された小テスト「デバッガーとはなにか」という問いに対する解答群に含まれる語である。実験に用いる 6 組の語と文は、すべてこの小テストの解答群のなかから選んだ。実験に用いた調査語、基準語、対照語およびそれらを使用した文の例を表 1 に示す。また、表 1 の例で、基準語や対照語を設定した例を表 2 に示す。

結果の比較は、一致率と相異率によって行う。人による判断と Word2vec での判断を比較して、双方の◎および×をつけた語が一致している割合を一致率、双方の◎および×の判断が正反対だった（どちらかが◎、もう一方が×）語の割合を相異率とする。一致率が高く、相異率が低ければ、Word2vec は類義語の集約に適するといえる。

表 1 動詞「進める」のアンケート内容

文章 A	
	開発を円滑に「進める」ためにバグを修正するツールである
1.	ソースコードを 1 つずつステップ実行し、該当する点で「促す」
2.	プログラミングを実行に「移す」
3.	ステップごとに分けてプログラムで表示される変数などを表示「させ」たりするステップを踏むことにより
文章群 B	4. 開発を円滑に「すすめる」ためのものであり
5.	プログラム中の不具合（バグ）の発見や修正を「行う」
6.	プログラムであるプログラムを一行ずつ実行させて間違いがあれば自動的に修正を「おこなう」

表 2 調査語、基準語、対照語の例

分類	語
調査語	「進める」
基準語	1. 「促す」
対照語	2. 「移す」 3. 「させ」 4. 「すすめる」
	5. 「行う」 6. 「おこなう」

3.3 結果と考察

表 3 には実験に使用した 6 つの語の平均一致率と平均相異率、表 4 に各語の結果をそれぞれ示す。

表 3 より、平均一致率はどちらも約 35% という結果となった。これは、Word2vec による結果が人の評価と一致していないことを示しており、提案法が不適切であることを示しているように見える。しかし、平均相異率に着目すると、動詞は 20% と低い数値であるが、名詞は 50% を超えている。これは、動詞については、人の判断と全く違う語は 20% 程度であり、提示の方法によっては使える結果であると考えられる。名詞については、相異率が高く一致率が低いいため、人の感覚と Word2vec で求めた結果は一致しているとはいえない。つまり、システムの改善手法として、Word2vec による類義語判定法は、動詞のみに適用するべきだといえる。

また表 4 から、個々の動詞と名詞の一致率・相異率に関しても考察する。動詞については、どの語も相異率は低い値である。しかし表 4 より、「調べる」という動詞の一致率が著しく低いことが読み取れる。この原因として、語によって Word2vec で算出した類似度の幅が違うことが挙げられる。「進める」と「用いる」では、抜き出した 6 語のうち類似度が最も高いものと低いものの差が 0.213、0.210 であったが、「調べる」は 0.119 であった。このように、類似度の幅が小さいことが原因で、被験者は語の類似関係をはっきりと認識できなかった可能性がある。

一方で、名詞に関しては、「構文」は一致率が、「認識」は相異率が低いなど、語によりさまざまな結果となっている。そのため、更に多くの名詞でも比較検討を行うべきであるといえる。

また加えて、word2vec に学習させるコーパスを改善することも検討すべきであろう。今回使用したコーパス (Wikipedia) 内における、各語の登場回数を表 5 に示す。表 5 から、明らかに動詞「調べる」と名詞「構文」の登場回数が少ないことが読み取れる。これらの語に関して、表 4 の一致率・相異率を見ると、他の語と比較して良い結果は得られていない。つまり、コーパス内での使用頻度が低かった語は、意味に基づくベクトル化が高精度でできなかった可能性がある。

表 3 平均一致率・平均相異率

	平均一致率[%]	平均相異率[%]
動詞	38.7	20.0
名詞	33.3	50.7

表 4 各語の一致率・相異率

語	一致率[%]	相異率[%]
動詞	20	28
	進める	44
	用いる	52
名詞	必要	40
	構文	8
	認識	52
		60
		68
		24

表5 各語のWikipedia内での登場回数

語	登場回数 [回]	
調べる	3299	
動詞	進める	13611
	用いる	47985
必要	71210	
名詞	構文	1134
	認識	11880

4. フレーズ表示インターフェイスにおける語の活用形に基づく改善方法

2.2節で示した言い回しの揺らぎについて、我々が解答群の分類方法として検討してきた方法⁽¹²⁾を用いて、語の活用形に基づき類似文節を集約することを検討する。この文献では、修飾関係にある2つの文節の関係を議論する際には、修飾されている側の文節にある助詞は無視してよいとしている。例えば、「CPUに」の被修飾節の「命令する」「命令して、」という2つの文節に関して、日本語形態素解析器 MeCab⁽¹³⁾を用いて品詞分解をそれぞれ実施すると、以下の表6, 7のようになる。接続助詞「て」は更に次の文節との関係を表すものであり、不必要であることがわかる。よって、助詞の違いに起因する表現のゆらぎの解消には、文献⁽¹²⁾の方法が適用できるであろう。

表6 「命令する」の品詞分解例

語	品詞, 活用形
命令	名詞
する	動詞 (基本, 自立)

表7 「命令した」の品詞分解例

語	品詞, 活用形
命令	名詞
し	動詞 (連用, 自立)
て	助詞 (接続)

5. おわりに

本稿では三段階表示システムについて、更なる改善を試みる手法を提案した。キーワード表示インターフェイスではWord2vecを用いる方法を検討し、有効性をアンケートにより評価した。また、フレーズ表示インターフェイスでは、語の活用形に基づく改善方法を提案した。これらの改善により、三段階表示システムにおいて、記述式であることが原因で生じる学生ごとの類似表現の揺らぎを集約できる可能性を示した。

謝辞

本研究は、JSPS 科学研究費 26350274 の助成を受けたものである。

参考文献

- (1) 文部科学省：“平成29年度国公立大学入学者選抜について”，
http://www.mext.go.jp/a_menu/koutou/senbatsu/_icsFiles/afieldfile/2016/10/06/1377882_3.pdf, 2017-6 参照
- (2) 小林 俊公, 柳沢 学, 池田 博一：“呼応度を用いた授業の分析”，工学教育, vol. 62-6, pp. 56-61 (2014)
- (3) 西森敏之：“大学生の授業における態度と数学教師の対策—日本数学会のある調査より—”，高等教育ジャーナル—高等教育と生涯学習—, vol. 6, pp. 1-31 (1999)
- (4) 井上 聡, 平林 茂：“授業改善策として学生に行う小テストの効果”，帝京科学大学紀要, Vol. 13, pp. 25-30 (2017)
- (5) Moodle.org：“open—source community—based tools for learning”，<http://moodle.org>, 2014 取得
- (6) 村山 航：“テスト形式が学習方略に与える影響”，教育心理学研究, vol. 51, pp. 1-12 (2003-3-30)
- (7) 大庭 知也, 高瀬 治彦, 川中 普晴, 鶴岡 信治：“多人数クラスにおける記述式小テストを支援するシステム：学生の理解状況をすばやく把握するためのインターフェイス”，Computer & Education, vol. 39, pp. 86-91 (2015)
- (8) 高瀬治彦, 川中普晴, 鶴岡信治：“記述式小テストの解答群の分析手法—解答群からのキーワード自動抽出—”，Computer & Education, vol. 34, pp. 46-49 (2013)
- (9) 伊藤 慎治, 高瀬 治彦, 川中 普晴, 鶴岡 信治：“記述式解答群の主要な内容の把握支援—三段階表示の効果の検証—”, 2016 PC Conference, pp. 43-44 (2016)
- (10) Tomas Mikolov, Kai Chen, Greg Corrado, Jeffrey Dean : Efficient Estimation of Word Representations in Vector Space, ICLR (2013)
- (11) 工藤 拓, 松本 裕治：“チャンキングの段階適用による日本語係り受け解析”，情報処理学会論文誌, pp. 1834-1842 (2002)
- (12) 伊藤 慎治：“記述式小テストの解答群の分類—内容によるタグ付け方法の検討—”, 平成27年度電子情報通信学会東海支部卒業研究発表会
- (13) MeCab: Yet Another Part-of-Speech and Morphological Analyzer,
<http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html>, 2016年3月8日参照