

記述式小テストの解答群からの類似解答の抽出法に関する一検討

服部 優季・伊藤 慎治・高瀬 治彦・北 英彦・川中 普晴
Email: 417m230@m.mie-u.ac.jp

三重大学工学研究科電気電子工学専攻

◎Key Words 記述式小テスト, 類似解答, 解答抽出

1. はじめに

講師は効果的な授業を行うために、学習者の理解状況を把握する必要がある。しかし、日本の教育現場では、学習者からの反応が得づらいという問題がある。そのため、学習者達の理解状況を把握するためには、講師からの能動的行動が必要である。

理解状況を把握するための代表的な能動的行動の一つとして小テストが挙げられる。ただし本研究では、小テストを成績評価のためではなく、学習者達の理解状況を把握するためのものとして扱う。小テストが出題された際、学習者は何らかの解答をしようとする。そのため、学習者からの反応が得やすく、理解状況を把握する手がかりとなる。実際に授業改善の方法として多くの教師が利用している⁽¹⁾。また、小テストは学習者に授業参加を促す一因となることも示唆されている⁽²⁾。

小テストの代表的な種類としては、多肢選択式・穴埋め式・記述式の3種類が考えられる。多肢選択式・穴埋め式では学習者が推測で解答できるため、理解状況を把握することは難しい。講師はそれを防ぐために入念な準備(選択肢の解答・穴埋め文章の作成)をする必要があり、大きな負担となる。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。対して記述式の小テストは、多肢選択式と反対の特徴を持つ。記述式では学習者が自分で文章を作成するため、推測だけでは解答が完成しない。加えて、解答文中に理解状況に関する情報が多く含まれる。しかし、講師が学習者達の理解状況を把握するためには、一人ひとりの解答文を読む必要があり、その負担は重い。また、解答を読むことに多くの時間を割いてしまうと、即時的なフィードバックが不可能になってしまう。

そこで我々は記述式小テストにおいて講師が解答を閲覧する際の負担を軽減する支援システムの構築をめざす。その第一歩として本稿では解答群から類似解答を抽出する方法について検討する。ある解答において講師が気になった部分をもとに、解答群から類似解答を抽出することができれば、講師は解答群を容易に整理することができるようになる。

2. 記述式小テスト実施時の支援

本章では、想定する記述式小テストの内容と支援内容について述べる。

まず、想定する記述式小テストの概要を以下に示す。

1. 講師は、講義中に学習者の理解状況を知りたくなった時、記述式小テストを出題する。
2. 学習者は、講師の出題を基に1~2分程度で簡単

な解答文を作成し、解答を出題する。

3. 講師は学習者ごとの解答文を読み、その場でフィードバックの内容を検討する。
4. 該当する学習者・または全員へフィードバックを返す。

これらの手順の中で、手順3.において、講師は学習者一人ひとりの解答文を読み、学習者の理解状況を把握する。この際、学習者一人ひとりの理解状況に加えて、講義全体での理解状況も把握する必要がある。

そのためには、講師がある誤りを含む解答を発見した場合、同じような誤り方をしている学習者がどの程度いるかといった情報が必要となる。解答を読むだけでなく、解答群の整理も行うことは講師にとって大きな負担となる。さらに、一つの解答に含まれる誤りは種類とは限らないため、整理すること自体が難しくなることも予想される。そこで、講師がある解答から気になる記述を発見した際に、同様の記述を含む類似解答を計算機により抽出することができれば、講師が解答を閲覧する際の負担の軽減が可能となる。

3. 類似解答の抽出方法

本章ではまず、類似解答の抽出における基本的な考えについて述べる。そしてその問題点と解決する手法を提案する。

3.1 単語による抽出方法

文章の内容は単語の集合 (bag of words) で表すことができる。そのため、類似解答を抽出する方法として、解答に含まれる単語を用いることが考えられる。講師が気になった内容を記述するために必要な単語を含んでいるかどうかにより、類似解答を判別する。判別する方法は、単語の一致数をもとに(1)式で表さる類似度を用いる。そして、類似度が同じ値になった解答を、講師が気になった内容の観点で、近い意味の解答をしている解答グループとする。これにより、講師が気になる内容の文と類似した解答の内容の解答を抽出することができる。

$$\text{類似度} = \frac{n(A \cap B)}{n(A)} \quad (1)$$

式中のAは講師の着目内容を構成する単語、Bは解答を構成する単語である。

しかし、記述式小テストにおいては、同じ講義を受けた学習者が解答文を作成するため、多くの学習者の解答文では似通った単語が使用される。また、語の使い方は個人によって異なるため、同じ単語が使用されていても同じ内容の記述とは限らない。そのため、単語の有無により類似解答の抽出を行うと、余分な解答まで抽出されてしま

う。そこで、単語そのものだけではなく、単語の使われ方にも着目する。

3.2 修飾・被修飾の関係を持つ文節対の活用

単語の使われ方を得る手法として、文節対に着目した手法⁴⁾が提案されている。この手法では修飾・被修飾の関係にある文節対をタグとし、これの集合が文章の内容を表すとしている。このタグを類似度の計算に用いることで、単語の使われ方も考慮した抽出が可能となる。

ただし、単にタグ付けを行うだけでは、多種類のタグが生成され、一致するタグが少なく検索できないといった問題がある。これは類義語の存在により、同じ意味合いを持つ記述に対しても別々のタグが生成されてしまうことに起因する。この問題を解決するために類義語の集約を行う。

3.3 類義語の集約

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がある。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での判断をする類義語辞書では、十分ではないと考える。

そこで、単語同士の類似度を算出できる WordNet⁵⁾に注目する。WordNet は概念辞書であるが、単語が synset と呼ばれる同義語のグループに分類されており、それらが他の synset (同義語のグループ) と意味的に結びついている。これらは上位下位のような関係性を持ち、構造のように存在し、概念の関係を表している。このような synset の仕組みを活かし、概念同士の距離を計算することで、類似度を得ることができる。

この類似度を活用することで、本来近い意味で使用したはずの語が、別の語の扱いにならずに同一視することができ、多様な記述への対応が可能となるだろう。

以上をまとめ、類義語を集約する処理を行ったタグを用いて類似解答の判別を行う。類似度は (1) 式の A を講師の着目内容を構成する文節対・単語に、B を解答を構成する文節対・単語として計算する。

4. 実際の解答群を用いた評価実験

本節では、提案手法として使用した、タグのしよと、類義語の集約の有効性を確認することを目標として、実験を行う。

4.1 実験内容・評価方法

比較対象として、表 1 のような 4 つの手法での比較を行う。これら 4 つで比較する意図としては、提案手法では文節対による単語の使われ方への着目と、類義表現への処理の 2 つの要素を含んでいるためである。そのため、各々の効果を確認すること、およびそれを組み合わせた提案手法、これらをせずに単語だけで検索した場合で比較を行うことが最適だと判断した。

本実験では、三重大学・工学部・電気電子工学科で実際に行われた講義「計算機工学 I」の中間テストで出題された問題を使用する。具体的な内容を表 2 に示す。今回の評価実験において、講師が実際に気になった記述内容は、「マクロ命令をマイクロ命令に変換する」であった。その

表 1: 比較対象とした手法

分類方法	文節対の使用	類義語の集約
提案手法	○	○
手法 A	○	×
手法 B	×	○
手法 C	×	×

表 2: 評価実験に使用した小テスト

問題	マイクロプログラム制御方式とはどのような制御方式かを、40 文字以上、80 文字以下で説明しなさい。ただし、「メインメモリ」、「マクロ命令」、「マイクロ命令」を説明に用いること。
模範解答	メインメモリに格納されているマクロ命令を CPU が読み込むたびに、マイクロ命令群からなるマイクロプログラムに変換して実行する方法。
出題年数	2017 年度
解答した学習者数	99 人

ため、講師には、筆者が解答群中から白紙ではない解答を無作為に 42 人分抜粋した解答に対して、「マクロ命令をマイクロ命令に変換する」ことが記入できているかどうか、一人分ずつ解答文を確認し、正誤の判断を行っていただいた。これら 42 人分の解答の中に、正しく記述できていると判断された解答は 8 人分であった。よって、今回の評価実験においては、これら 8 人の解答を適切に抽出できることを確かめる。また、本実験において、学習者の推敲した解答文は誤字・脱字を含めて変更せず、提出された文章をそのまま用いている。

評価には、(2)、(3) 式に示す再現率 X、適合率 Y を用いる。再現率 X は、講師が気になった内容を解答文中で記述している学習者を多く抽出できるかどうかの評価をする。この値が小さい場合、類似解答の抽出として不十分だといえる。適合率 Y は、解答の分類が過不足なく行えていることを評価する。この値が小さい場合、適切な学習者を分類できていても、見なければいけない解答総数は多くなり、講師の負担軽減としての効果が期待できない。また、再現率 X と適合率 Y は相反するものであり、解答を多く分類すれば、再現率は上昇する一方で、適合率は下降する。そのため、最も適切な分類ができていないか確認する方法として、これら 2 つの指標を合わせた評価としたい。

$$\text{再現率 } X = \frac{n(C \cap D)}{n(C)} \quad (2)$$

$$\text{適合率 } Y = \frac{n(D)}{n(E)} \quad (3)$$

式中の C は講師が気になった解答の総数、D は適切に分類できた解答の総数、E は類似度が最上位に分類された解答の総数である。

表 3: 評価結果

	提案手法	手法 A	手法 B	手法 C
適合率[%]	100	100	50	75
再現率[%]	50	25	37.5	37.5

表 4: 正しく分類できなかった解答一覧

学習者	解答例
イ	メインメモリに格納されている命令によってマクロ命令が作られる。マクロ命令には数個のマクロ命令に分割される。
ロ	メインメモリに格納されているマクロ命令を取り出し、そのマクロ命令を執行するのに必要なマイクロ命令を取り出し実行する。よって複雑な操作を行うことができる。
ハ	与えられたマクロ命令を、メインメモリ内でさらに多くのマイクロ命令に分けて実行する制御方式のこと。
ニ	メインメモリに格納されたマクロ命令を制御メモリに格納されるマイクロ命令にしてから実行する制御方式。

4.2 再現率・適合率による評価

実験のそれぞれの方法ごとの評価結果を表 3 に示す。再現率・適合率に関して、それぞれの結果を述べる。適合率に関しては、提案手法と手法 A (文節対を用いた分類) がともに 100% となった。再現率に関しては、提案手法が 50% と一番高い結果となった。文節対を用いた分類の適合率がともに 100% となっていることから、文節対は単体でもある程度の効果をもたらしている。これに対し類義語の集約は文節対と組み合わせることで効果をもたらした。

次に再現率は提案手法が最も高いが、それでも講師が気になった内容の解答の内、半分しか適切に分類できていない。ただし、再現率と適合率は相反するものであるため、再現率を高めるためにはいかに適合率を下げないかが重要となる。

4.3 正しく分類できなかった解答に対する考察

本節では、今回の実験結果から、再現率が 50% になったことに関する考察をする。理由として、解答群中で使われていた語が、意味が近いと判断するのが難しい語であり、類義語の処理が上手くできていなかったことが考えられる。表 4 は、適切に分類できなかった解答の詳細な内容である。今回、実験の際に出題者が気になった内容としたものは「マクロ命令をマイクロ命令に変換する」ことである。表 4 を見ると、学習者イ、ハは「変換する」ことを「分割する」「分ける」と述べており、学習者ロ、ニは「変換する」を「取り出す」「する」と述べていることがうかがえる。つまり、出題者は小テスト内での語の意味として、これらを同義と見なしてもよいと判断しているといえるが、日本語 Wordnet による類義語の判断ができていないのが原因だろう。これらの語は使う場面に応じて意味が変化してしまう可能性が高いと考えられ、自動的に処理することは難しい。よって、このような判断の難しい類義語をどのように扱っていくかも今後の課題である。

5. まとめ

本研究では、類似解答の抽出方法として、タグ付けと類義語の集約の 2 つの技術を利用する手法を提案・検証を行った。この結果、講師が気になった内容を含む類似解答を抽出することができた。この抽出結果を基に適切な解答群を講師に提示すれば、講師の解答閲覧の負担が軽減できる可能性を示した。

参考文献

- (1) 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)
- (2) 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)
- (3) Moodle.org:“open—source community—based tools for learning”, <http://moodle.org>, (2014)
- (4) 伊藤慎治, 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類 -内容によるタグ付け方法の検討-”, IEICE 卒業研究発表会, D-4-2 (2016)
- (5) Kyoko Kanzaki, Francis Bond, Noriko Tomuro, Hitoshi Isahara: "Extraction of Attribute Concepts from Japanese Adjectives.", In LREC-2008, Marrakech. (2018)
- (6) 川島貴広, 石川勉:「言葉の意味の類似性判別に関するセンサーラスと概念ベースの性能評価」, 人工知能学会論文誌, 20 巻 5 号 B, pp.326-336 (2005)