記述式演習における解答群からのフレーズによる類似解答検索手法の比較検討

二村駿輔，高瀬治彦，北英彦
Email: 422M240@mie-u.ac.jp

三重大学大学院工学研究科電気電子工学専攻

◎Key Words　記述式演習，自然言語処理，類似判定

1. はじめに

教育において学習の指導を行う講師は学習者に対して質の高い講義を行うことが求められている。そのためには学習者の理解状況を講師が把握し，それに適したフィードバック（解説、改善）を行うことが必要である[1]。しかし，現在の教育においては多人数講義が主であり，多人数講義において学習者一人一人の理解状況を把握することは講師の負担，講義時間を考えると非常に困難である。その中で問題演習を課し，その解答から学習者全体の理解状況の把握を行っている。問題演習にも，穴埋め式や記述式，選択式などの種類がある。

特に，記述式演習が学習者本人の言葉で解答をするため理解状況が解答に現れやすく，理解状況の把握に適している。講師は，把握した状況をもとに適切なフィードバックを行うことができる。しかし，記述式演習は個々の解答を読まなければならず，多人数講義においては，講師が要点を把握することに時間がかかり，学習者へのフィードバックが遅れてしまう。このことによる学習効果の低下と講師への負担が大きいという点が問題である。

これをふまえて，本研究では記述式演習において講師が示したフレーズをもとに解答群からそのフレーズの意味を含む解答群を抽出することを最終目標する．本稿では特に指定したフレーズと解答の間の含意関係判定手法の検討を行う。この結果，講師が気になったフレーズを含む解答群を簡単に得られるようになり，学習者への素早いフィードバックを行えるようになる。

2. 類似文書の抽出

2.1 単語による類似文書抽出

指定したフレーズ(クエリ)を含む解答を抽出するタスクは，類似文書検索の一種とみなすことができる。古典的な類似文書の検索の手法のひとつに，クエリに含まれる単語をもとに，単純に単語検索をしたり，Bag of Words による単語出現回数をもとにした文書ベクトルのコサイン類似度による抽出したりする手法がある。これらの手法は，使用されている単語が共通であれば類似している文書であるという考えに基づいている。これらの手法では検索対象の文書群が多岐にわたる分野の文書からなる際には有効である。しかし，単語の有無に着目しているため，同義語の判断ができない，語順による意味の異なりを判定することができないといった欠点がある。

2.2 大規模言語モデルによる類似文書抽出

大規模言語モデルとは，多量の文書からそこに含まれる言語的な規則・意味を獲得した深層学習モデルである。BERT [2]，GPT [3] などさまざまなアーキテクチャ・学習方法に基づいたモデルが開発されており，自然言語処理分野の様々なタスクに利用され好成績をおさめている。特に，英語圏における自然言語処理の標準ベンチマークになっている GLUE [4] のうちのひとつである STS-B(テキストの意味的類似度判定) で高いスコアを出しており，本稿でめざす含意判定への適用も期待できる。

ただし，大規模言語モデルそれ自体は，主に与えられた文に続く文を生成するものであり，含意判定に用いるためには，判定用のニューラルネットワークを追加し，これを学習（ファインチューニング）することで対応する。

3. 記述式演習における類似解答の抽出

前節から本稿でめざす含意判定を，大規模言語モデルを使用し実現するためには，ファインチューニングが必要である。ファインチューニングは少量のデータで可能であると説明されることが多いが，それは大規模言語モデルを学習する際のデータと比べたものである。本稿で対象としているのは，授業中の演習の解答群であり，更に少ないデータ量である。講師の負担・文書のバリエーションを考慮したうえでファインチューニングが可能であるか不明である。

そこで次節では，実際の授業(数回分)で行った演習の解答群を用いファインチューニングした場合の，含意判定の性能を評価する。

4. 実験

今回は大規模言語モデルのひとつである BERT による含意関係判定を行い，その判定精度を評価する。また，単語に基づく判定手法で抽出することが困難な同義語や語順の変化による判定能力についても検討する。

4.1 使用する解答データ

評価に用いるデータおよび含意判定を行うためのファインチューニングに用いるデータには，実際の講義で行った演習の解答を用いた。用いたのは，三重大学工学部で実施した，2021 年度および 2022 年度の計算機工学の授業の演習の解答である。2021 年度の解答(256 文)を学習に用い，2022 年度の解答(151 文)を評価に用いた。なお，実験で用いた検索フレーズに対する正解含意判定結果として，授業を行った講師とその授業を受講したことのある生徒によって手作業でラベル付けを行った。使用した演習は「ハーバードアーキテクチャの長所と短所を記述せよ」

という問いで検索対象のフレーズは 5 種類とした。フレーズ 1 を「ボトルネックを回避する」，フレーズ 2 を「データのビット幅をそろえる必要がない」，フレーズ 3 を「ハードウェア構成が複雑になる」，フレーズ 4 を「コストが大きくなる」フレーズ 5 を「命令用とデータ用のメモリが分かれている」とした。フレーズと解答の含意関係の例を下記に示す(文頭〇は検索フレーズの意味を含み，×は意味を含まないとラベル付けされたことを示す)。

フレーズ 1（ボトルネックを回避する）

〇長所はフォン・ノイマンのボトルネックを回避できる

×長所としては、命令用とデータ用のバスを分離することで高速な動作を実現している

〇ノイマン型とは命令用のプログラムメモリとデータ用のデータメモリを分離している点が異なり、フォン・ノイマンのボトルネックを回避できる点やプログラムとデータのビット幅を揃える必要がないが優れているがその反面回路が複雑になりやすいという欠点がある

4.2 モデルのファインチューニング

使用した大規模言語モデル(事前学習モデル)は文献 [5] の日本語学習モデルを使用した。

ファインチューニングの手順として教師入力データとして，判定対象の解答文と検索フレーズを SEP トークンで繋げたものを入力し，人手で付与したラベルを教師出力として学習を行った。この際，最適化手法は最適化手法として AdamW を利用し，学習率は 2×10^{-5}，バッチッチサイズは 32，エポック数は 50 とした。

4.3 フレーズに対する検出性能

4.2 節でファインチューニング結果(モデル A)を利用して 2022 年データの含意関係判定を行った。学習時と同様に解答を入力し，判定結果を得た。フレーズごとの正答率を表 1 に示す。いずれのフレーズに対しても高い正答率を示している。

表 1　フレーズごとの正答率

フレーズ	正答率
ボトルネックを回避する	0.97
データのビット幅をそろえる必要がない	0.99
ハードウェア構成が複雑になる	0.99
コストが大きくなる	0.99
命令用とデータ用のメモリが分かれている	0.97

4.4 同義語に対する検出性能

単語を基準にした検索では検索することができない同義語に対する検出性能を確かめるために，疑似的に解答を作成し，モデル A を用いて含意判定を行った。「コスト」や「大きくなる」という単語と同義語であり 2021 年度の解答に含まれていない語である「費用」，「経費」，「負担」「増加」「掛かる」を含む解答を作成し「コストが大きくなる」という意味を含意していると検出できるか判定を行った。その解答と結果を表 2 に示す。

表 2　疑似解答の判定結果

疑似解答	含意判定	
	人手	自動
コンピュータの設計のための費用が増加する	〇	〇
ハードウェア作成にかかる費用が大きくなってしまう	〇	〇
欠点としてコストが増加してしまう点が挙げられる	〇	〇
バスを二本にしているためより複雑になり、設計の負担が大きくなる	〇	〇
ノイマン型に比べて多くの経費が掛かってしまう	〇	〇

「コスト」，「大きくなる」が同義語に入れ替わっても含意判定ができていた。この結果から学習していない同義語に対する検出も可能だといえる。

4.5 語順の違いによる検出性能

次に使用単語が同じでも語順により文意が変化する場合について調査するために，疑似的に解答を作成し，モデル A を用いて含意判定を行った。作成した疑似解答は，「ハードウェア」，「複雑」というふたつの単語を文に含みながら「ハードウェア構成が複雑になる」という意味を含まない解答を作成した。その解答と判定結果を表 3 に示す。

表 3　疑似解答の判定結果

疑似解答	含意判定	
	人手	自動
新しいハードウェアの組み立ては、予想以上に複雑な手順を要します	〇	〇
複雑なデータを扱う上に、ハードウェアに負荷がかかる	〇	〇
複雑な動作を実現できるがハードウェアのコストが大きくなってしまう	〇	〇
ハードウェアの構成を大きく変えて、複雑なデータを扱っている	〇	〇
ハードウェア構成が大きく変わることなく、メモリが複雑に分かれていない点	〇	〇

表 3 から，語順により意味が変化した場合は，正しく判定できなかったことがわかる。この結果から語順による意味の違いには対応が困難だと考える。ただし，2.2 節で示したように，ベンチマークでは高い含意判定性能を示しているので，学習したデータの量・学習手法の変更により改善する可能性はあるだろう。

4.6 許容しない同義語の取り扱い

最後に，講師によって異なる含意の基準に対応できるのかを検討する。モデル A では，フレーズ 3 において，「ハードウェア構成」と「回路構成」を同義語とみなして学習した。講師により「回路」を許容しない場合を想定して，「回路構成」と記した解答はフレーズ 3 を含意しないとラベル付けしたデータを用いて学習したモデル(モデル

B)を用いて，2022 年度の解答群についてフレーズ 3 を含意しているか判定した。モデル A を用いた場合の判定果を表 4 に，モデル B を用いた場合の判定果を表 5 に示す。なお，表 5 の正解ラベルも，モデル B の学習データと同じ基準でつけ直している。

表 4　モデル A によるフレーズ 3 の判定結果

		予測		解答合計
		○	×	
正解	○	71	0	71
	×	1	80	81

表 5　モデル B によるフレーズ 3 の判定結果

		予測		解答合計
		○	×	
正解	○	18	3	21
	×	1	130	131

表 4，5 より，学習段階から「ハードウェア」と「回路」を同義語でないとして学習したモデル B では含意しているものを含意していると予測できる数が減少した。個々の事例を見ていくと，解答の中に「ハードウェアの回路構成が複雑になる」といった二つの意味を含んでいる解答で誤って予測を行っていた。そのほかの解答では「ハードウェア」と「回路」を異なるものと認識として含意判断を行えていた。このことから，許容しない同義語の指定は，学習データを適切に用意することで可能になると考えられる。

5.　考察

前節の実験結果より，今回ファインチューニングで用いた程度のデータ量(1 年分)であっても，学習したフレーズに関連した含意判定は，同義語を許容する部分までは可能になった。ただし，語順による意味の変化には対応できなかったので，今後の検討が必要であろう。また，ファインチューニングの際に学習しなかったフレーズへの対応もできていないので，この点についても併せて検討が必要である。

6.　おわりに

本稿では記述式演習において講師が示したフレーズをもとに解答群からそのフレーズの意味を含む解答群を抽出するためにフレーズと解答の含意判定を目的とし，実際の講義の解答データをもとに深層学習由来の BERT による手法の検討を行った。この手法では単語検索や Bag of Words では検索が困難な同義語による類似解答の検索の可能性が示すことができた。しかし，教師データ作成のコストや教師データのないものに関してはまだまだ課題がある。今後は学習するデータ量や，種類に着目して語順による意味の異なりが判別できるかも検討していきたい。

参考文献

(1) 中島英博：“多人数講義で学生の深い学習を促す教員の特質”，名古屋高等教育研究，Vol.15，pp.161–177 (2015).

(2) Jacob Devlin， Ming-Wei Chang， Kenton Lee and Kristina Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding， Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies， Vol.1， pp.4171-4186 (2019)

(3) Alec Radford， Karthik Narasimhan， Tim Salimans and Ilya Sutskever. Improving Language Understanding by Generative Pre-Training 2018

(4) Alex Wang， Amanpreet Singh， Julian Michael， Felix Hill， Omer Levy & Samuel Bowman1. GLUE: A MULTI-TASK BENCHMARK AND ANALYSIS PLATFORM FOR NATURAL LANGUAGE UNDERSTANDING， Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP， pp.353-355 (2018)

(5) GitHub - cl-tohoku/bert-japanese: BERT models for Japanese text. https://github.com/cl-tohoku/bert-japanese 閲覧日 2023 年 6 月 30 日