

# 文章構造グラフを用いた 国語記述式答案への自動フィードバック生成

岩瀬裕哉<sup>1,2</sup> 舟山弘晃<sup>1,2</sup> 松林優一郎<sup>1,2</sup> 乾健太郎<sup>1,2</sup>

<sup>1</sup> 東北大学 <sup>2</sup> 理化学研究所

{yuya.iwase.t8,h.funa}@dc.tohoku.ac.jp {y.m, kentaro.inui}@tohoku.ac.jp

## 概要

学習者は自らのアウトプットに適切なフィードバックを受けることで効果的に学習を進められることが知られている。本研究では、国語科目的長文読解問題における記述式答案に対してフィードバック文を生成することで、学習者のより良い学びと教育指導者の負担軽減を目指す。その手法として、問題本文の文章構造や論理関係を利用して、答案の部分箇所に対し採点項目ごとにフィードバック文を生成する手法を提案する。実験では、既存の採点済みデータセットを用いて、この手法を用いたフィードバック生成が現実的に実現可能かを確かめた。

## 1 はじめに

数十語程度を答案に記す記述式問題は、教育現場で広く用いられる問題形式であるが、教育指導者にとって記述式答案の採点や答案にフィードバックを与える作業は多大な労力を伴う [1]。適切なフィードバックは学習者の学びに有益である反面、実際にそれを得られる機会はコストの問題で限定されている [2]。本研究では、学習者が指導者の負担無しに効率的に学習を進められるフィードバック生成システムの構築を目指す。

記述式答案の自動採点技術は実用化に向け着実に研究が進められている [3, 4, 5]。水本らの研究 [3] は、採点項目を利用して国語記述式問題の自動採点を定式化した。同研究では、複数ある採点項目ごとに採点の根拠箇所と得点を出力する。これによって、学習者は採点結果を見てどの部分を修正すれば良いかというフィードバックを得ることができる。本研究では、この水本らの研究を出発点として、より学習者の学びを助けるフィードバック文の生成を目指す。

フィードバックの学習への効果について詳細に検討した Hattie and Timperley 2007[6] の研究によれば、

フィードバックは目標と現在の理解の間の乖離があるとき学習に効果的に働く。ここで述べられる効果的なフィードバックとは、目標が何で、現時点でどこまで進んでいて、次に何をすれば良いか、という情報を含むものである。本研究はこの効果的なフィードバックの枠組みを国語記述式問題のフィードバックに応用することを目指す。ある答案を入力として、効果的なフィードバックに必要な情報を得る手段として、記述式問題に用いられる本文の文章構造を利用するなどを提案する。このアイデアは、国語の授業中や問題集の解説書等で行われている、本文を分割し、それぞれの部分の間の論理関係を構造化して問題を解説する方法に着想を得ている [7]。

文章に記された情報を構造的に理解する能力については、近年の高等学校での論理国語科目導入などに見られるように注目を集めている。記述内容を正確に読み取る能力は、日常生活でも必要となる重要な能力であり、それらを効率的に訓練できるツールには一定のニーズがあると考えられる。本研究では、このような背景から、評論文や論説文と呼ばれる比較的論理構造が明確な文章を題材とした問題へのフィードバック文生成に限定して取り組む。

学習者に対するフィードバック文生成に関する研究として、英語学習者のライティングにおける文法的な誤りへの解説文生成が挙げられる [8]。Lai and Chang 2019 [9] は、テンプレートを利用した解説文生成という点で本研究と関連性がある。

本研究では国語記述式問題のフィードバック生成研究の第一歩として、既存の国語記述式問題のデータセットを利用して、問題本文の構造を利用したフィードバック生成というアイデアが技術的に可能であるかを確かめる。構造を利用するための方法として、答案と本文の間で、それぞれを表現したベクトルのコサイン類似度を計算し、答案が本文中のどの部分を参照して記述されたものか推定する

図 1 データセット内のデータ例

手法を提案した。そして、本文の構造から得られた情報と、フィードバック文のテンプレートを用いてフィードバック文を生成した。結果として一部生成に適さない問題と答案はあるものの、アイデアが現実的に実現可能であることが確かめられた。

## 2 データセット

本研究では、理研記述問題採点データセット代々木ゼミナールデータ [3, 4] を題材として記述式答案に対するフィードバック生成を行う。このデータは、高校生を対象に行われた模擬試験の国語長文読解記述式問題について、問題と答案、および、人間の採点者がつけた採点項目ごとの得点と、採点の根拠箇所についてのアノテーションを含むものである(図 1)。採点項目とは採点の際に利用される採点基準のこと、問題ごとに 5 項目程度存在する。採点は、各採点項目で指定された内容が答案に記述されているかという観点で行われる。採点の根拠箇所として、採点項目ごとに答案中のサブワードに対して、点数の根拠となった部分に 1、その他に 0 が付与されている。データセットは論説文や小説など複数の文体を題材とした問題を含み、また問題の出題形式も本文中に引かれた傍線部を言い換えて説明するものや、傍線部がなく、本文から要約的にまとめるもの等いくつかの形式がある。本研究ではその中でも、問題本文の論理構造を利用したフィードバックというアイデアに即して、論説文を扱ったもので、かつ本文中の言葉を使い、傍線部について説明するような問題 3 題を扱う。

### 3 フィードバック生成タスク

上述のデータセットの採点は、各採点項目に対して記述内容に応じた部分点を与える方式である。本研究で答案に対し生成するフィードバックは、この採点の設定に即し、採点項目ごとに output する。ある問題に対する答案が与えられたとき、答案は採点項

目ごとに関連する内容が記述された箇所が特定され、その内容に応じて点数が付けられる。フィードバックの完全自動化のためには採点および根拠箇所の特定についても何らかのモデルを開発する必要があるが、本研究では文章構造を用いたフィードバック生成という我々のアイデアの実現可能性を検証することに焦点を絞り、採点根拠箇所についてはデータセットに付与されている人間の採点者によるアナテーションを利用する。したがって、本研究で設定するフィードバック生成タスクは、答案と、ある採点項目のペアに対し、その採点項目に対して特定される根拠箇所の文字列が与えられるとき、その根拠箇所の内容に対して適切なフィードバック文を生成することと定義される。

## 4 手法

我々の手法の外観を図 2 に示す. 本研究では, 設問で問われている内容を本文に即して説明するような記述式問題を扱う. この問題形式では, 答案が得点するために, 本文に書かれている情報を答案に含む必要がある. したがって, 得点の根拠箇所は本文に含まれる特定の箇所を参照している. そこで, 我々はまず本文を適切な単位に区切り, 答案の根拠箇所と, 本文の各分割単位との類似度を計算することで, その根拠箇所が本文のどの部分を参照しているのか特定する(図 2(1)).

次に、本文について、分割された各部分をノードに持ち、関連する部分同士の論理関係を示すラベルをエッジに付与した文章構造グラフを作成しておき(図 2(2))、この文章構造グラフの上で、答案内に理想的に含むべき情報が書かれている位置と、答案内に書かれている内容の位置がどのような関係ラベルで接続されているかという情報に基づき、フィードバック文のテンプレートを選択し、そこからフィードバック文を生成する(図 2(3))。

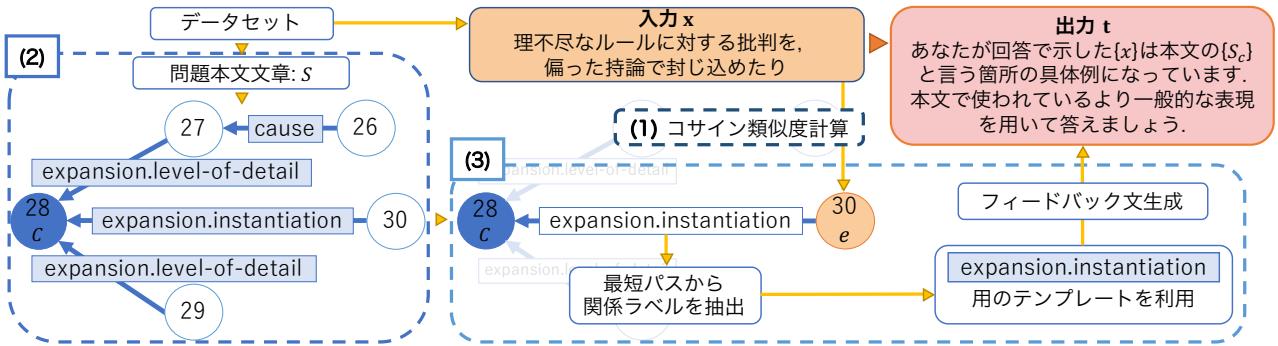


図2 提案するフィードバック生成手法の外観図

#### 4.1 答案根拠箇所の本文参照箇所推定

答案の、ある採点項目に対する根拠箇所文字列  $x$  を入力として、その根拠箇所が本文中のどの部分を参照して記述したものか推定した結果  $e$  を出力とする。 $e$  は本文の分割単位それぞれに割り振られたインデックスを表す。

答案の根拠箇所  $x$  を訓練済み日本語 BERT[10] に入力し、最終層の埋め込み表現を抽出する。埋め込み表現のベクトルの和をとることで、 $x$  に対する根拠箇所ベクトル  $y$  を得る。

本文を分割して作成された文字列集合  $S = \{s_1, s_2, \dots, s_m\}$  に対しても同様の手続きで本文ベクトル集合  $H = \{h_1, h_2, \dots, h_m\}$  を得る。なお、 $y$  および  $h_m$  は次元数 768 のベクトルであり、 $m$  は本文の分割数を表す。根拠箇所ベクトル  $y$  と本文ベクトル集合  $H$  の各要素との間でコサイン類似度を計算し、 $y$  に対して最も高い値を返した  $h_e$  を、 $x$  が参照した本文箇所の推定結果  $e \in \{1, 2, \dots, m\}$  とする。

#### 4.2 本文の構造グラフを用いたフィードバック文生成

**本文構造グラフ** フィードバック生成のために用いる本文構造グラフ  $G$  を人手で作成する。分割された本文の集合  $S$  の要素をグラフのノードとする。グラフの作成に当たり、まず採点基準や模範解答をもとに答案が満点を獲得するために参考すべき箇所を本文から特定し、そのノードのインデックスを正解  $c$  とする。次に、本文中で談話関係のあるノードペアについて、関係の種類を表すラベルをエッジに付与する。関係ラベルには Penn Discourse Treebank Version 3.0 のタグセットを採用した[11]。

本文全体にこのようなアノテーションを行うには多大な労力を要する。また、各採点項目において答案が得点するために含むべき情報は本文中の一部分

に集中的に記載されている。そこで、採点基準をもとに各採点項目と関連が強い箇所に限定して談話関係アノテーションを行い、部分構造グラフ  $G'$  を採点項目ごとに作成した。図2(2)はそのような部分構造グラフの例である。図中の青いノードは正解ラベルのノードであり、各エッジにはそれぞれ関係ラベルが付与されている。

**フィードバック文生成** 答案の根拠箇所が参照している本文ノード  $e$  と、正解ノード  $c$  の間を結ぶエッジに付与された関係ラベルに基づき、あらかじめ定めておいたフィードバック文のテンプレートのうち適切なものを選択する。具体的には、本文部分構造グラフ  $G'$  上で、根拠箇所が参照しているノード  $e$  と正解ノード  $c$  の間の最短パスを求める。次に、パスに含まれるエッジのうち、正解ノードと隣接するノードを結ぶエッジに付与されたラベルを得る。得られたラベルによりフィードバックの生成に用いるテンプレートを選択する。

さらに、各答案の内容に応じて柔軟にフィードバック文を生成するために、答案の根拠箇所の文字列  $x$  と正解ノード  $c$  の本文文字列  $s_c$  をテンプレートに埋め込み、最終的なフィードバック文  $t$  とする。ただし、答案の根拠箇所の文字列  $x$  と最も対応する本文ノード  $e$  がこの採点項目のために設計した部分構造グラフ  $G'$  に存在しない場合、および  $e$  が正解ノードと一致する場合には、それぞれ固定文のフィードバックを生成した。テンプレートの具体例は付録Aに示す。

## 5 実験

### 5.1 実験設定

本研究の目的は学習者に取って有益なフィードバックを生成することであるため、対象の採点項目

**表1** 答案根拠箇所に対し生成されたフィードバック文の例

| 答案根拠箇所                       | 生成されたフィードバック文   |
|------------------------------|---|
| 理不尽なルールに対する批判を偏った持論で封じ込めたり   | あなたが回答で示した「理不尽なルールに対する批判を偏った持論で封じ込めたり」は本文の「(本文正解箇所)」と言う箇所の具体例になっています。より本文で使われている一般的な表現を用いて答えましょう。 |
| メディアに有利なルールを作り               | あなたが回答で示した「メディアに有利なルールを作り」は本文の「(本文正解箇所)」と言う箇所の具体例になっています。より本文で使われている一般的な表現を用いて答えましょう。             |
| 理不尽なルールができてしまったりルールを変えてしまったり | 本文の参照箇所は合っています。問題文をよく読んでより適切に答えましょう。問題で問われている事に答えていません。本文と問題文を丁寧に読みましょう。                          |

に対して満点を得ている答案は扱わない。また、0点の答案についても根拠箇所が付与されないため扱わない。適切に提案手法の評価を行うために、上記の条件を満たす答案を10件以上含む採点項目のみを用いた。**3**節で述べたように、本研究の焦点は我々が設計したフィードバック生成システムが実現可能か検証することにある。したがって、各採点項目の根拠箇所情報は、データにあらかじめ人手で付与されたものを入力として用いて実験を行った。

## 5.2 参照箇所推定結果の妥当性の確認

**4.1**節で説明した方法によって、 $e$ をどの程度正しく推定できるかを確かめるために、入力の各根拠箇所  $x$  に対して正解の  $e$  の値を付与した評価データを人手で作成し、推定結果との一致率を計算した。評価データは、対象の採点項目に対する根拠箇所からランダムに20個サンプリングしたもの(random)と、根拠箇所文字列の出現頻度を降順に並べ上位20種を抽出したもの(unique)について人手でラベルを付与した。結果を表**2**に示す。randomの結果からは、全6個の採点項目のうち、5個の採点項目においては7割程度の一致率で $e$ を推定できることが確かめられた。また、uniqueの結果からは、答案によく書かれる記述に関する精度にも採点項目ごとにばらつきがあることが示唆された。適切に推定できなかっただけでなく、根拠箇所が答案に含まれる部分文字列であることに対して、比較を行う本文のノードは文または節のように意味的に一つのまとまりを持つ単位であったという点が考えられる。また、問題2の一致率が低い原因是、当該の問題形式に起因していると考えられる。問題2は本文中のある文が示す内容について、本文に即して説明する問題であるが、得点するためには本文中の複数の箇所を参照して答案を作成しなければならない問題である。本手法では、ある根拠箇所が参照している本文中の分割単位が一つに定まる事を仮定して、手法の設計を

**表2** 参照箇所推定の精度。 $A=1$ は採点項目  $A$  が1点である答案を表す。uniqueの括弧内の値は頻度上位20件の表現が対象データ全体に占める割合。

|        | 問題1   |       | 問題2   |       | 問題3   |       |
|--------|-------|-------|-------|-------|-------|-------|
|        | B=2   | D=1   | B=3   | A=1   | D=2   | E=2   |
| random | .70   | .70   | .25   | .95   | .58   | .68   |
| unique | .60   | .75   | .25   | .71   | .74   | .85   |
|        | (.51) | (.71) | (.91) | (1.0) | (.25) | (.25) |

行ったため、このような問題において適切に本文の参照箇所を推定することができなかったと考えられる。このような種類の問題に対するフィードバック生成手法の構築は、今後の重要な課題である。

## 5.3 フィードバック生成

我々が**4**節で提案した手法によってフィードバックの生成を行い、その結果を検討した。表**1**に生成されたフィードバック文の具体例を示す。このうち、上の2つの例は、 $e$ が適切に推定できた場合に生成されるフィードバック文を表している。下の2つは適切に $e$ を適切に推定できなかった場合の結果を示している。この結果より文字列として似た答案であるにも関わらず、推定された $e$ の内容に応じて生成されるフィードバックが大きく異なってしまっている。したがって、我々が提案するフィードバックの生成手法では、 $e$ を適切に推定することが鍵である。今後は、学習者を巻き込んだテストを行うことを計画しており、生成されたフィードバックの効果を実践的な形で計測する予定である。

## 6 おわりに

本研究では、答案の根拠箇所と本文の文章構造を用いてフィードバック文を生成するという手法を提案し、その手法の実現可能性について確かめた。実験により、学習者の学びの助けになりそうなフィードバック文を生成できる可能性が確かめられたと同時にさらなる検討課題が明らかになった。

## 謝辞

実際の模試データを提供していただいた学校法人高宮学園代々木ゼミナールに感謝します。本研究はJSPS 科研費 JP19K12112, JST 次世代研究者挑戦的研究プログラム JPMJSP2114 の助成を受けたものです。

## 参考文献

- [1] Anna Filighera, Siddharth Parihar, Tim Steuer, Tobias Meuser, and Sebastian Ochs. Your answer is incorrect... would you like to know why? introducing a bilingual short answer feedback dataset. In **Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)**, pp. 8577–8591, Dublin, Ireland, May 2022. Association for Computational Linguistics.
- [2] Valerie J. Shute. Focus on formative feedback. **Review of Educational Research**, Vol. 78, No. 1, pp. 153–189, 2008.
- [3] Tomoya Mizumoto, Hiroki Ouchi, Yoriko Isobe, Paul Reisert, Ryo Nagata, Satoshi Sekine, and Kentaro Inui. Analytic score prediction and justification identification in automated short answer scoring. In **Proceedings of the Fourteenth Workshop on Innovative Use of NLP for Building Educational Applications**, 2019.
- [4] Hiroaki Funayama, Tasuku Sato, Yuchiroh Matsubayashi, Tomoya Mizumoto, Jun Suzuki, and Kentaro Inui. Balancing cost and quality: An exploration of human-in-the-loop frameworks for automated short answer scoring. In **Artificial Intelligence in Education**, 2022.
- [5] Haruki Oka, Hung Tuan Nguyen, Cuong Tuan Nguyen, Masaki Nakagawa, and Tsunenori Ishioka. Fully automated short answer scoring of the trial tests for common entrance examinations for Japanese university. In **Artificial Intelligence in Education: 23rd International Conference, AIED 2022, Durham, UK, July 27–31, 2022, Proceedings, Part I**, p. 180–192, 2022.
- [6] John Hattie and Helen Timperley. The power of feedback. **Review of Educational Research**, Vol. 77, No. 1, pp. 81–112, 2007.
- [7] 野矢茂樹. 論理トレーニング 101 題. 産業図書, 2012.
- [8] Kazuaki Hanawa, Ryo Nagata, and Kentaro Inui. Exploring methods for generating feedback comments for writing learning. In **Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing**, pp. 9719–9730, 2021.
- [9] Yi-Huei Lai and Jason Chang. TellMeWhy: Learning to explain corrective feedback for second language learners. In **Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP): System Demonstrations**, pp. 235–240, 2019.
- [10] 日本語 bert 訓練済みモデル. <https://github.com/cl-tohoku/bert-japanese>.
- [11] Rashmi Prasad, Bonnie Webber, Alan Lee, and Aravind

## A 作成したフィードバック文テンプレート例

表3 作成したテンプレート例

| ラベル                       | テンプレート  |
|---------------------------|---|
| Expansion.Instantiation   | あなたが回答で示した「 <u>答案根拠箇所</u> 」は本文の「 <u>本文正解箇所</u> 」と言う箇所の具体例になっています。より本文で使われている一般的な表現を用いて答えましょう。 |
| Expansion.Level-of-detail | あなたが回答で示した「 <u>答案根拠箇所</u> 」は「 <u>本文正解箇所</u> 」を詳細に記述したものです。より一般的な表現で答えましょう。                    |

本研究で作成したフィードバックの一例を示す。表3に示した通り、グラフ上に出現する関係ラベルに対して、答案の根拠箇所と本文の答案の参照にすべき正解箇所を埋め込むことができるフィードバック文を作成した。

表4 テンプレートに追加した固定文の例

| 推定ノードの位置     | テンプレート文                              |
|--------------|--------------------------------------|
| グラフ外にある場合    | 問題で問われている事に答えていません。本文と問題文を丁寧に読みましょう。 |
| 正解ノードが一致した場合 | 本文の参照箇所は合っています。問題文をよく読んでより適切に答えましょう。 |

本研究では、正解ラベルが付与されたノードと本文参照箇所推定結果ノードとの最短パス上にあるラベルを利用する手法を取るため、表4に示されたような場合にはパス上のラベルを得ることができず、手法通りにフィードバック文を生成することができない。そのため、それらの場合に対しては、表の通り埋め込み箇所のない固定文のフィードバックを返す設定とした。