E-008

BERT を用いた記述式問題の自動採点に関する研究
A Study on Automatic Grading of Short Answer Questions Using BERT

藤野 準平† **徳山 豪**†
Jumpei Fujino Takeshi Tokuyama

1 はじめに

2021 年度より，大学入学試験において大学入試センター試験に代わる大学共通テストが実施された. 大学共通テストには選択マーク問題に代わり記述式問題が導入される予定であった. 文部科学省は記述式問題を導入する意図として，学生の論理的思考力や判断力，表現力を評価することができるとしている. しかし，大学入学試験として実施するには様々な問題が考えられる. 大学受験のような受験者数が大規模な試験においては，記述式問題を採点する際，採点基準の一貫性や，採点者の労力，時間的なコストなど様々な問題が考えられ，短期間で数十万人もの解答に対して公平に採点することは非常に困難である. そこで，近年では記述式問題に対して機械学習や深層学習を用いて自動採点する研究が広く行われている．水本らの文献 [1] では，LSTM を用いた深層学習モデルを拡張しアテンション機構を利用することで，項目点ごとに採点しフィードバックを可能とするシステムを提案し非常に高い精度で自動採点することに成功している. また，文献 [2] では入力される学生の解答を BERT[3] を用いて分散表現ベクトルに変換し，Bi-LSTM に入力し全体点を出力するモデルの提案と，事後確率を用いて自動採点システムにおける採点の信頼性を表す確信度の有効性を示し予測の信頼性を推定できることを確かめている. 高い精度で採点できていることや，確信度を導入することからシステムの信頼性を担保することができているが，本当に文脈や否定語などの影響を考慮できているのかを確認する必要があると考えた.

そこで本研究では，自動採点モデルとして水本らのモデル [1] を参考に BERT を用いて文脈を得たのち，採点項目の数だけネットワークを付け Fine-tuning することで点数付けするモデルを実装した. そのモデルの評価といくつかのテストケースを用意し，モデルが自動採点する際，文脈や否定のようなものを考慮し採点できているのかを採点の違いから確認する.

2 BERT を利用したモデル

本研究にて扱うモデルは，水本らの提案していたモデル [1] を参考に BERT のみを利用し，事前学習済みの BERT に項目ごとのネットワーク層を追加し Fine-tuning した自動採点モデルである. 図 1 に，本モデルの概要を示す.Fine-tuning する際，事前学習済みの BERT は学習率を低めの 5e-5，追加した層の学習率は 1e-4 とした. 最終層である 12 段目の BERTLayer の Self-Attention を取り出し，それらを利用すると BERT モデルがどの部分に着目して採点したのかを可視化することができる.

3 実験

実装したモデルが自動採点に適しているかを評価する. 続いて，そのモデルが本当に文脈を読んだ上で自動採点しているのかをいくつかのテストケースを用意して確認する.

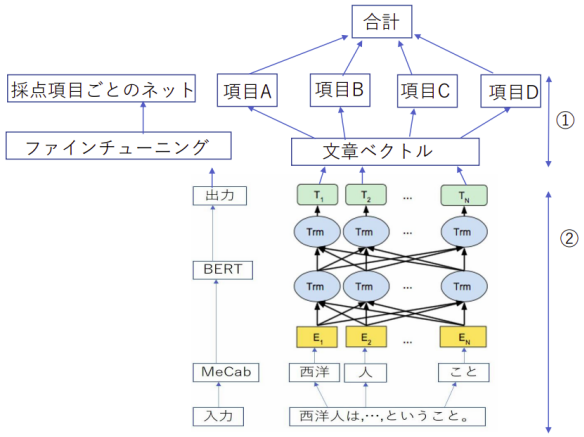

図 1 BERT を用いた自動採点モデル

3.1 実験の設定

データセットとして，理研記述式問題データセット [4] を利用する. データ数は 2000 件のうち，1600 件を学習データ，200 件を検証データ，200 件をテストデータとして 5 分割交差検証を行う. BERT は東北大学が公開している日本語の Wikipedia で事前学習を行ったモデルを用いている. 形態素解析は MeCab を用いて文章を分割する. 学習させる際，epoch 数は 15，データのバッチサイズは 16，損失関数は交差エントロピー誤差，最適化には AdamW を用いて学習を行う.

3.2 評価

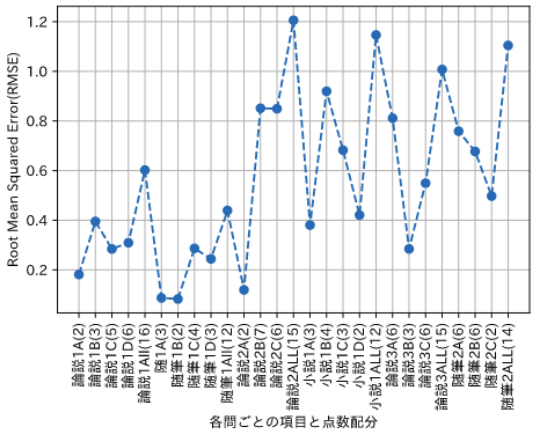

図 2 自動採点モデルの RMSE

評価としては，Root Mean Squared Error (RMSE) と Quadratic Weighted Kappa (QWK) を用いる. 図 2，3 より，本モデルは論説 1 において全体では RMSE が 0.6 点，QWK が 0.98 となり非常に高い精度で自動採点できていた. 文献 [1] にて報告されているモデルには劣るが，

† 関西学院大学大学院理工学研究科 Graduate School of Science and Technology, Kwansei Gakuin University

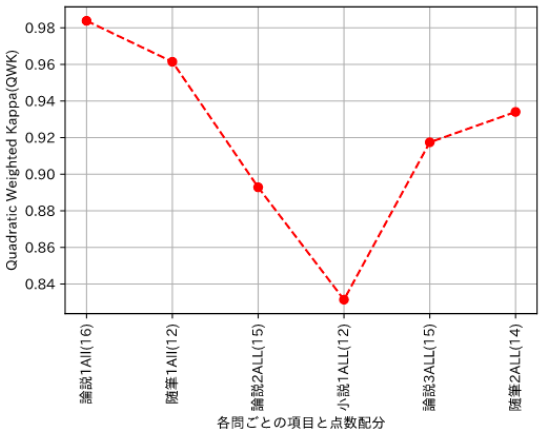

図 3　自動採点モデルの QWK

BERT を用いて Fine-tuing するモデルも自動採点には有効であることがわかった. また可視化した Attention 部分を見ると，解答の中の採点項目や似たような部分に注目していることが多く，点数分類の判断として有効であった. その他の問についても，比較的高い精度で自動採点することができたが文献 [1] でも報告されたように，自動採点が苦手な部分も存在した. 次節では論説 1 を自動採点したものを用いて，文章の文脈を変更した際の点数の影響を確認する.

3.3　文脈を見ているかのテスト

BERT は文脈を読むことを可能としており，書籍 [5] では英語の文章における二重否定を読み取ることができていると報告している. 日本語版の自動採点 BERT モデルが入力された文章の否定語などによって採点が影響するかを見る必要がある. そこで次の項目に関するテストケースをいくつか用意し，自動採点した場合採点結果がどのようになるかを調査する.

- 解答に関係のない日本語文章
- 文章として成立していない文章
- 模範解答に否定語などを導入し，意味的な文脈部分を変更した文章
- 採点項目に重要なキーワードをランダムに置換した文章
- 採点項目のみの羅列

解答に関係のない文章や文章として成立していない文章については，複数のテストケースですべて 0 点と採点した. 模範解答に否定語を導入し，文脈を変更した文章は導入していない場合と同点を出力した. 重要なキーワードを別のものに変換する. 例えば，”西洋では”を”東洋では”に変換したり，その他ランダムに置換した文章についても置換しない場合と同点を出力した. 採点項目のみ羅列した場合は満点を与えることはなかったがいくつか点数を出力し，キーワードから点数を得ているような結果がみられた. これらのことから，文脈的な効果は少なく，文章の構成や類似性から採点している可能性が高いことがわかった. 図 4 では，文章を変更した例とその Attention 部分を可視化したものの例であり，それぞれ同じ 16 点と出力された. 自動採点モデルの精度は高いが，文脈や否定語といったものに強いわけではなく学生にフィードバックを与える際，間違えた判断の上でフィードバックを与えてしまう可能性があることがわかった.

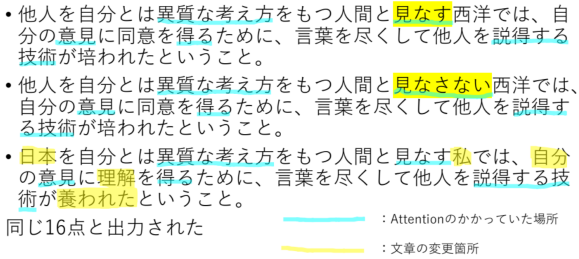

図 4　模範解答と語句をいくつか変更した際の自動採点例

4　まとめ

本研究では，BERT のみを用いて Fine-tuning したモデルを自動採点モデルとして評価し，自動採点する際に文章の文脈を読めているか確認した. 模範解答に否定語を加えることで文脈的な意味を変換したとしても，自動採点した際に点数の変化がなかったことから文章の類似性などを見ている可能性が高く、否定語などの影響は少ないと思われた. しかし，実用的なことを考えると否定語による文脈の変化やキーワードの違いから点数が変化することは必要である. 今後，これらの問題に対応できるように否定語や文脈の変化による影響度を考慮したい.

謝辞

本研究では，国立情報学研究所の IDR データセット提供サービスにより国立研究開発法人理化学研究所から提供を受けた「理研記述問題データセット」を利用した. また事前学習済みの BERT を提供してくださった東北大学乾研究室の方々に深く感謝いたします. この研究は科学研究費基盤研究 B20H04143 の支援をうけた.

参考文献

[1] 水本智也, 磯部順子, 関根聡, 乾健太郎. 採点項目に基づく国語記述式答案の自動採点. 言語処理学会第 24 回年次大会発表論文集, pp. 552-555, March 2018.

[2] 船山弘晃，佐々木翔大，水本智也，三田雅人，鈴木潤，松林優一郎，乾健太郎. 記述式答案自動採点のための確信度推定手法の検討. 言語処理学会第 26 回年次大会 発表論文集，, pp.997-1000, March 2020.

[3] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanoval. BERT: Pre-training of deep bidirectional transfomers for language understanding. 2018. https://arxiv.org/abs/1810.04805

[4] 理化学研究所 (2020):理研記述問題採点データセット. 国立情報学研究所情報学研究データレポジトリ. データセット: https://doi.org/10.32130/rdata.3.1

[5] 小川雄太郎 Pytorch による発展ディープラーニング 第 8-4 章 BERT の学習・推論、判断根拠の可視化を実装 株式会社マイナビ出版 東京都 2020 年 4 月 10 日