

短答記述式問題データセットに対するフィードバック生成手法の検討

A Feedback Generation Approach for Short Answer Question Datasets

PHAN Huy Thanh^{*1}, 谷 文^{*1}, 太田 光一^{*1}, 長谷川 忍^{*1}
Huy Thanh PHAN^{*1}, Wen GU^{*1}, Koichi OTA^{*1}, Shinobu HASEGAWA^{*1}

^{*1}北陸先端科学技術大学院大学

^{*1}Japan Advanced Institute of Science and Technology

Email: hasegawa@jaist.ac.jp

あらまし：本稿では短答記述式問題データセットを対象としたフィードバックの生成手法を提案する。適応的なフィードバックを実現するため、本研究では自動採点手法、類似正解文選択手法、差分フィードバック生成手法を組み合わせたフィードバックシステムを開発している。これにより、ドメイン知識を持たない短答記述式データセットのみでも学習者の支援に利用可能なフィードバックの実現が期待できる。

キーワード：短答記述式問題、データセット、フィードバック生成

1. はじめに

本文から指定された考え方や解釈を比較的短い文章で解答する短答記述式問題は様々な試験において利用されている。特に、大規模試験への適用時の採点の時間的・経済的コストの高さに対応するため、様々な自動採点手法が提案されている⁽¹⁾⁽²⁾⁽³⁾。近年、採点精度は非常に高くなっており、実環境での運用が期待されている。一方で、自動採点手法のアウトプットは一般に得点のみであり、学習者毎にどのように修正すればよいのかについて自動的にフィードバックを行うことは想定されていない。

こうしたフィードバックを実現するためには、大別すると、学習者の解答と教員のフィードバックをセットにしたデータセットを利用する方法と、本文等をドメイン知識として学習者の解答と教員の採点結果からフィードバックを生成するアプローチが考えられる⁽⁴⁾。しかしながら、前者については現時点で公開されているデータセットは限られており、後者についても問題文は本文の著作権の関係で公開されていないことが多い。そこで、本研究では学習者の解答と教員の採点結果を組とするデータセットから修正フィードバックを生成する手法について検討する。

2. 対象データセット

本研究では、大学受験予備校や教育関連会社において行われたテストの記述問題の解答に対し、理化学研究所において採点アノテーションを付与した理研記述問題採点データセット⁽⁵⁾を利用した。このデータセットでは、50字程度の解答テキストに加えて、採点根拠のアノテーション（採点基準と照らし得点の根拠になるテキストに1、それ以外に0をつけたもの）、全体得点、項目得点（採点基準毎の得点）から構成され、JSON形式で格納されている。また、参考情報として、問題文、解答・解説はあるが本文は含まれていない。問題は17問あり、1問につき500～2,000件のデータが含まれている。

3. フィードバックデザイン

図1に本研究で対象とする問題例と模範解答、解答例とその得点を示す。解答例を見てわかる通り、同じ得点でも解答の内容には多様性がある。そのため、全ての学習者にフィードバックとして模範解答を示すよりも、それぞれの学習者の解答の文脈を考慮したフィードバックが行えることが望ましい。

問 傍線部「こうした緊張したスタンスこそが饒舌な西洋文化を導いてきた」とあるが、それはどういうことか。句読点とも七〇字以内で説明せよ。

模範解答	他人を自分とは異質な考え方もつ人間と見なす西洋では、自分の意見に同意を得るために、言葉を尽くして他人を説得する技術が培われたということ。
解答 13点	A. 西洋人は、他人は自分とは別の考え方をしていると考えるので、相手に自分の考え方を説明し、同意してもらうために説得に努めるということ。
	B. 西洋は他人を異人と見なし、自分の考えを相手に理解させ受け入れさせようと言葉で対決する姿勢が説得という西洋文芸の本質を生み出したということ。
	C. 西洋は民族や言語や文化の異なる国が多く雑居しているため他人の考え方と自分の考え方が違うと思い、相手を説得して同意させようということ。
解答 8点	D. 西洋人は基本的に他人は異人と見なすため、自分の考えに他人を同意させる必要があるから西洋文化の基底には「対決」のスタンスがあるということ。
	E. 相手は自分とは別の考え方をしている「異人」なのだから、自分の考えを説明する必要があるという考え方が西洋文化の基底になっているということ。
	F. 様々な民族など、文化の違う国の人々に自らの考えを伝えるにはきちんと説得をする必要があり、言葉を必須としてきた為、意思が強く表れる文化になった。

図1 問題例、模範解答、解答例⁽⁵⁾

本研究では上記のような観点から、以下のアプローチでフィードバックを生成する。

- (1) 対象解答文から自動採点手法により得点を予測
- (2) (予測得点が低い場合) データセット中の高得点群の解答の中で意味的類似度が最も高い類似正解文を選択
- (3) 類似正解文に含まれて、対象解答文に含まれな

い差分フィードバックを生成

4. フィードバックシステム

フィードバックシステムは図2に示す通り、複数のモジュールを組み合わせて実現される。各モジュールの概要は以下の通りである。

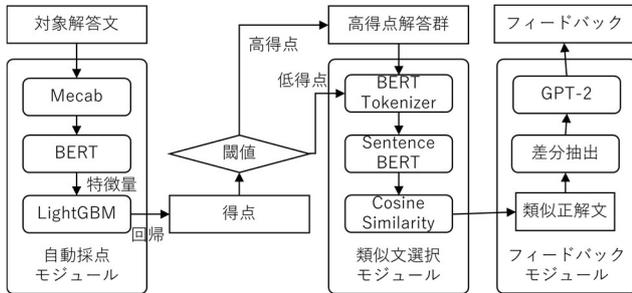


図2 フィードバックシステムの概要

4.1 自動採点モジュール

入力された対象解答文に対して、Mecab⁽⁶⁾による形態素解析、記号や句読点の削除などの前処理を行った後、日本語事前学習済みのBERTモデル⁽⁷⁾を利用して文脈的特徴量を抽出する。データセットの解答文のBERT特徴量と得点からLightGBM⁽⁸⁾による回帰予測モデルを構築し、このモデルを適用することで自動採点を行う。

4.2 類似正解文選択モジュール

対象解答文の予測得点が低い場合、対象解答をBertJapaneseTokenizerによりSentence BERT Japaneseモデル⁽⁹⁾の形式に変換する。その後、対象解答文とデータセット内の高得点群解答の間のコサイン類似度を計算し、最も類似度の高い高得点群解答を類似正解文として選択する。

表1に図1で示した解答を利用した類似正解文選択のシミュレーション結果を示す。低得点群(D, E, F)に対する模範解答と高得点群(A, B, C)のコサイン類似度を比較すると、(大きな差ではないが)模範解答は最も類似度が低く、最も類似度の高い類似正解文もそれぞれ異なっていることがわかる。

表1 類似正解文選択のシミュレーション結果

	模範	A	B	C
D	0.9121	0.9432	0.9479	0.9296
E	0.9073	0.9583	0.9398	0.9367
F	0.9050	0.9272	0.9318	0.9384

4.3 フィードバック生成モジュール

前節で選択された類似正解文は対象解答文に近い表現に加えて、より高い得点を得るために必要な内容が含まれている。現在開発中であるが、差分要素を日本語で事前学習したGPT-2⁽¹⁰⁾を利用して、表2に示すような形で学習者にフィードバックする予定である。

表2 フィードバック生成例

D	「相手への説得」の観点
E	「相手の同意と説得」の観点
F	「相手を同意させる」の観点

5. おわりに

本研究では、ドメイン知識(本文やフィードバック情報)を持たない短答記述式データセットを対象として、学習者の解答内容を考慮した改善フィードバックを生成するシステムを開発した。現時点では各モジュールに対するFine Tuningを行っていないため、それらを実施して個別モジュールについての評価を行う予定である。なお、フィードバックの妥当性に関する客観的な指標の検討も課題の一つである。また、低得点群の対象解答文ではフィードバックのポイントが複数になる可能性が高いため、重要度を設定して段階的なフィードバックを生成できる手法についても検討したい。

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。

参考文献

- (1) T. Mizumoto et al., “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, pp. 316–325, (2019).
- (2) S. Takano and O. Ichikawa, “Automatic scoring of short answers using justification cues estimated by BERT,” in Proceedings of the 17th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2022), pp. 8–13, (2022).
- (3) 内田優斗, 宇都雅輝: “受験者の能力を考慮した深層学習ベース短答記述式問題自動採点手法,” 教育システム情報学会誌 Vol. 38, No. 3, pp.218-228, (2021).
- (4) G. Deeva, et al., “A review of automated feedback systems for learners: Classification framework, challenges and opportunities,” Comput Educ, vol. 162, 104094, (2021)
- (5) 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>
- (6) <https://github.com/SamuraiT/mecab-python3> (2023.6.7 アクセス)
- (7) <https://github.com/cl-tohoku/bert-japanese> (2023.6.7 アクセス).
- (8) <https://github.com/microsoft/LightGBM> (2023.6.7 アクセス)
- (9) <https://github.com/colorfulcoop/sbert-ja> (2023.6.7 アクセス)
- (10) <https://github.com/rinnakk/japanese-pretrained-models> (2023.6.7 アクセス)