

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

内田 優斗^{*,**}, 宇都 雅輝^{*}

A Deep Neural Network-Based Automated Short Answer Grading that Considers Examinee Ability

Yuto UCHIDA^{*,**}, Masaki UTO^{*}

Recently, automated short-answer grading (ASAG) methods based on deep neural networks (DNN) have attained high scoring accuracy. However, the accuracy requires further improvement especially for large-scale and high-stakes tests because a slight scoring error will strongly influence many examinees. To improve the accuracy, this study proposes a new DNN-based ASAG method that utilizes examinees' abilities which are estimated using an item response theory model from their true-false responses for objective exam questions offering with a target short-answer question.

キーワード：自動採点，深層学習，短答記述式問題，項目反応理論，テスト理論

1. はじめに

近年，国内外のさまざまな試験において短答記述式問題の利用が注目されている。例えば，TOEFL や Graduate Management Admission Test (GMAT)，Medical College Admission Test (MCAT) などの世界のさまざまな大規模試験では短答記述式問題がすでに活用されており，日本でも大学入試センター試験に代わる大学入学共通テストで導入が検討されてきた。一方で，受験者数が数千～数万となるような大規模試験に短答記述式問題を導入する場合，採点に要する時間的・経済的コストの高さが問題となる。このような問題を解決する手法の一つとして自動採点技術が注目されている⁽¹⁾。

短答記述式問題の自動採点手法は 2000 年頃から活発に研究がなされており，伝統的には，専門家が設計した特徴量（文字数や文法エラー率など）を解答文か

ら抽出し，回帰モデルや分類モデルなどを用いて得点を予測するアプローチが一般的であった⁽²⁾。一方で近年では，深層学習を用いた自動採点モデルが注目を集めている。深層学習自動採点モデルでは，特徴量の事前設計や取捨選択を行うことなく，高精度な自動採点を実現できる。深層学習自動採点モデルとしては，畳み込みニューラルネットワークを用いたモデル⁽³⁾や Long Short Term Memory (LSTM) を用いたモデル^{(4)~(6)}，自己注意機構で構成される Transformer を用いたモデル^{(7)~(8)}など，多様なモデルが提案されている。深層学習自動採点モデルは，多段階得点の予測において 70%以上の正答率を^{(5)~(6)~(9)~(10)}，正誤の 2 値分類においては 90%以上の正答率を達成する^{(3)~(4)~(7)~(8)}といった事例が報告されており，高い精度を達成していることがわかる。

このように深層学習自動採点モデルは高精度な得点予測を実現しているが，特に大学入試や資格試験など

* 電気通信大学 (The University of Electro-Communications)

** 株式会社ベネッセコーポレーション (Benesse Corporation)

受付日：2020 年 6 月 22 日；再受付日：2020 年 10 月 28 日；採録日：2020 年 12 月 24 日