

記述式解答の機械認識による採点支援と自動採点

中川正樹^{†1}

概要: 本稿では、学習者・受験者の考える力を問うためには記述式問題が不可欠であるとの認識に立ち、筆記答案採点の信頼性を高め、労力と時間を軽減できる採点支援、採点の正しさを保障あるいは確認できる採点確認機能を前提とした自動採点を提案する。採点支援・自動採点は、学習者・受験者には記憶の新しいうちに採点を返して復習効果高められる。また、学習者の自学自習にも有効である。

キーワード: 記述式問題, 採点支援, 自動採点, 手書き認識

Computer-assisted and automated marking of descriptive answers by machine recognition

MASAKI NAKAGAWA^{†1}

Abstract: Assuming questions requiring descriptive answers rather than selecting answers are indispensable to test “ability to think” by learners or examinees, we propose to develop computer-assisted marking of handwritten answers to make the marking reliable and decrease the marking burden and time as well as automatic marking combined with function to confirm correct marking. Computer-assisted and automated marking will also enable learners or examinees to receive marked answers immediately after exams and review their answers when their memories are fresh. They will be also useful for self-learning.

Keywords: Descriptive questions, Computer-assisted marking, Computer-automated marking, Handwriting recognition

1. はじめに

新学習指導要領では「考える力」が重視され、それを問うために、記述式問題の必要性への認識が高まっている。そして、大学入学希望者学力評価テスト（仮称。以下、大学新入試）で記述式問題が導入される。大学新入試での記述式問題の採用は初等中等教育に大きな波及効果を生む。しかし、記述式解答の採点には多大な労力と時間を要する。そして、人による採点にもミスがある。また、別の課題として、受験者の進路が決まっただいぶ後になって、入試の採点ミスが判明する不祥事が後を絶たない。

大学入試改革に関する中央教育審議会答申(中教審第177号)が示されて以後、「新大学入試コンピュータ画面で」読売新聞(H27.7/14)、「大学新入試24年度から長文記述採点にコンピュータも」朝日新聞(H27.6/14)など、情報技術の導入が期待されているのは課題認識の表れでもある。文部科学省も、情報技術の利用可能性に言及している[1]。

本稿では、本研究会での石岡の招待講演とそれをもとにした解説[2]を踏まえ、中川ら[3]にそって講演する。まず、学習者の学習成果や能力を測る観測系として学力試験を捉え、記述式試験の必要性和重要性を確認する。次に、記述式試験の採点支援や自動採点によって得られる利点とそれに至る課題を整理する。そして、これらの課題を克服するための戦略を提案する。

2. 抜粋スライド

背景

- **社会背景**
 - CBT, WBTの普及、しかし、考える力や深い理解を問うためには記述式問題が不可欠
 - 大学新入試にも記述式問題 ⇒ 模擬試験などに波及
 - タブレット、電子ペーパー、依然として紙（2020年の大学新入試）
 - 記述式での採点コスト・時間をどうするか？特に紙の試験で
 - 採点の返却が遅れると復習効果が低下
 - 人による採点ミス、バラつき→最悪、進路が決まったあとで発覚
- **技術背景**
 - 深層学習の台頭。画像にはCNN、時系列には(B)LSTM
 - 手書き認識、特に画像認識は難しい。深層学習でも誤認識半分が現状。
 - しかし、手法の併用は誤認識、誤採点を減らせる
 - 手書き認識技術も、評価関数の高次化で着実に進歩
 - キーボード入力の英文エッセーには自動採点が実用化
 - 文字パターンデータベース：UNIPEN (韓国), NIST (米国), CASIA (中国), IAMonDO (スイス), ETL (日本), HP-JEITA (日本), TUAT.

試験・採点の関数モデル

生徒の理解 や考える力 x	試験 g	答案 $g(x)$	採点 f	採点結果 $f(g(x))$				
 小数点のある 掛け算ができる	記述式 考える力 深い理解 を問える	3.6 $\times 2.7$ 2.542 7.2 9.72	手採点 時間 大 採点ミス 有	3.6 $\times 2.7$ 2.542 7.2 9.72				
	選択式 観測系 として 不十分	問題ア ① 7.3 ② 9.72 ③ 97.2	答案 (マーク) ア ① ② ●	機械照合 時間 高速 誤り 僅少	<table border="1" style="font-size: small;"> <tr><td>答案</td><td>判定</td></tr> <tr><td>ア</td><td>×</td></tr> </table>	答案	判定	ア
答案	判定							
ア	×							

副作用

^{†1} 東京農工大学
Tokyo Univ. of Agri. & Tech.

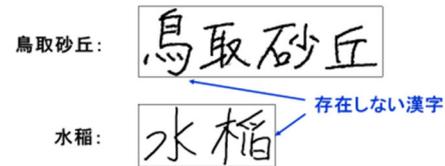
採点支援

- 方式
 - ・複数採点の一つに機械採点（先行，並行，後行），その後人によるチェック
 - ・一定比率を人手採点⇒分類器を学習，その後自動化
 - ・機械認識でクラスタリング⇒その後，人手採点（大学新入試で検討）
- 文科省本年8月31日報道発表「高大接続改革の進捗状況について」
 - ・採点の課題⇒AIの活用も視野に。
 - ・日本語文章のクラスタリングによる事前グループ分け@大学入試センター⇒採点のばらつきが減少，採点時間も70%に削減
 - ・数式，英語のフィージビリティを2017年度中の試行試験で示す必要性
数式：パターン，分割，記号認識からの特徴抽出とクラスタリング
英語：語彙駆動で認識，語彙自由で採点注視点を付加
採点の質が高められる可能性
- 教育サービス業界（学習塾，通信教育，教育出版社など）
 - ・学習履歴として利用，個別指導
 - ・ベテランの採点者の減少，ノウハウの再利用，差別化，…

TAT

偽陽性誤採点の分析

- 文字列認識で誤答を正答に判定してしまった1.8%はすべて存在しない漢字を強制的に認識，誤字や違う字はすべて検出



- ◆ これらは，人間の採点でもほとんど正答
- ◆ これら以外では，誤答は誤答と判定
- 採点基準に応じて，これを許したり，あるいは，純粋文字認識のスコアから採点注視点として指摘

TAT

自動採点

- 方式
 - ・機械採点⇒棄却と採点照会を人が採点
 - ・採点照会機能が前提
 - ・正答を誤答に採点する誤り（偽陽性誤採点），誤答を正答にする誤り（偽陰性誤採点）。後者は照会されにくい。
 - ・誤採点，特に偽陽性誤採点の低減を目的関数とし，誤採点が最小化する手書き認識・採点方式を機械学習によって求める。
 - ・質の認識方式の併用により誤採点を低減，一定比率の棄却が容認
 - ・照会・質問への応答は普通に採点するより時間がかかる。本当に個別に回答すべきこと以外は，共通する質問に一括回答などで効率化
- 教育サービス業界
 - ・キーボード入力のCBT/WBTだけでは難しい
 - ・記述式での採点コスト・時間を削減したい
 - ・早く返却したい

TAT

採点支援・自動採点の段階

活用段階	認識確認採点	採点支援	自動採点
概要	学習者がコンピュータの認識結果を確認・訂正	採点者による採点をコンピュータで支援	コンピュータによる採点
媒体	タブレット	タブレット&紙	タブレット&紙
利点	その場採点と確認即時フィードバック	採点ミスの防止採点注視点の指摘質の保証	即時フィードバック採点労力コスト軽減
技術的課題	正認識率と正採点率の向上，ユーザインタフェース	クラスタリング正採点率の向上誤採点率の低減	誤採点率の低減誤採点照会や質問への効率的な回答
効果的な利用分野	書写学習，算数・数学採点など	ほぼ全分野	採点照会機能と連携すればほぼ全分野*

* 長文の論述問題と図形の証明問題，グラフや化学式を書く問題などは当面除く。

TAT

予備実験

- 文脈・語彙の利用方法の違い（online）
- 純粋認識：MRFのオンライン認識，MQDFオフライン認識の統合スコアを出力（字形だけで判断するので，日本語の文字なのに，ギリシャ文字や数字などに誤認識することも）。
 - 文脈認識：3-gramを利用して文脈付きスコアを出力（誤認識を防げるが，それでも語句としていないものを出力したりする。しかし，辞書にない語句でも認識）。
 - 語彙認識：認識を登録語彙に限定（認識性能が一番高いが，正しくない手書きでも，それに近い語句に強制認識）。
 - 探索（純粋/文脈）認識：正解が分かっていることを利用して，それが純粋認識あるいは文脈認識の候補に含まれていなければそのスコア，含まなければ所定の最小スコアを出力。

TAT

オンライン認識例



TAT

予備実験結果（棄却あり）

アルバイトで，中学生社会・理科に出現する90語句を，正しく，及び，わざと（被験者自身が考えて）1，2文字間違えて記入
全体で100%，誤答と正答が同頻度として計算

データ	正答の手書きパターン（900）		誤答の手書きパターン（900）	
	正答と判定	誤答と判定	正答と判定	誤答と判定
純粋認識	71.7%	28.3%	3.8%	96.2%
純粋認識 + 文脈認識 + 探索文脈認識	35.8%	14.2%	1.9%	48.1%
1位と2位の差が 小さいものを棄却 棄却率：37%	27.5%	2.5%	1.8%	31.2%
純粋認識	73.5%	26.5%	3.8%	96.2%
純粋認識 + 文脈認識 + 探索純粋認識	36.7%	13.3%	1.9%	48.1%
1位と2位の差が 小さいものを棄却 棄却率：31%	27.0%	2.9%	1.8%	37.3%

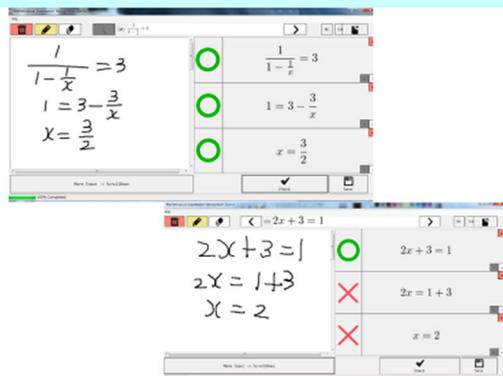
TAT

オフライン認識例



TAT

数式展開の正誤判定



TAT

オンライン試験の目標

採点支援

項目	3年後目標	5年後目標
採点時間短縮	80%	70%
採点ミスの低減	2/3	1/2

自動採点

項目	3年後目標	5年後目標
採点率	70%	85%
棄却率	30%	15%
誤採点率	3%	2%
採点照会・質問応答・個別指導	$(3+3)*3=18\%$ <small>採点時間</small>	$(2+2)*2.5=10\%$ <small>採点時間</small>
人による採点との比率	$30+18=48\%$	$15+10=25\%$ (約1/4)

TAT

Sep. 23, 2016 All right reserved by Nakagawa Lab.

10

オフライン試験の目標

採点支援

項目	3年後目標	5年後目標
採点時間短縮	80%	70%
採点ミスの低減	2/3	1/2

自動採点

項目	3年後目標	5年後目標
採点率	60%	70%
棄却率	40%	30%
誤採点率	5%	3%
採点照会・質問応答・個別指導	$(5+5)*3=30\%$ <small>採点時間</small>	$(3+3)*3=18\%$ <small>採点時間</small>
人による採点との比率	$40+30=70\%$	$30+18=48\%$ (約1/2)

TAT

3. おわりに

本講演は、記述式問題の採点支援や自動採点の可能性を論じた。記述式解答の試験的な採集、自動認識・自動採点コンテスト、実際の入試での解答パターン・技術・ノウハウの蓄積などを進められれば、10年後には効果が見えてこよう。試験とその採点の質、そして、効率の向上に向けて、真の情報化が進展することを願って止まない。

謝辞 本稿は、課題設定型産業技術開発費助成金の下、大学発ベンチャ iLabo との共同研究で実施している「タブレット上で筆記された回答の自動採点システム・サービス

の開発」の成果を基にしている。堀口昌伸社長、松下朋泳技術部長に感謝する。また、いつもご議論頂く信州大学の平井祐樹先生、大学入試センターの大久保智哉先生、そして、アイデアを形にしてくれる研究室のメンバーに深謝する。

参考文献

- [1] 文部科学省：報道発表「高大接続改革の進捗状況について」、(2016.8.31).
- [2] 石岡恒憲：コンピュータで実施する記述式試験－エッセイタイプ、短答式、マルチメディア利用について－，電子情報通信学会誌，**99**, (10), 1005-1011 (2016).
- [3] 中川正樹，平井祐樹：記述式解答の採点支援・自動採点に向けて－手書き認識からの挑戦－，情報処理学会誌，**57**, (9), 920-924 (2016).