

[特別講演] コンピュータで実施する記述式試験 ～ 人工知能活用の観点から ～

石岡 恒憲[†]

[†] 大学入試センター 研究開発部 〒153-8510 東京都目黒区駒場 2-19-23

E-mail: [†] tunenori@rd.dnc.ac.jp

あらまし 自由形式で解答を記述するいわゆる記述式試験の最新の状況について報告する。文章を書くタイプの記述式試験にはエッセイタイプのもので短答式タイプの2つがあるが、前者については代表的なシステムを紹介し、後者についてはまだ公的な試験に使われるほどには実用化が進んでいないことから、現在われわれが開発を進めている「人工知能を活用した採点支援システム」を紹介する。また、全米学力調査 (NAEP : National Assessment of Educational Progress) で実施された最新の記述式試験のコンピュータによる出題形式を動画を交えて報告する。

キーワード 自動採点システム, エッセイ (小論文) 採点, 短答式テスト採点, 記述試験, 全米学力調査(NAEP)

Computer-based Writing Tests

— From the point of view of artificial intelligence utilization —

Tsunenori ISHIOKA[†]

[†] Research Division, The Center for University Entrance Examinations. 2-19-23 Komaba, Meguro-ku, Tokyo, 153-8501
Japan

E-mail: [†] tunenori@rd.dnc.ac.jp

Abstract Writing tests include essay and short-answer tests. We present both of their procedures and scoring systems. In particular, we explain the latest technologies using recognizing textual entailment in short-answer scorers. A writing test format that made full use of multimedia materials for the National Assessment of Educational Progress in 2011 is also presented.

Keywords Automated Scoring System, Essay Scoring, Short-answer, Writing Test, National Assessment of Educational Progress

1. はじめに

文部科学省・高大接続システム改革会議の最終報告書[1]においてセンター試験にかわる新共通テストにおいて記述試験の導入と、その採点業務の効率・安定化のためにコンピュータの効果的な活用や人工知能の利用が盛り込まれた。記述式テストはいままで多肢選択テストに比べ、より正統あるいは真正(authentic)で信頼できると広く考えられており、採点のための技術的な課題が克服されてさえいれば、これを使う潜在的な需要は(センター試験に限らず)計りしれない。

しかし記述式テストの採点を巡る課題、たとえばその体制が取れるかや事務の負担増の問題などは、入試に関わる者によって多くの懸念となっている。そもそも記述式テストには大別して、「正解のないエッセイタイプ」のもので、「正解のある(したがって正解との意味の合致を判定する)短答式タイプ」のもの2つがあるが、いま新共通テストに求められている記述式は後者である。採点における技術的な困難から、質問の

解答は数十字から多くとも60字程度であると予想されている。

エッセイタイプについてはこれまでも、特にアメリカにおいて多くの自動採点システムが開発され実用に供されてきた[2]。我が国においても、我々のグループが日本語で初めての小論文自動採点システム Jess[3]を開発し、実用化の段階にある。

一方、短答式テストの自動採点については、その重要性は認められているものの技術的にさまざまな課題が未解決のままである。我が国においても国立情報学研究所の「ロボットは東大に入れるか」プロジェクト[4]で、受験ロボットがセンター試験の選択肢の正誤を、教科書からの知識源との含意関係認識技術を用いて解くということを試みてきた。その結果わかってきたことは、完全な含意関係認識技術すなわち正しい意味理解は現時点では困難[5]であり、いくつかの手法を組み合わせて半ばアドホックに解くことで、あたかも人間が解答したかのようにみせかけることがせいぜいであるという事実である。我々もいままでの科研「H23～

25 ; 短答式記述テストにおける実用的な自動採点システムの開発」[6]で自動採点に向けた要素技術の習得やシステム試作に努めてきたが、システムの限界もわかってきた。

このため著者らは、短答式記述テストをコンピュータで自動採点し、その自動採点に至った根拠を示すことで人間（採点者）が自動採点を修正することのできる採点支援システムを試作した[7]。

本発表では、最初にエッセイタイプの試験を採点する自動採点システムについて、その評価観点と評価手法について紹介する。つぎに、短答式記述の試験について、現在、われわれが開発を進めている採点支援システムを紹介する。人工知能の活用の名に値するシステムといっても過言ではないだろう。最後に、全米学力調査(National Assessment of Educational Progress, NAEP)[8]で実施された最新の記述試験のコンピュータによる出題形式について報告する。ただしNAEPの採点は自動ではなく、現在は人が行なっている。

2. エッセイ（小論文）採点システム

2012年にヒューレット財団がスポンサーとなりAutomated Student Assessment Prize(ASAP)と呼ばれるKaggle（企業や研究者による最適モデルのコンペ）が実施された。201の応募者が参加し、8つの論題に対して自動採点システム(Automated Essay scoring System: AES)の性能を競った。以下の9つのAESベンダーは招待された。順にシステム名と開発元を示す。

- ① AutoScore : American Institutes for Research
- ② LightSIDE : カーネギーメロン大学 TELE-DIA 研究室
- ③ Bookette : CTB McGraw-Hill
- ④ E-rater : Educational Testing Service (ETS)
- ⑤ Lexile Writing Analyzer : MetaMetrics
- ⑥ Project Essay Grade (PEG) : Measurement Inc.
- ⑦ Intelligent Essay Assessor(IEA) : Pearson Educational Technologies
- ⑧ CRASE : Pacific Metrics
- ⑨ IntelliMetric : Vantage Learning

うち1つ(LightSIDE)はオープンソースであり商用ベンダーは8つであるが、この8つで現在のアメリカにおける自動採点市場の97%を占めるとされている。この9つのAES調査の結果[9]はAESが人間の評価者に比べ信頼できるというものであるが、それに対しては多くの反論(例えば[10])がある。これらについても解説する。

3. 短答式記述採点システム

我々の研究グループは比較的解答の短い問題で、かつ高度な判断による寄与の少ない理科や社会の問題に対して、たとえば30字程度内外の解答に対する自動採

点、および自動採点システムの試作を完成させた。このシステムの最大の特徴は完全な自動採点システムとせず人間（採点者）が修正しうる余地を残すところにある。現在の技術レベルでは正確な意味理解はまだ難しいことによる。より具体的には以下の手順で採点を行う。

(1) 設問ごとに作題者が用意した「採点基準」に従い、用意された解答がその採点基準の各項目に合致しているか否かを表層的な観点からシステムが自動判定する。

(2) その自動判定に基づいた仮採点（チェック得点と称する）を与えるほか、システムはすでに専門家が採点した採点データ（教師データ）をもとに機械学習器で与えられる予測得点（システム得点と称する）の両方を与える。

(3) 人間の採点者は、これらの情報を参考に、システムの提示する得点を認証することができるほか、自らの判断でこれを修正・上書きできる。

完成システムのスクリーンショットを図1に示す。採点者が自動採点の結果に同意するならば、採点者はデフォルトで表示される採点結果を承認するだけで採点をすすめることができる。

なお、システムの開発に当たり、学研グループより記述式問題（2015年度「学研全国総合模試」）、解答、採点結果を提供いただいた。ここに記して感謝申し上げる。

4. 全米学力調査における記述テスト

2011年の全米学力調査(National Assessment of Educational Progress, NAEP)[8]では、8年生と12年生に対して、作文テストが従来の紙筆テストに換わり、初めてコンピュータによって実施された。そのテストは、単に従来の紙と鉛筆をコンピュータに置き換えただけのものではない。現在のデジタル技術が十分に活用できるよう、取り扱う作文のタイプには、テキストによる問いかけのほか、写真を含むもの、音声によるもの、ビデオを見て問いかけに答えるものの4つがある。加えて、多くの学生が受験できるようユニバーサルデザインについての配慮がなされている。たとえば、問題文についての音声読み上げやフォントサイズの変更が可能である。また電子上のスペルチェックが利用できる。

発表では、この初めてのコンピュータ化された作文試験について、その仕様を動画を交え示すとともに、今後の方向性や果たすべき課題について論考する。

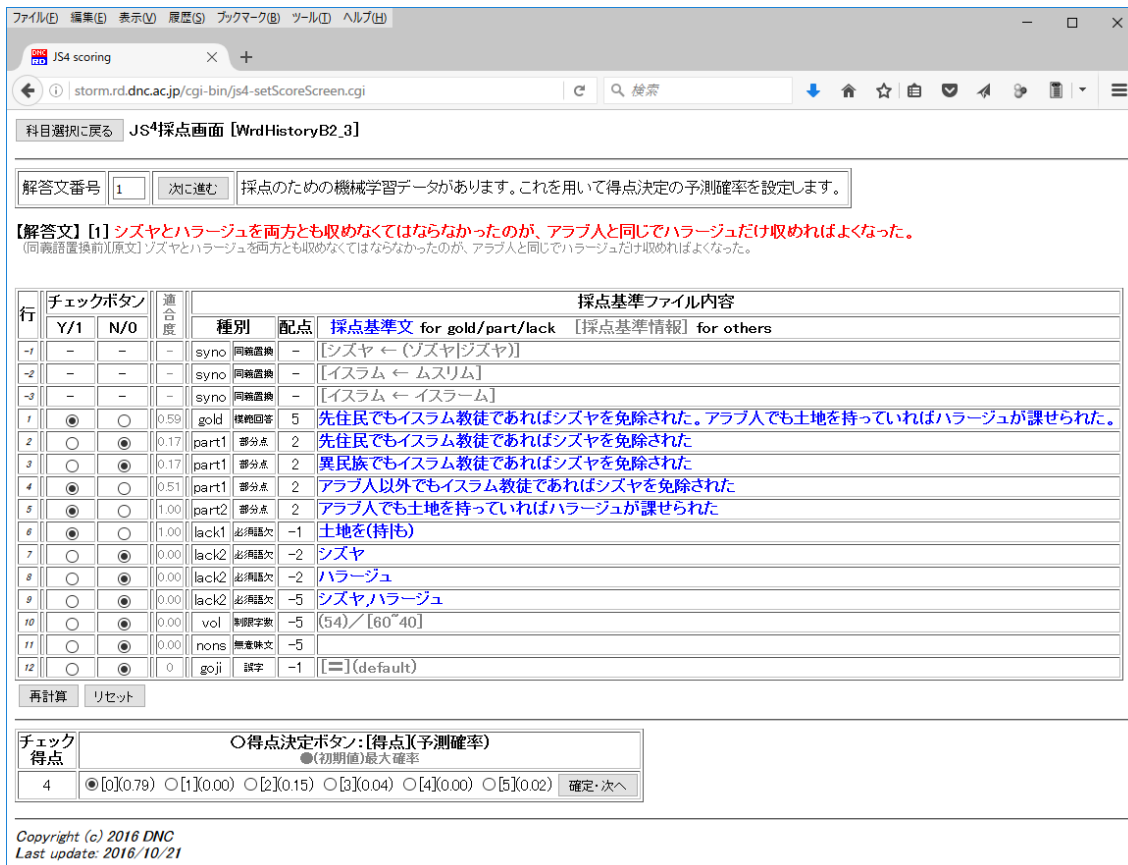


図 1 : 短答式記述テストにおける採点支援システム画面 [7]

文 献

- [1] 文部科学省, 高大接続システム改革会議「最終報告」の公表について, Mar. 2016.
http://www.mext.go.jp/b_menu/shingi/chousa/shouga/i/033/toushin/1369233.htm
- [2] 石岡恒憲, “コンピュータ上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, Vol. 99, No.10, pp.1005-1011, 2016.
- [3] T. Ishioka and M. Kameda, “Automated Japanese essay scoring system based on articles written by experts,” Coling-ACL 2006, no. P06-1030, pp.233-240, 2006.
- [4] ロボットは東大に入れるか Todai Robot Project, <http://21robot.org>
- [5] 国立情報学研究所, “問われるのは意味を理解する力. 暗記だけでは解けない社会科科目, [特集] 人工頭脳プロジェクト「ロボットは東大に入れるか」,” NII Today, no. 60, pp. 8-9, 2013,
- [6] 短答式記述テストにおける実用的な自動採点システムの開発, 研究課題:23650558, 科学研究費助成事業データベース
<https://kaken.nii.ac.jp/grant/KAKENHI-PROJECT-23650558/>
- [7] 石岡恒憲, 亀田雅之, 劉東岳, “人工知能を利用した短答式記述採点支援システムの開発” 信学技報, 言語理解とコミュニケーション研究会 (NLC), 2016年12月22日 (刊行予定).
- [8] National Assessment of Educational Progress, NAEP (2011). <http://nces.ed.gov/nationsreportcard/>
- [9] M.D. Sharmis and B. Hamner, Contrasting State-of-Art Automated Scoring of Essays: Analysis, Contrasting Essay Scoring, 2012.
- [10] Les C. Perelman, “Critique of Mark D. Sharmis & Ben Hammer: Contrasting State-of-Art Automated Scoring of Essays: Analysis,” The Journal of Writing Assessment, Vol. 6, Issue 1, Aug. 2013.