

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

石岡 恒憲[†] 亀田 雅之[‡] 劉 東岳[¶]

[†] 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

[‡] 大学入試センター 研究開発部付 〒153-8510 同上

[¶] 学研ホールディングス 学研教育総合研究所 〒141-8510 東京都品川区西五反田 2-11-8

E-mail: [†] tunenori@rd.dnc.ac.jp, [¶] DL.Liu@mmf.gakken.co.jp

あらまし センター試験など大学入試試験レベルの短答式記述試験の自動採点および人間による採点を支援する実用可能なシステムを目指して、その試作および実装をした。自然言語におけるテキスト間の同義や含意、推論を公的な試験に採用するほどの精度はこの数年では不可能であるという判断のもとに、採点は設問ごとに作題者が用意した「採点基準」に従いシステムがある程度の精度をもった採点計算（自動採点）を基本とし、その結果を人間が確認・修正できるものとする。システムは「(予め用意された) 模範解答」と「(被験者の実際の) 記述解答」との(ある程度の) 意味的同一性や含意性を判定するほか、プロンプトと呼ばれる素材文と解答文との意味的近似性なども考慮する。また採点結果は多値分類であることから、自然言語処理でしばしば用いられているサポートベクターマシンではなく、多くの予測変数において有効に機能するランダムフォレストによる機械学習分類を使う。試作システムは Linux 上での Web システムとして動作する。大学入試の模試として実施された社会科の記述式問題 8 題に対して、3-6 点の配点の場合にシステムの提示する自動採点と人間の採点を比較すると、高度な意味的判断の要さない大半の設問においては、おおよそ 7 割から 9 割が 1 点差に収まっている。

キーワード 短答式記述テスト, 自動採点, 機械学習, ランダムフォレスト, 含意関係認識, 文科省, 高大接続システム改革会議

AI-based Japanese Short-answer Scoring and Support System

Tsunenori ISHIOKA[†], Masayuki KAMEDA[‡] and Dongyue Louis LIU[¶]

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

[‡] c/o Research Division, The Center for University Entrance Examinations. Same as the above

[¶] Gakken Research Institute for Learning and Education, 2-11-8 Nishi-Gotanda, Shinagawa-ku, Tokyo, 141-8510 Japan

E-mail: [†] tunenori@rd.dnc.ac.jp, [¶] DL.Liu@mmf.gakken.co.jp

Abstract We have developed an automated Japanese short-answer scoring and support machine for new National Center Test written exams. Our approach is based on the fact that recognizing textual entailment and/or synonymy has been almost impossible for scoring systems for several years. The system generates automated scores on the basis of evaluation criteria or rubrics, and human raters revise the scores. The system determines semantic similarity between the model answers and the actual written answers as well as a certain degree of semantic identity and implication. Owing to the need for the scoring results to be classified at multiple levels, we use random forests to effectively utilize many predictors rather than use support vector machines. An experimental prototype operates as a web system on a Linux computer. We compared human scores with the automated scores for the case in which 3-6 allotment points were placed in 8 categories of the social studies test as a trial examination. The differences between the scores were within one point for 70-90 percent of the data when high semantic judgment was not needed.

Keywords Writing test, Automated scoring, Machine learning, Random forests, Recognizing textual entailment, Ministry of education, university and high-school articulation reform system conference

1. はじめに

文部科学省・高大接続部会の答申[1]において、新テ

ストに記述式テストが導入されることが決定し、高大接続システム改革会議の最終報告書において採点業務

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。記述式テストには大別して、「正解のないエッセイタイプ」のものと、「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」のものの2つがあるが、いま新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。短答式テストはいままでの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)確かにある。

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システムが開発され実用に供されてきた(サーベイ論文については[2]など)。我が国においても、著者ら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。

しかし今ここで開発しようとしている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。Vigilante[4]は世界最大のテスト機関である ETS とニューヨーク大学とで、この分野における共同研究を行い、最初の報告をした。Leacock & Chodorow [5]は、ETS が開発した c-rater の最新の仕様について報告している。Pulman & Sukkarieh [6]は、情報抽出技術に隠れマルコフモデルなどの自然言語処理を用いて、システムが用意する正解文と同じ意味の文を幾つか自動生成する試みについて述べている。しかしその性能は、エッセイの自動評価採点システム(e-rater)に比べ、人間による専門家との一致率は10%以上も小さい。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[7]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[8]であり、いくつかの手法を組み合わせれば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。著者らもいままでの科研「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」[9]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着想した。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。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のは、現在の技術レベルでは正確な意味理解はまだ難しいとの判断による。もっともセンター試験の後継となる新テストの記述には比較的自動採点が容易と思われた(教科書に書かれている事実との意味的照合で対応可能な)理科や社会科の問題ではなく、国語に記述問題が採用された。国語の問題では構文的に使われている語彙にもほとんど違いはないにもかかわらず、比較する2文の意味的な違いを検出しなければならず難しい。このため現時点では比較的取り組みやすい解答の短い問題で、かつ高度な判断による寄与の少ない社会科の問題に対してアプローチを試みる。

2節では大学入試のための模試で用いられた記述問題の内容とその回答例(模範解答)を示す。3節では採点支援システムの仕様を示す。4節には社会科の問題8題についてのシステム採点について定量的に評価する。5節はまとめである。

2. 模試で用いられた記述問題

今回、解析の対象とするのは、2015年度の「学研全国総合模試」で出題された地理・歴史4科目において出題された短答式記述解答部分(1科目あたり各2問、計8問)である。問われる「内容」と、模試受験者に配られた「解答と解説」にある「正解」を表1に記す。

表1: 記述問題の内容と正解例

科目	設問	内容と正解
世界史 B 〔②〕	① 問 8 (3点)	〔内容〕 古代ギリシャ: ソロンの財産政治 〔正解〕 財産権の多少に応じて、市民の参政権を定めた。(22字)
	③ 問 5 (5点)	〔内容〕 イスラーム: アッバース朝の租税制度 〔正解〕 征服地の先住民でもイスラム教徒であればジズヤを免除され、アラブ人でも征服地に土地を持っていればハラージュが課せられた。(60字)
日本史 B 〔①〕	② 問 5 (3点)	〔内容〕 日露戦争の講和条約(領土面の変更) 〔正解〕 ロシアからサハリンの南部が日本に割譲された。(22字)
	④ 問 5 (5点)	〔内容〕 第2次大戦後の財閥解体 〔正解〕 独占禁止法によって持ち株会社やカルテル・トラストなどを禁止し、過度経済力集中排除法によって巨大独占企業の分割が行われた。(59字)
日本史 B 〔②〕	① 問 11 (3点)	〔内容〕 元明天皇: 律令制度における税負担 〔正解〕 都での労役の代わりに、一定の布を納める。(20

字)

- ③ 問 11 (6点) **〔内容〕** 鎌倉時代後半：都市の商業活動 **〔正解〕** 問と呼ばれる運送業者や借上げと呼ばれる高利貸業者が現れ、金銭を輸送する代わりに手形で決済する為替が使われるようになった。(59字)

- 地理 B** ① 問 5 (3点) **〔内容〕** ナミブ砂漠の2つの成因 **〔正解〕** 沖合を流れる寒流と中緯度高压帯のため。(19字)

- ④ 問 6 (6点) **〔内容〕** 世界の人口：インドと中国の比較 **〔正解〕** 中国では一人っ子政策を実施したが、インドでは厳しい産児制限に失敗したため、中国よりも出生率が高くて人口増加が著しいから。(59字)

◎ 学研

世界史 B 〔②〕 ③問 5 の正解について、少し補足する。イスラーム世界においてジズヤは非ムスリムに信仰の代償として課せられる人頭税で、征服地の先住民がイスラームに改宗すれば本来は課税されないものであった。またハラージュは地租（土地税）でムスリムであろうとなかろうと、土地を所有しているものには等しく課税されるはずのものであった。それにもかかわらずウマイヤ朝の時代には、先住民がイスラーム教に改宗してもジズヤが課され、土地を所有していてもハラージュが免除されていた。アッバース朝になって本来の正しい状態になった（表 2）。本設問はその内容を問うものであるが、採点にはその正確な意味理解が必要である。

表 2：ウマイヤ朝→アッバース朝の租税制度

	ムスリム (イスラーム教徒)	非ムスリム
ジズヤ	課す→免除	課す
ハラージュ	課す	免除→課す

3. 採点支援システムの仕様

3.1. 概要

我々のアプローチは、

(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し、

(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、

(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・上書きできるものとする。た

だしシステムの提示値は、ある程度の精度（適合度）を有するものでないと、採点の手間を軽減するという点で有効とはいえず、80%以上の精度を一応の目標とする。機械学習に際しては、「採点基準」への表層的な一致に加え、「模範解答」と「(実際の)記述解答」との意味的近似性など、深層的な変量も考慮するものとする。

本システムの構成は図 1 の通りである。

- ① 動作前準備として人間の採点者による採点データを多数収集し、ランダムフォレスト (RF) による学習器にかける。学習には採点基準ファイルとの適合の程度も必要である。この学習結果をもとに新規の解答に対して採点結果を返すための採点エンジンを構築しておく。
- ② あらかじめ用意した採点基準ファイルからユーザ（採点者）が用いる HTML による採点画面を自動的に生成する。
- ③ ユーザ（採点者）は自らの端末から Web ブラウザを用いて②の採点画面を開く。すると cgi プログラムが動作する。そこには①の採点エンジンの結果であるところのシステムの推奨値が表示される。採点結果はファイル/データベースに格納される。ユーザはこの採点操作を必要な分だけ繰り返す。

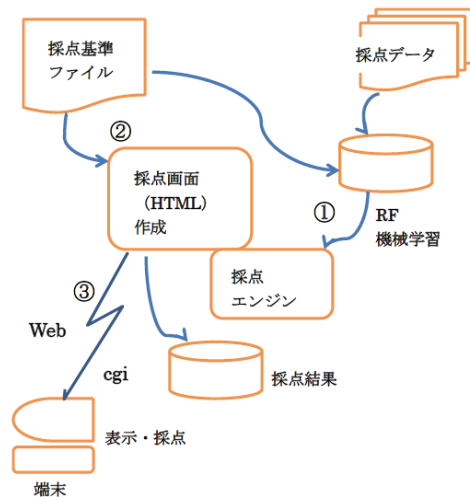


図 1：システム構成図

3.2. 採点画面

完成システムのスクリーンショットを図 2 に示す。システム上部に「採点すべき解答文」があり、採点基準には「(許される)同義語や異表記」、「模範解答」、「部分点解答」、「必須語」などがある。「模範解答」、「部分点解答」には「採点すべき解答文」のシステムによる適合度判断や、「必須語」の有無などが自動的に判断、マークされる。他にもシステムは、「制限字数」を超えたか否か、あるいは「意味のある文」になって

いるか否かを判断する。問題の「タイプ」によっては、理由を答えるべき設問に対してそのような回答形式になっているか、あるいは名詞や名詞句を問う設問に対してそのように答えられているかをシステムは自動判断する。この判断は Y/N の 2 値判断であるが、その判断の程度を「適合度」として [0-1] の連続値で示す。採点者はこの数値を確認の上、修正することができる。

画面下部の「チェック採点」は「チェックボタン」による Y/N の 2 値判断とそれに付与される。一方、「得点決定ボタン」では採点者がその決定点を選択・決定

する。その初期値は“過去の採点結果に基づく（いわゆる機械学習でいうところの教師学習による）”予測確率が最大になる得点に予めマークされており、それぞれの得点の確率も同時に示される。

教師学習データがない場合、すなわち該当する設問についての採点データが存在しないときは、その旨がメッセージ欄に表示され、予測確率も表示されず、初期値マークもされない。

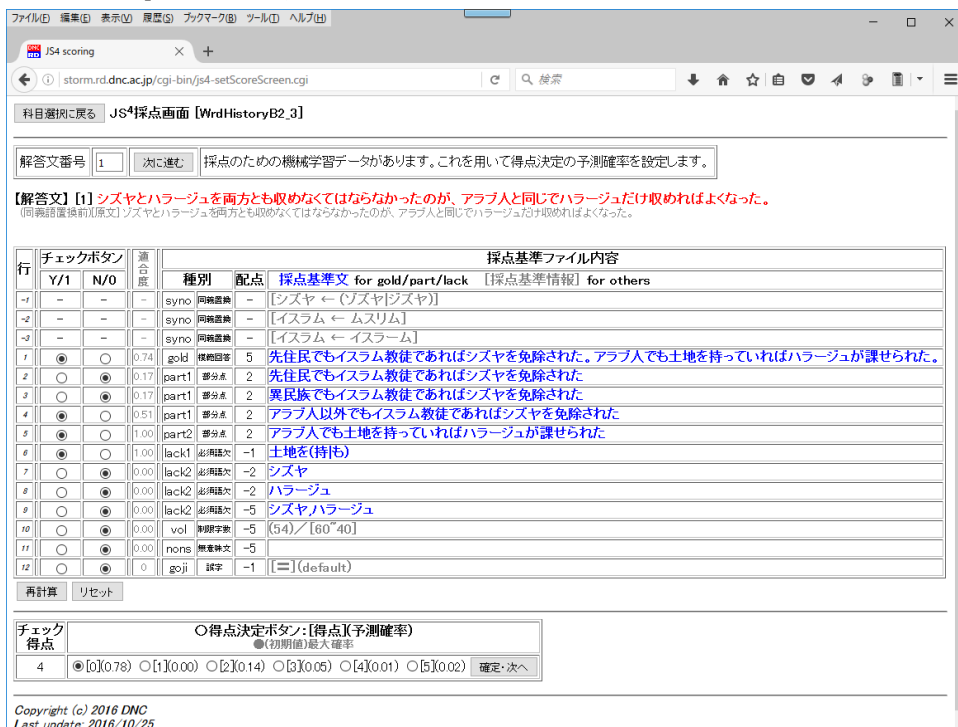


図 2：採点支援システム画面(世界史 B [②]③問 5 の場合)

3.3. チェック得点と機械予測得点との差異

図 2 に示されている例は世界史 B [②]③問 5 の事例である。ここに示されている解答「ジズヤとハラージュを両方とも納めなくてはならなかったのが、アラブ人と同じでハラージュだけ納めればよくなった」は、模範解答「先住民でもイスラム教徒であればジズヤを免除された。アラブ人でも土地を持っていればハラージュが課せられた」と比較し、一見、字面は共通性があるものの意味的には全く適合していない。このため、表層的な語句の一致に基づきチェックボタンの判定から得られる「チェック得点」は 4 点（配点 5 点）と高いが、表層的側面以外の要素を加味して機械学習を行った予測得点は 0 点で、これを推奨値（その予測確率 0.79）として提示している。機械学習の効果が適切に機能していることがわかる。

3.4. 採点基準ファイルから画面の自動作成

本システムは Web アプリケーションであり、図 2 に

示される画面は HTML (Hyper-Text Markup Language) と呼ばれる、表示文に適当な指定文を埋め込んだ（マークアップした）ファイルにより生成されている。我々はこの HTML ファイルを（コンピュータ初心者によってわかりやすく取り扱い易い）プレーンな採点基準ファイルから自動的に作成する仕組みを構築した。

以下に図 2 を作成するもとになったプレーンな採点基準ファイルを示す（図 3）。

syno	シズヤ	(ズズヤ ジズヤ)
syno	イスラム	ムスリム
syno	イスラム	イスラーム
gold	5	先住民でもイスラム教徒であればシズヤを免除された。アラブ人でも土地を持っていればハラージュが課せられた。
part1	2	先住民でもイスラム教徒であればシズヤを免除された
part1	2	異民族でもイスラム教徒であればシズヤを免除された

part1	2	アラブ人以外でもイスラム教徒であればシズヤを免除された
part2	2	アラブ人でも土地を持っていればハラージュが課せられた
lack1	-1	土地を(持)も
lack2	-2	シズヤ
lack2	-2	ハラージュ
lack2	-5	シズヤ,ハラージュ
vol	-5	60~40
nons	-5	
goji	-1	

図 3: 採点基準ファイル (ラベル名, 配点, 表記をタブ区切りで書く)

syno には同義語や異表記を, gold には模範解答を, part には部分点を記す. 部分点には加算が可能なものと, 最大値をとるものとの 2 種類がある. 同じ表記 (たとえば 3 個ある part1 など) では最大値をとり, 異なった表記(part1 と part2)では加算できる. lack は必須語で, これがないと減点される. これらの表記には正規表現が可能である. vol は字数制限を示す. nons は無意味文, goji は誤字による減点を示す.

gold で指定された「模範解答」文, 及び part で指定された「部分点」文と解答文との関連度を本システムでは「適合度」と呼ぶが, ここでは筆者(亀田)が重要キーワード/重要文/関連文の抽出のために提案した, キーワード候補間の関連度の手法(擬似キーワード相関法)[10]に基づき, 文内のキーワード候補群間で定義した文間の参照関連度と被参照関連度及びそれらの調和平均(F 値)[11]を活用した(参照関連度と被参照関連度の関係は, 各々再現率(recall)と正解率(precision)の関係と同じとみることができ). これらの適合度は 0~1 の間の値になるが, 適合度を参考値として示した上で, 該非(初期値)は適合度を四捨五入して, 1(該当)/0(非該当)として表示する.

4. 性能評価

4.1. クラス分類の評価

図 1 で示されるそれぞれの「適合度」に加え, 日本語 Wikipedia を言語集合とした場合に python で利用できる gensim ライブラリーを用いた「解答と模範解答」および「解答と問題文」との意味的空間の近さ(コサイン類似度)などを予測変数(predictor)とし, ランダムフォレスト V3.1[12,13] (以下 RF と略す) による機械学習で予測モデルを構築した. 多くの機械学習の手法の中で RF を用いるのは, この方法が(1) 多くの予測変数を用いた分類で有効に機能する (場合が多い) (2) 分類に有効な予測変数への寄与の程度を定量的に評価することができる (3) 得点は 0~3, あるいは 0~6 の多値分類が必要でありこれに適している, ことによる.

表 1 に示す 8 つの設問に対し, 予測モデルによる予

測値と正解値を比較した. 交差行列の詳細については紙面の都合で割愛するが, 一致しない確率(誤分類率)は決して良いとはいえない. ただ, 正解と予測の違いが 1 点差内で収まる確率を求めると表 3 のようになり, 正確な意味理解が必要な世界史 B [②]③を除けば 71% から 95% とかなり使えるレベルであることがわかる.

表 3: 予測が正解の 1 点差内に収まる確率

設問	確率	設問	確率
世界史 B [2] ①	0.75	日本史 B [2]④	0.86
世界史 B [2] ③	0.48	日本史 B [2]③	0.71
日本史 B [1] ②	0.76	地理 B ①	0.91
日本史 B [1] ④	0.88	地理 B ④	0.95

なお RF ではエラー率を計算するのに交差検証を行う必要がない. つまり分離したテスト集合を用意する必要がない[13]. RF を実行すれば自動的に得られるからである. RF ではオリジナルデータの 2/3 のブートストラップ標本で木を作り, 残りの 1/3 をテストデータとする. R による RandomForest パッケージでは, デフォルトで 500 回繰り返し, 最大投票(most votes)によって最終的なクラス分類を決定する. これによりテスト集合についてエラー率を決定する. 表 3 はこのブートストラップ繰り返しにより得られた数値である. 最終的に得られた RF モデルにオリジナルのデータを与えてその予測値と正解を比べるならば, 本実験では標本サイズに比べ予測変数が多いこともあり, ほぼ 100% に近い一致率が得られる.

4.2. クラス分類に寄与する変数

ランダムフォレストは判別(クラス分類)における変数の重要度をジニ係数の値で計算する. ジニ係数の値が大きいほど判別への寄与が大きい. 図 4 にジニ係数の大きい順に変数を並べ, その値を横軸に目盛ったグラフを示す. 紙面の都合で)世界史 B [2] ③の場合のみ示す. 寄与の大きな上位 3 つは以下の通りである.

QA_sim : 問題文と解答とのコサイン類似度

SA_sim_1 : 正解 1 と解答とのコサイン類似度

sa_jpkwrel_Fvl_gold_std01 : 正解 1 と解答とのキーワード一致における F 値

大文字で始まる変数名は日本語 Wikipedia で用いられている語彙で言語的な意味空間を構築し, その意味空間におけるコサイン類似度や正解率(precision), 再現率(recall), 正解率と再現率の調和平均である F 値 (より正確には F_1 尺度), およびそれらに配点を乗じたものである. 小文字で始まる変数は, 3.4 節で述べている比較的表層的な関連度に関する幾つかの指標である. 用いている変数の総数は(採点基準の大きさに依存するが)大よそ 40~70 程度である.

図 4 で示される指標を各設問で比較すると, 設問によって判別へ寄与する変数は一定ではなく多様であっ

た。その反面、模範解答（正解）と記述解答との意味空間における F 値やコサイン類似度はクラス分類に最も寄与する場合が多いことも確認された。

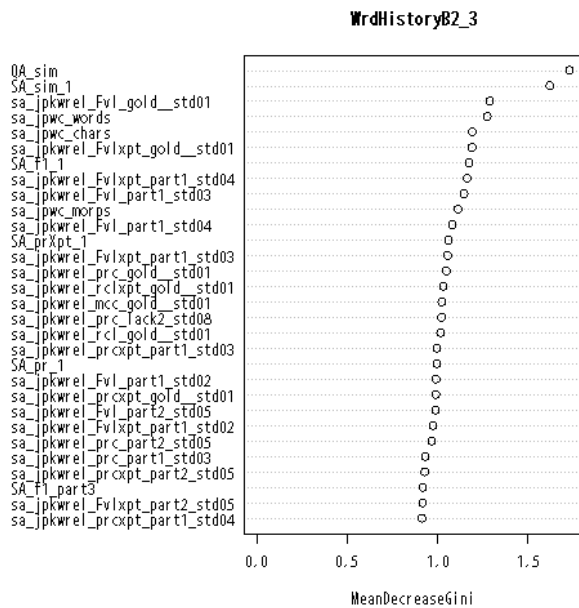


図 4：判別への寄与（世界史 B [2] ③の場合）

5. おわりに

模範解答（正解文）と記述解答文の意味的同一性チェックは、構文的・意味的レベルの複雑な照合が必要でまだまだ技術的に困難であるが、現在の双方のキーワード群同士の照合は、「表層的」な側面だけでなく「意味的」な側面も予測変量として加えており、その第一歩のアプローチとしては、十分に現実的だと判断できる。文章の正確な意味理解は現時点では難しく、最終判断を人間に委ねる形は妥当であろう。本事例では社会科（地理・歴史）の問題を事例に評価したが、国語の記述採点においても、正解に多くの異表記の（それでいて意味の同一な）正解文を複数用意することで、それなりに対応できると考えている。現在の我々のシステムは（採点基準ファイルの）同じラベルに対してはその適合度合いの最大のものを選択する仕組みになっているので、現状の仕様で対応可能である。また例えば「父の気持ち」といった一般には重要な単語でなくてもこれが内容的に重要なフレーズであるならば、これを「必須語」として採点基準ファイルに記録し欠落時に適当な配点を減らすことができる。「必須語」の登録機能は既実装済みであり、現在の仕様において採点基準ファイルの書き方を丁寧に行えば、ある程度の予測妥当性を保持できるのではないかと考える。

もっとも本実験では教師データと呼ばれる人間の採点データの数が十分な学習を行うほどには多くが確保できなかった。実際の記述採点では無回答や 0 点（が付与される）解答が多く、ある程度の得点のパラツキ

をもった適当な回答を数多く確保することは今後も必要であろう。

文科省・高大接続システム改革会議の最終報告に短答式記述試験の採用が提言されてことから、現在この短答式記述採点支援および自動採点にはわかに注目を浴びている。本システムがこの分野の研究を行う研究者や実務者の参考になることを期待する。

謝辞

本システムの開発に当たり、記述式問題、解答、採点結果の提供をいただき、貴重なコメントを賜った学研グループの皆様、特に学研プラスの松岡嗣直氏、西郷直子氏に心より感謝申し上げる。

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/shougai/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] Vigilante, R., “Online Computer Scoring of Constructed-Response Questions,” Journal of Information Technology Impact, 1(2), pp.57-62.1999.
- [5] Leacock, C. and Chodorow, M., C-rater: Automated Scoring of Short-Answer Questions, Computers and the Humanities, Springer Netherlands, ISSN 0010-4817 (Print) 1572-8412 (Online), 37(4), pp.389-405, 2003.
- [6] Pulman, S. G. and Sukkarieh, J. Z., “Automatic Short Answer Marking,” Proceedings of the 2nd Workshop on Building Educational Applications Using NLP, ACL 2005, pp.9-16, 2005.
- [7] ロボットは東大に入れるか Todai Robot Project, <http://21robot.org>
- [8] 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,
- [9] 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/>
- [10] 亀田雅之, “擬似キーワード相関法による重要キーワードと重要文の抽出,” 言語処理学会 第 2 回年次大会 発表論文集, pp.97-100. Mar.1996.
- [11] 亀田雅之, “段落間及び文間関連度を利用した段落シフト法に基づく重要文抽出,” 情報処理学会自然言語処理研究会 121-17, 情報学基礎 47-9 (共催), pp.119-126, Sep.1997.
- [12] Breiman L, Random Forests, Machine Learning, 45 (1), pp.5-32, 2001.
- [13] Breiman L. and Cutler, A. Random Forests, <http://www.stat.berkeley.edu/~breiman/RandomForests/> updated March 3, 2004.