

Dec. 5-8, 2018

NTCIR 13@NII, Japan

Rubric-based Automated Japanese Short-answer Scoring and Support System Applied to QALab-3

Only Evaluation-method Task



Tsunenori Ishioka (Japan)

Kohei Yamaguchi (Japan)

Tsunenori Mine (Japan)



Outline

- ❑ Status of Written Test
- ❑ Specification of our System
- ❑ The Performance Evaluation
- ❑ Summary



© edsurge.imgix.net

Writing test



© www.writeawriting.com

□ Essays

- No model answer; 300~ words
- Rhetoric, logics, contents
- Practical application: **e-rater**, **Intellimetric**, and **Jess** (for Japanese)

□ Short-answers

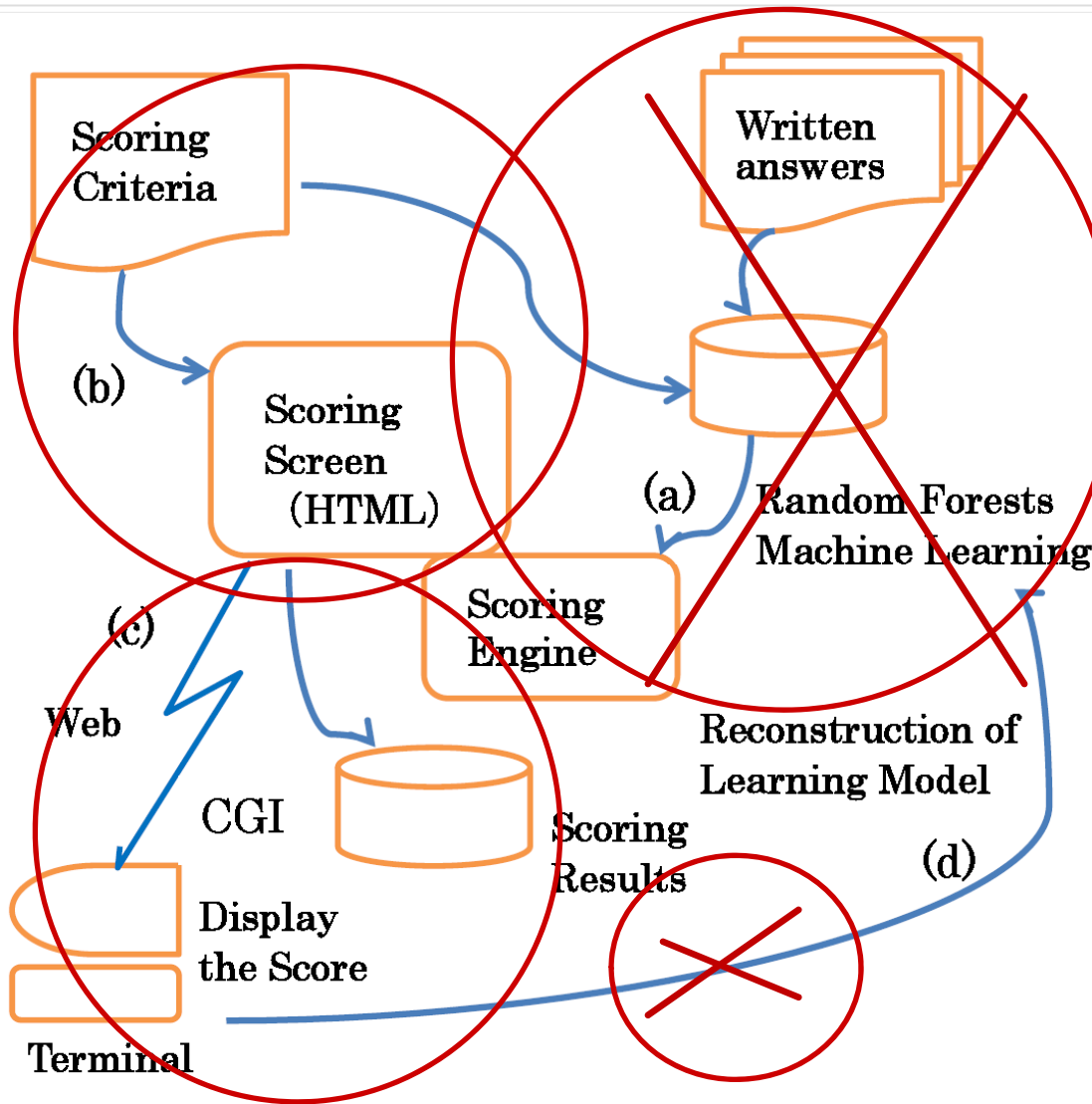
Our task

- **Model answer**(s); 1~2 statement(s)
- Semantic identity / Recognizing textual entailment
- Technical difficulty
- AES: Not applied for high-stakes test(s)

Support system for short written tests

- Accurate language understanding is almost impossible
 - Agreement to scoring rubric
 - Automated Scoring by rubric
 - ~~Classification scoring by ML~~
 - ~~Overwriting the score~~ ← Human rater
- } Machine
- Simple scoring rubric description
 - Automatic creation of scoring screen

System Components



- ~~1. Scoring by Professional human raters; DB; Scoring model by RF~~
2. Automatic creation of scoring screen
3. Scoring via Web; Results in DB
- ~~4. Reconstruction of new learning model~~

Scoring criterion file

tab

tab

syno ムハンマド・アリー ムハンマド=アリー synonym

gold 20 古代エジプトは、ナイル川を中心に長らく独立王朝が栄えたが、アレクサンドロスなどの征服を受ける。プトレマイオス朝エジプト

Nugget sentence

model answer

part1 2 古代エジプトは、ナイル川を中心に、古王国から新王国まで、長らく独立王朝が栄えた。 partial points

part2 2 古代エジプトは、アレクサンドロスの征服を受けた。 Max. Vote: 2 = {2,2,3} ス大

part88 3 ナセルは、スエズ運河の国有化、アラブ連合共和国の合邦など、多くの事績をあげた。

lack1 -5 アクティウムの海戦 mandatory words; When missing, reduce the points

vol /2 540

volume operation max [- min]

Scoring Screen

JS4 – Score Input Screen for [B792W10_1] / scored by [unknown]

解答文番号 採点のための機械学習データがありません。得点決定の予測確率は設定されません。

← Message Window

【解答文】 [1] オスマン帝国 (オスマントルコ語: Devlet-iAlye-iOsmaniye、現代トルコ語: Osmanlımparatorluğuまたはオスマントルコとしても知られています。また、タハリール広場から南のナイル川沿いはガーデン・シティと呼ばれ、イギリス統治時代にエジプト総督府がおかれ開発が進められたエリアである。とくに19世紀後半のエジプト太守イスマーイール・パシャは近代化に熱心であり、スエズ運河の開通にあわせてナイル川東岸の低湿地を開発して、パリの都市計画に倣った新市街を旧市街の西側に建設した。やがて西方を拠点としたオクタヴィアヌスと東方に拠るアントニウスの対立がおこり、オクタヴィアヌスは、プトレマイオス朝エジプトのクレオパトラと結んだアントニウスを前31年アクティウムの海戦でやぶった。しかし、オクタヴィアヌスは、プトレマイオス朝エジプトの女王クレオパトラと結んだため、オクタヴィアヌスはこの連合軍を、前31年、アクティウムの海戦でやぶった。

← Answer sentences

行	チェックボタン		適合度 個別得点	適合度	採点基準ファイル内容	
	Y/1	N/0			種別	配点
-	-	-	-	-	syno	同義置換 - [ムハンマド・アリー ← ムハンマド・アリー] ← Synonym
1	<input checked="" type="radio"/>	<input type="radio"/>	13	0.67	gold	模範回答 20 古代エジプトは、ナイル川を中心に長らく独立王朝が栄えたが、アレクサンドロス、プトレマイオス朝エジプトのクレオパトラは、地中海の覇権国であったローマの内乱にクタイアヌスに対抗するが、アクティウムの海戦に敗れ、エジプトはローマの属州ア半島を統一したイスラム教勢力は、東ローマ帝国とササン朝ペルシアの対立にを征服した。イスラム帝国の版図の一部を継承したファーティマ朝では、1196年にサラディンが宰相となり、イスラム教勢力から聖地エルサレムを奪還しようとする十字軍に対抗した。同じく、十字軍に対抗する中でエジプトに建てられたマムルーク朝は、オスマン帝国の侵攻を受けて滅ぶ。近代には、イギリス本国とインドとの連絡線を断つために、フランスのナポレオンがエジプトに進軍し、解放者を称するが、民衆の抵抗に遭った。対仏戦争に際して、オスマン帝国がエジプトに派遣したムハンマド・アリーは、同地に王朝を建てる。このムハンマド・アリーによる内政への介入を受けるようになり、その保護国となる。ナセルは、1952年に革命を起こしてラエルとの戦争を指導した。
2	<input checked="" type="radio"/>	<input type="radio"/>	15	0.73	gold	模範回答 20 川流域に穀倉地帯を形成して文明を発展させたが、ヒクソスやアケメネス朝などの異民族の支配を受けた。アレクサンドロス大王死後はプトレマイオス朝がこの地を支配したが、クレオパトラがアクティウムの海戦で敗北し、ローマ帝国の属州となった。7世紀以降はアフリカにまで勢力を拡大したイスラム教徒の支配を受け、地中海交易とインド洋を結ぶ交通の要衝として諸王朝が繁栄し、この地を中心に帝国を建設したアイユーブ朝のサラディンは十字軍の撃退にも成功している。その後オスマン帝国の支配を受けたが、19世紀以降はアジア航路の中継地としてヨーロッパからの関心が高まり、ナポレオンはイギリス経済への打撃を狙ってエジプトに遠征した。この混乱後、エジプト総督に就いたムハンマド・アリーは独立とシリア領有を求めたが、列強の干渉により実現しなかった。スエズ運河が開通するとエジプトの重要度は高まり、ウラービー運動鎮圧後にイギリスはエジプトを事実上支配した。第一次世界大戦後にはワフド党を中心に独立が達成されたが、スエズ運河への駐兵権はイギリスに認められていた。エジプト革命を主導したナセルはスエズ動乱でスエズ運河国有化を実現し、イギリスの影響下から脱することに成功している。
						エジプトはナイル川が育んだ肥沃な土壌により紀元前3000年頃から文明が栄え、ピラミッド・太陽暦・神聖文字といった優れた文物を生み出したが、その豊かさに着目した外部勢力の侵入を度々受けた。例えば、紀元前17世紀頃に侵入したヒクソス、紀元前6～4世紀にエジプトを支配したアケメネス朝ペルシアとアレクサンドロス帝国、

↑ Initial Points

再計算 リセット

チェック	適合度	適合度	期待値
得点	得点2	得点	得点
0	0.0	0.0	

← Tentative Score

8

Performance Statistics

Issue	\hat{x} (n=4)	ME $\frac{\sum(x - x_0)}{n}$	MSE $\frac{\sum(x - x_0)^2}{n}$	Ref. all our predicted values (n=18~19)
B	0,0,0,2	0.50	1.00	0 × 11, 2, 8, 14, 15 × 4
C	0,0,0,0	0.00	0.00	0 × 13, 3, 9, 12 × 2, 18 × 2
G	0,0,0,3	0.75	2.25	0 × 10, 2, 3, 7, 8, 9, 19 × 4
L	5,0,0,4	2.25	10.3	0 × 9, 4, 5, 8, 9, 11, 12, 14 × 2, 19 × 2
P	0,4,0,4 ^{.5}	2.13	9.06	0 × 8, 4, 4 ^{.5} × 6, 5 × 2, 7 ^{.5} , 9

- A full mark is 20 points
- Professional evaluations were all zero.

Comments on Evaluation Indicators (ρ or τ)

- How to predict professional evaluations well
 - When $x_0 \equiv 0$, indices based on the correlation are **inappropriate**. $\because \sigma = 0$
 - ρ or τ with only 4 data has almost **no meaning**; the D.F. is only 2 ($=4-2$).
 - The residential errors are **natural** and **proper**.

Conclusion

- Our system can show a certain degree of validity
 - Returned a score close to zero
- Our technique
 - Based on the scoring rubric
 - Considering **superficial** and **semantic aspects (LSI)**
 - Sufficiently suitable

Thank you for your attention.