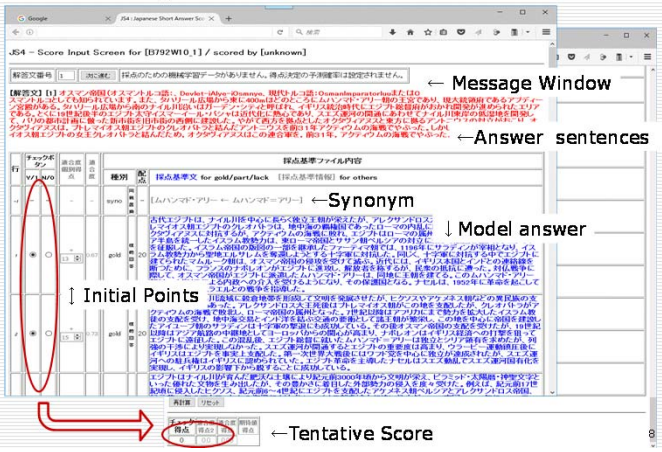
Rubric-based Automated Japanese Short-answer Scoring and Support System Applied to QALab-3

Tsunenori Ishioka (Japan)
Kohei Yamaguchi (Japan)
Tsunenori Mine (Japan)

Writing test

- 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)

Scoring Screen



Support system for short written tests

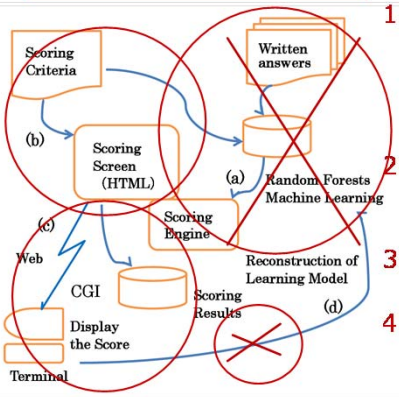
- Accurate language understanding is almost impossible
  - Agreement to scoring rubric
  - Automated Scoring by rubric } Machine
  - ~~Classification scoring by ML~~
  - ~~Overwriting the score~~ ←Human rater
- Simple scoring rubric description
  - Automatic creation of scoring screen

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	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.

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~~

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.

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、アレクサンドロス大王などの征服を受ける。プトレマイオス朝エジ… model answer

Nugget sentence

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… ス大王の征服を受けた。

Max. Vote: 2 = {2,2,3}

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]

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