Automatic Feedback Generation for Short Answer Questions Using Answer Diagnostic Graphs

Momoka Furuhashi [1,2,♣], Hiroaki Funayama [1,2], Yuya Iwase [1,2], Yuichiroh Matsubayashi [1,2], Yoriko Isobe [2], Toru Nagahama [1], Saku Sugawara [3], Kentaro Inui [4,1,2]

[1] Tohoku University, [2] RIKEN, [3] National Institute of Informatics, [4] MBZUAI   ♣ furuhashi.Momoka.p4@dc.tohoku.ac.jp

Goal : Generate feedback to promote the understanding of logical structure of prompt text

Abstract

- Develop a system that uses Answer Diagnostic Graphs (logical structure graphs) to generate feedback on students responses to reading comprehension questions.
- Conduct a practical experiment with high school students to evaluate the potential effectiveness of our feedback

Background

- Individual feedback is necessary because of the variety of errors in students' responses.
- Since only scores are returned, it is difficult for students to understand the reasons for errors in their answers.
- Answer Guides for review are not individualized.

Short-reading Comprehension Questions

- Prompts require students to explain reasons or details for a specific sentence in the prompt text
- Students derive appropriate responses by **understanding the logical structure of the prompt text**

Prompt text is **critical essay**.

Idea

- By graphing the sentence structure, it is able to **clarify the difference between student's responses and the correct answer**. [Roehling+, 2017]
- The errors in students' responses **can be classified into several types.**

Prompt ：傍線部(3)「...世界と私の間に橋がかかってしまうので，距離も生まれてしまう．」とあるが，どういうことか．本文に即して句読点と共に70字以内で説明せよ．
(What does mean in the phrase "...a bridge is built between the world and me, distance is created as well."? Please answer in 70 characters.)

Student response ：言語によって何かを表現することは...世界との生々しい一体感を失うこと．
(Expressing something through language means that ... it leads to losing the vivid sense of unity with the world.)

Before

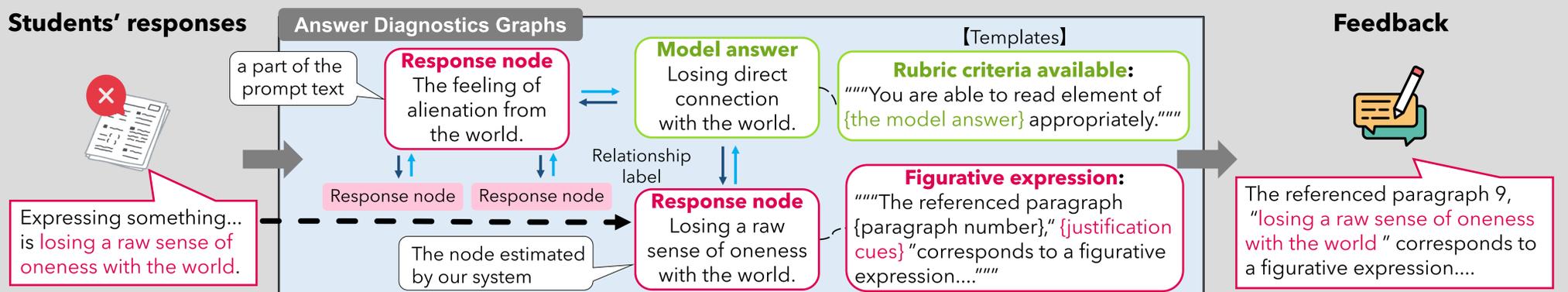I don't understand what is wrong with my response.

8/15 Returns by **score only**.

With our feedback

I understand where to focus in prompt text.

Our System: Automatic Feedback Generation using Answer Diagnostic Graphs

Students' responses

Expressing something... is losing a raw sense of oneness with the world.

Answer Diagnostics Graphs

a part of the prompt text

Response node — The feeling of alienation from the world.

Model answer — Losing direct connection with the world.

Response node   Response node

Relationship label

Response node — Losing a raw sense of oneness with the world.

The node estimated by our system

【Templates】

Rubric criteria available: """You are able to read element of {the model answer} appropriately."""

Figurative expression: """The referenced paragraph {paragraph number}," {justification cues} "corresponds to a figurative expression....."""

Feedback

The referenced paragraph 9, "losing a raw sense of oneness with the world " corresponds to a figurative expression....

Empirical Study

Experimental Setting

- Participants : 41 Japanese high school students
- Two reading comprehension questions
- Oracle feedback

What we research

- Whether our feedback can help participants understand logical structure of the prompt text.
- Whether participants can be motivated to revise their responses.

Results and Challenges

- ✔ The use of our feedback helped participants understand why their responses were wrong and where to focus on in prompt text.
- ✗ There was no significant difference in score improvement between answer explanation cond. and feedback cond. when re-answering.
- ✗ No significant differences in understanding of sentence structure occurred in the answer explanation and feedback conditions
- ✗ Feedback was unusual for participants, so there may have been a positive bias.

Divided equally into **two groups** based on total scores.

A

B

A. Feedback cond.

B. Answer explanation cond.

A. Answer explanation cond.

B. Feedback cond.

Questionnaire

Answer explanation cond. : only answer guide
Feedback cond. : answer explanation & feedback

Future Work

- How do you set up **a bias-free environment**?
- What type of **experimental setting** is best?
  - For example, comparative experiments, etc.
- **What kind of feedback is effective** for reading comprehension questions?
- **What kind of feedback is effective** to promote **understanding of logical structures**?
- What metrics are good for evaluating **engagement**?