Inject Rubrics into Short Answer Grading System

Tianqi Wang\textsuperscript{1,3} Naoya Inoue\textsuperscript{1,3} Hiroki Ouchi\textsuperscript{3} Tomoya Mizumoto\textsuperscript{2,3} Kentaro Inui\textsuperscript{1,3}
\textsuperscript{1}Tohoku University \textsuperscript{2}Future Corporation \textsuperscript{3}RIKEN Center for Advanced Intelligence Project

**Experimental Results**

**Task Introduction**

Short Answer Grading (SAG) is a task of assessing the correctness of short answers to questions automatically.
- Answers are scored by graders with rubrics
- Time consuming especially when limited graders are available

**Key Idea**

- Consider student answers as combination of multiple key concepts.
- Answers are scored based on key concept identification

**Contribution**

- The first study that explores how to incorporate rubric information into neural SAG
- A general framework to extend existing neural SAG models with a component for exploiting rubric information

**Proposed Model**

**An example of SAG**

- Prompt: Starting with mRNA leaving the nucleus, list and describe four major steps involved in protein synthesis.

**Experimental Results**

- Dataset:
  - ASAP-SAS (5 prompts where key elements are explicitly provided)
  - 2226 answers for each prompt on average:
    - 1,704 answers as training set
    - 522 as test set
  - Train the model with various size of training data

**Analysis of $\lambda$**

- Value of $\lambda$ learned from different size of training data
- Higher $\lambda$ means less contribution from rubrics
- Rubric component contributes more when less training data is available

**Instance of attention weights**

- Left: The model successfully found words and phrases most related to the key element, helping the model improve the performance.
- Right: The model incorrectly aligned words in the answer and key element.