TYPIC: A Corpus of Template-Based Diagnostic Comments on Argumentation

Shoichi Naito¹,²,⁵, Shintaro Sawada³, Chihiro Nakagawa³,², Naoya Inoue⁴,²
Kenshi Yamaguchi¹, Iori Shimizu³, Farjana Sultana Mim¹, Keshav Singh¹, Kentaro Inui¹,²
¹Tohoku University, ²RIKEN, ³Osaka Prefecture University, ⁴Stony Brook University, ⁵Ricoh Company, Ltd.

Overview

Goal
• To provide specific diagnostic comments on the students’ counterargument for fostering critical thinking skills

Original argument
We believe homework should be abolished. Forceful study by parents deteriorates family relationships and is one of the reasons why students dislike studying.

Counterargument
They said homework can worsen family relationships. However, it is not true. Asking parents about homework can rather create an opportunity for communication with the family.

Challenge
• What task setting should be designed?
• No dataset with specific diagnostic comments annotated on an argumentation

Contribution
• Propose task setting for Template-Based Diagnosis and evaluate the quality of the induced template set
• Release TYPIC corpus, consisting of counterarguments, some of them are annotated with specific diagnostic comments in both natural language and template format

Dataset

Original argument
Counterargument

Template set
(total 24 patterns)

Diagnostic comments

Natural language
Insufficient explanation of why homework is necessary as a topic of communication between parents and children, as other topics seem to exist as well.

Template format
It lacks an explanation of why homework is a better method to realize family communication instead of other topics.

Task Setting

Template-Based Diagnosis

Step 1. Template Selection
It lacks an explanation of why homework is a better method to realize family communication instead of other topics.

Step 2. Slot Filling
It lacks an explanation of why homework is a better method to realize family communication instead of other topics.

Motivation behind the task setting
• Reliability of automatic evaluation
• When formulated as a generative task, reliable automatic evaluation is problematic
• Considering the characteristics of this task, frequently occurring diagnostic comments are limited and the templates can cover many of them
• Template Selection can be reliably evaluated by metrics such as F1 and Accuracy

Quality of induced template set
• 92.2% of comments can be represented using our template set
• 78% of comments are judged to have the same specificity compared to natural language
• The IAA of template selection is 0.517 for Cohen’s Kappa

Statistics
# of Original argument 10
# of Counterargument 1,000
# of Diagnostic comments 1,082
Diagnostic comments are annotated on 197 counterarguments

Link to the dataset