

UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Formulating Textual Difficulty of Questions as Population who Answer Correctly

Permalink

<https://escholarship.org/uc/item/6np0c8wf>

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 45(45)

Author

Ehara, Yo

Publication Date

2023

Peer reviewed

Formulating Textual Difficulty of Questions as Population who Answer Correctly

Yo Ehara

Tokyo Gakugei University, Koganei, Tokyo, Japan

Abstract

This study proposes a novel approach for extracting the textual difficulty of test questions from learner-adaptive language models that predict the response patterns of learner test-takers. The proposed method uses neural language models, such as BERT, to analyze question texts and formulate the problem of difficulty estimation by estimating the distribution of the number of test-takers who answered correctly in a certain set of test-takers. By utilizing a Poisson binomial distribution, this method can extract the difficulty level of texts that are intuitively interpretable from the fine-tuned model. The proposed method is model-agnostic and can be applied to most language models. This method can also select good questions among those of similar difficulty by selecting the smallest variance in the number of test-takers predicted to be able to answer the questions correctly. Our approach is highly interpretable and achieves high predictive performance owing to the use of neural language models.