UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Using Natural Language Processing Models to Evaluate STEM Book Coherence

Permalink

https://escholarship.org/uc/item/0p39z0kh

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 42(0)

Authors

Miller, Hilary

Wolff, Phillip

Publication Date

2020

Copyright Information

Peer reviewed

Using Natural Language Processing Models to Evaluate STEM Book Coherence

Hilary Miller

Emory University, Atlanta, Georgia, United States

Phillip Wolff

Emory University, Atlanta, Georgia, United States

Abstract

Learning in the STEM disciplines depends on high-quality STEM books, but choosing a textbook can be difficult in the absence of objective measures of text quality. Here we compared two natural language processing approaches for evaluating text cohesion. In Coh-Metrix (Graesser et al. 2004), text cohesion is indicated by the mean cosine value of the all possible pairs of sentence vectors, with sentence vectors based on LSA. We introduce a new method for measuring text coherence based on the deep learning language model RoBERTa (Liu et al., 2019). In this new approach, coherence is measured by determining the average predictability of all of the words in the text, with word predictability a function of each words linguistic context. Coherence as measured by RoBERTa more closely matched the coherence ratings of human judges than did Coh-Metrix. Implications for the assessment and categorization of STEM books are discussed.