UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Analyzing Reliability of Interpretable Parameters in Deep Learning Language Models

Permalink

https://escholarship.org/uc/item/30j787wb

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 44(44)

Author

Ehara, Yo

Publication Date

2022

Peer reviewed

Analyzing Reliability of Interpretable Parameters in Deep Learning Language Models

Yo Ehara

Tokyo Gakugei University, Koganei, Tokyo, Japan

Abstract

Deep-learning has made remarkable progress in recent years. However, its parameter estimation methods are very different from those of traditional statistical models. They are typically iteration-update formula, such as Adam, designed to avoid falling into a local solution even when there are a large number of parameters. While deep-learning-based models still have interpretable parameters such as difficulty or ability, how reliable are these parameters estimated by such estimation methods? In this study, we compared the estimation methods used in conventional statistical models, such as Marginalized Maximum Likelihood Estimation and MCMC, with those used in deep learning, such as Adam, for simple item response models. In experiments, the parameter values estimated by the methods using Adam and others and the parameter values estimated by the methods used in conventional statistical packages showed statistically significant correlation. This result supports that the parameter estimates by deep learning can be trusted to some extent.