THE UNIVERSITY OF **CHICAGO** | **DEPARTMENT OF STATISTICS**

Master's Thesis Presentation

Jin Sung Kim

Department of Statistics
The University of Chicago

"Efficiency of Various BERT Models in Identifying and Classifying Political Bias in Korea Media"

Friday, November 10, 2023, at 9:00 AM (CT), **via Zoom**
Zoom information provided in email announcement

Abstract

Bidirectional Encoder Representations from Transformers ("BERT"), developed by researchers at Google in 2018, has become one of the most popular models to use in various Natural Language Processing ("NLP") tasks due to its state-of-the-art performance. In this paper, I analyze the efficiency of BERT models on the Korean language, specifically with the task of classifying political bias in different Korean media outlets. Using a bag-of-words logistic regression as a baseline method of classification, I analyze how differently fine-tuned Korean BERT models perform in classifying Korean news articles and op-ed pieces, and how they compare to labels scored by a human. Experiments show that as expected, fine-tuned BERT models in general offer better accuracy than logistic regressions. Logistic regression, however, still offers better accuracy than BERT models that have not been fine-tuned or initialized with sub-optimal weights, albeit if and only if the data being predicted has a similar set of features to the training data. In addition, I find that there is little difference in whether the BERT models have been fine-tuned on a more granular level of data or not; and that while fine-tuned BERT models do a good job of identifying documents with a strong political bias or a total lack of political bias, it is not good at identifying very moderate or "centrist" bias.