

Automated L2 Evaluation Using Machine Learning: Integrating Large Language Models with Lexical and Syntactic Features

by

Youngmeen Kim

Abstract

Recent advances in natural language processing (NLP) and artificial intelligence (AI) have sparked significant interest in various disciplines, yet the automated evaluation of second language (L2) proficiency using machine learning (ML) is still nascent. Traditional approaches have analyzed lexical and syntactic features in L2 learners' essays, correlating these with proficiency levels through statistical models (e.g., Crossley & McNamara, 2012). While these methods have identified distinct linguistic characteristics at different proficiency levels, they often overlook critical level-specific features due to rigid statistical assumptions that vary across proficiency levels. Additionally, these methods typically rely on group-based comparisons, potentially missing unique individual features that do not conform to group norms (Dwyer et al., 2018).

This study addresses these limitations by employing ML classification with advanced NLP and text vectorization techniques, including n-gram, TF-IDF (Term Frequency-Inverse Document Frequency), part-of-speech tagging, dependency labeling, and BERT (Bidirectional Encoder Representations from Transformers) embedding, which encompasses a broad spectrum of linguistic features, while also leveraging the robust capabilities of large language models (LLMs). A subset of the Education First-Cambridge Open Language Database (EFCAMDAT-2; Huang et al., 2018), which includes essays by L1 Korean, Japanese, and Chinese learners of English, was used. For the ML classification, the support vector machine (SVM) algorithm was chosen for its interpretability and effectiveness in multi-class classification (Joachims, 1998), aiming to identify key linguistic predictors of proficiency levels. Additionally, this approach enabled the identification of specific linguistic features significant in proficiency classification, thereby providing deeper insights into the linguistic characteristics of texts across different levels. This research not only enhances the efficacy of automated language assessments but also sheds light on the distinctive linguistic characteristics associated with different proficiency levels.