Zero Shot Neural Machine Translation System for Low Resource Languages

Abstract

Neural Machine Translation (NMT) systems perform exceptionally well in resource-rich scenarios, however, their performance is sub-optimal in low-resource settings particularly when translation systems are optimized for domain-specific translations. Moreover, the quality of translations also suffers when working within specialized domains. The primary factor of this performance degradation is the scarcity of high-quality parallel training examples for low resource languages and domain-specific settings. To address this challenge, this research introduces a novel methodology, for developing in-domain training data, leveraging the transformative potential of Large-scale Language Models (LLMs). We demonstrate that LLMs excel at selecting potential parallel sentences from comparable corpora as well as serve as effective builders of machine translation models. Additionally, the study makes a significant contribution to exploiting the potential of LLMs for domain adaptation in machine translation, while focusing on the biomedical domain.