Breaking Language Barriers With AI: Towards Neural No-Resource Translation

Thakur, Madhavendra (School: Trinity School)

No-resource languages—those with minimal or no digital representation—pose unique challenges for machine translation (MT). Unlike low resource languages, which rely on limited but existent corpora, no-resource languages often have fewer than 100 sentences available for training. This work explores the problem of no-resource translation through three distinct workflows: fine-tuning of translation-specific models, in-context learning with large language models (LLMs) using chain-of-reasoning prompting, and direct prompting without reasoning. Using Owens Valley Paiute as a case study, we demonstrate that no-resource translation demands fundamentally different approaches from low-resource scenarios, as traditional approaches to machine translation, such as those that work for low-resource languages, fail. Empirical results reveal that, although traditional approaches fail, the in-context learning capabilities of general-purpose large language models enable no-resource language translation that outperforms low-resource translation approaches and rivals human translations (BLEU 0.45-0.6); specifically, chain-of-reasoning prompting outperforms other methods for larger corpora, while direct prompting exhibits advantages in smaller datasets. As these approaches are language-agnostic, they have potential to be generalized to translation tasks from a wide variety of no-resource languages without expert input. These findings establish no-resource translation as a distinct paradigm and propose a state-of-the-art solution for no-resource translation, providing practical and theoretical insights for language preservation and the learning capabilities of LLMs.

Awards Won: