IN-CONTEXT LEARNING AND PIVOT-BASED TRANSLATION COMPARISON FOR LOW-RESOURCE MACHINE TRANSLATION

Andre Hutagaol, Tyra Silaphet, Yen-Shuo Su

Georgia Institute of Technology
{ahutagaol3, tsilaphet3, ericsu}@gatech.edu

1. OBJECTIVES

1.1. Motivation

Neural machine translation (NMT) has achieved state-of-the-art (SOTA) performance for high-resource languages, where large parallel corpora enable robust model training, but translation quality declines significantly for low-resource languages, which lack sufficient bilingual data [1]. Recent advances in multilingual machine translation (MMT), such as Meta AI's No Language Left Behind (NLLB) model, have significantly improved translation quality for many underrepresented languages by leveraging large-scale multilingual training. Despite its broad coverage, NLLB's performance varies across different low-resource language pairs, especially when parallel training data is extremely scarce [2].

Two strategies have emerged to address this issue:

- Pivot-based translation: a widely used technique where a low-resource language is translated into a high-resource language before being translated into the target language.
- In-Context Learning (ICL): Large language models (LLMs) like GPT-4 can perform translations without fine-tuning, using set of sample translations to generalize from. Recent studies have demonstrated that ICL can match fine-tuned NMT models for some language pairs [3].

This study aims to compare NLLB's direct and pivot-based translation aginst ICL-based translation to determine which approach is most effective for low-resource machine translation.

1.2. Task Definition

1.2.1. Problem Statement

This project will investigate the following research questions:

- How does NLLB perform in direct low-resource translation?
- Does pivot-based translation improve translation quality?
- Can pivot-based translation be an augmentation approach?
- Can in-context learning outperform pivot-based translation?

1.2.2. Dataset

For our project, we decided to use FLoRes-200 and WMT21 datasets. We will use English-Nepalese, English-Sinhala, English-Russian, and Nepalese-Sinhala language pairs to train and evaluate our model.

2. RELATED WORK

Multilingual Machine Translation models leverage cross-lingual transfer to improve low-resource translation. Meta AI's NLLB [2] supports 200 languages with direct translation, reducing reliance on English as a pivot. However, performance varies depending on data availability and linguistic similarity. The impact of pivot-based translation on NLLB's effectiveness remains an open question.

Pivot-based translation improves translation quality by first translating a low-resource language into a high-resource intermediary before translating to the final target language. This approach has been effective for low-resource pairs but suffers from error propagation, where mistakes in the first step may affect the second.

An alternative strategy is in-context learning (ICL), where LLMs translate by leveraging a small number of sample translations in the input prompt rather than undergoing fine-tuning [4]. Chen et al. (2023) demonstrated that ICL can match or outperform fine-tuned NMT models when example selection is optimized, especially for low-resource pairs [3]. However, their study did not compare ICL to pivot-based translation, leaving open questions about its relative effectiveness.

3. APPROACH

Our goal is to translate between two low-resource languages, specifically from Nepalese to Sinhala. We will use four different approaches to train this model and we will compare the performance of these approaches using the BLEURT metric [5].

- Direct translation: This approach will serve as the baseline for the Nepalese to Sinhala translation task, in which NLLB is used to perform direct translation on the languages.
- Pivot-based without fine-tuning: In this approach, we utilize NLLB in a two-step process: one for translating from Nepalese to English, and the other for translating from English to Sinhala to evaluate the result of translating from Nepalese to Sinhala. We will use the FLoRes-200 dataset to test the performance of this approach.
- Pivot-based with fine-tuning: This approach uses NLLB to generate translation pairs of English-Nepalese, English-Sinhala, English-Russian, English-German, etc. We can then use these paired translations, such as Nepalese-Sinhala, Nepalese-Russian, German-Sinhala, to fine-tune the model. This will allow the model to have more exposure to these low-resource languages and improve translation performance.
- In-context learning technique: We will perform in-context learning to perform translation between the Nepalese-Sinhala pair directly while incorporating only a few examples from the FLoRes-200 dataset with the GPT-4 model.

Our contribution to this topic lies in our attempt to build a model that can translate from one low-resource language to another. While existing research primarily focuses on translating from high-resource languages, such as English, to low-resource languages [6], our work aims to address the pair of low resource languages. Additionally, we will compare four different approaches to perform the low-resource language translation task in our project.

4. REFERENCES

[1] Madhavendra Thakur, "Towards neural no-resource language translation: A comparative evaluation of approaches," 12 2024.

[2] NLLB Team, Marta R. Costa-jussà, James Cross, Onur Çelebi, Maha Elbayad, Kenneth Heafield, Kevin Heffernan, Elahe Kalbassi, Janice Lam, Daniel Licht, Jean Maillard, Anna Sun, Skyler Wang, Guillaume Wenzek, Al Youngblood, Bapi Akula, Loic Barrault, Gabriel Mejia Gonzalez, Prangthip Hansanti, John Hoffman, Semarley Jarrett, Kaushik Ram Sadagopan, Dirk Rowe, Shannon Spruit, Chau Tran, Pierre Andrews, Necip Fazil Ayan, Shruti Bhosale, Sergey Edunov, Angela Fan, Cynthia Gao, Vedanuj Goswami, Francisco Guzmán, Philipp Koehn, Alexandre Mourachko, Christophe Ropers, Safiyyah Saleem, Holger Schwenk, and Jeff Wang, "No language left behind: Scaling human-centered machine translation," 2022.

[3] Yufeng Chen, "Enhancing machine translation through advanced in-context learning: A methodological strategy for gpt-4 improvement," 2023.

[4] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei, "Language models are few-shot learners," 2020.

[5] Thibault Sellam, Dipanjan Das, and Ankur P. Parikh, "BLEURT: learning robust metrics for text generation," *CoRR*, vol. abs/2004.04696, 2020.

[6] Xavier Garcia, Yamini Bansal, Colin Cherry, George Foster, Maxim Krikun, Melvin Johnson, and Orhan Firat, "The unreasonable effectiveness of few-shot learning for machine translation," in *Proceedings of the 40th International Conference on Machine Learning*, Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett, Eds. 23–29 Jul 2023, vol. 202 of *Proceedings of Machine Learning Research*, pp. 10867–10878, PMLR.