International Journal of Interdisciplinary Innovative Research & Development (IJIIRD)
ISSN: 2456-236X
Vol. **11** Special Issue **02** | **2026**

Enhancing Accessibility of Government Notices through LLM-based Multilingual Translation

Mr.Rohit.V.Talele[1], Ms.Shruti.S.Ujjainkar[2], Mr.Shiva.D.Hinge[3], Mr.Saurabh.R.Patil[4], Prof. Manjiri.U.Karande [5]

[1,2,3,4]*Student , Computer science and Engineering, Padm. Dr.V. B.Kolte College of Engineering, Malkapur*
[5]*Professor , Computer science and Engineering, Padm. Dr.V. B.Kolte College of Engineering, Malkapur*

DOI: 10.5281/zenodo.19254264

ABSTRACT

In the field of computational linguistics, addressing machine translation (MT) challenges for low-resource languages remains crucial, as these languages often lack extensive data compared to high resource languages. General large language models (LLMs), such as GPT-4 and Llama, primarily trained on monolingual corpora, face significant challenges in translating low-resource languages, often resulting in subpar translation quality. This study introduces Language-Specific Fine-Tuning with Low-rank adaptation (LSFTL), a method that enhances translation for low-resource languages by optimizing the multi-head attention and feed-forward networks of Transformer layers through low-rank matrix adaptation. LSFTL preserves the majority of the model parameters while selectively fine-tuning key components, thereby maintaining stability and enhancing translation quality. Experiments on non-English centered low-resource Asian languages demonstrated that LSFTL improved COMET scores by 1-3 points compared to specialized multilingual machine translation models. Additionally, LSFTL's parameter-efficient approach allows smaller models to achieve performance comparable to their larger counterparts, highlighting its significance in making machine translation systems more accessible and effective for low-resource languages.

***Keywords:** - Machine translation, low-resource languages, large language models, parameter- efficient fine-tuning, LoRA*.

1. INTRODUCTION

The Indian government places significant reliance on the distribution of notices as a primary means of conveying vital and pertinent information to the masses. This method has proven to be a robust and dependable way to establish a direct connection with a large and diverse population, ensuring that every citizen can remain informed about the government's ongoing activities and developments. However, a notable challenge arises in this process– the government often employs the language that is predominantly spoken within a particular state. While this choice effectively reaches the majority of the population, it leaves a significant portion of the minority population struggling to grasp the content of these notices. The linguistic diversity of India is one of its defining characteristics, with numerous regional languages spoken across the nation. Consequently, a substantial portion of the populace may not be fluent or familiar with the language chosen for official notices in their respective regions. This language barrier can be a substantial impediment to the government's objective of ensuring that all citizens, regardless of their linguistic background, can access critical information. To address this challenge, a systematic solution is imperative. The government should develop and implement a robust system capable of seamlessly translating and converting local languages into various Indian languages. By doing so, these notices can be made accessible to the minority population, transcending linguistic boundaries and enabling everyone to comprehend and engage with the government's communications effectively. Such a system would not only promote inclusivity and equal access to information but also uphold the principles of democracy by ensuring that every citizen can actively participate in the governance process. It would reflect the government's commitment to fostering unity in diversity and strengthen its relationship with all segments of society, reinforcing the bonds of a vibrant and pluralistic nation. In essence, the development of a comprehensive language conversion system is a vital step towards creating a more connected, informed, and equitable society in India.

2. LITERATURE REVIEW

Literature review related to exiting system/methodology

The Machine Translation System using Open NMT has been designed specifically to facilitate learning in native Indian languages by translating English into 22 different Indian languages. However, the system's effectiveness is limited when faced with complex English sentences, as it has primarily been trained on simple query phrases found in the dataset. [1] The paper explores the application of DNN networks and the Word2vec tool in multilingual language translation. Machine translation faces several limitations, including problems associated with a restricted vocabulary, data sparsity, and the preservation of the history of vector values. [2] The proposed system utilizes a combination of RBMT(Rule-based Machine Translation) and SMT (Statistical Machine Translation) techniques to leverage the advantages of each approach for English to Marathi translation. Open NLP (Open Natural Language Processing) is employed to construct the hybrid machine translation model, which effectively integrates RBMT and SMT. [3]

Sr. No.	Title of Paper	Review	Analysis/Limitations
[1]	Real Time Machine Translation System for English to Indian language	The Machine Translation System finds how to translate English Language into 22 Indian Languages to help learning in native languages using OpenNMT. [1]	The system has been developed only for English to Indian languages. The model is not able to translate the complicated English sentences due to simple query phrases in the dataset.
[2]	Machine Translation using Deep Learning: An Overview	The paper discusses the use of DNN net- works in translation of multilingual language along with the use of Word2vec tool.[2]	The challenges associated with machine translation are insufficient vocabulary, data sparsity, and maintaining a history of vector values. The use of RNNs is very complicated for large datasets which require a high GPU.
[3]	Hybrid machine translation for English to Marathi: A research evaluation in Machine Translation: (Hybrid translator)	The system proposed uses the combination of RBMT and SMT to utilize the advantage of each machine translation technique for English to Marathi Translation. Open NLP is used for constructing a hybrid machine translation model [3]	Vuforia is a popular freemium platform where using Vuforia Object Scanner, we can scan and create object targets. The recognition process can be implemented using the database (local or cloud storage).

3. METHODOLOGY

Users upload images or PDFs to the website, where the text is extracted and processed. Extracted text is structured into sentences, then segmented into word arrays. These arrays are fed into a language model (LLM) for translation, with the model leveraging training data. Predicted translations are reconstituted into sentences and displayed to users. This process enhances accessibility and comprehension of Marathi government documents, fostering citizen engagement and alleviating complexities. It combines Optical Character

Recognition (OCR)for text extraction, LLM for translation, and user-friendly web interfaces for effective information delivery.

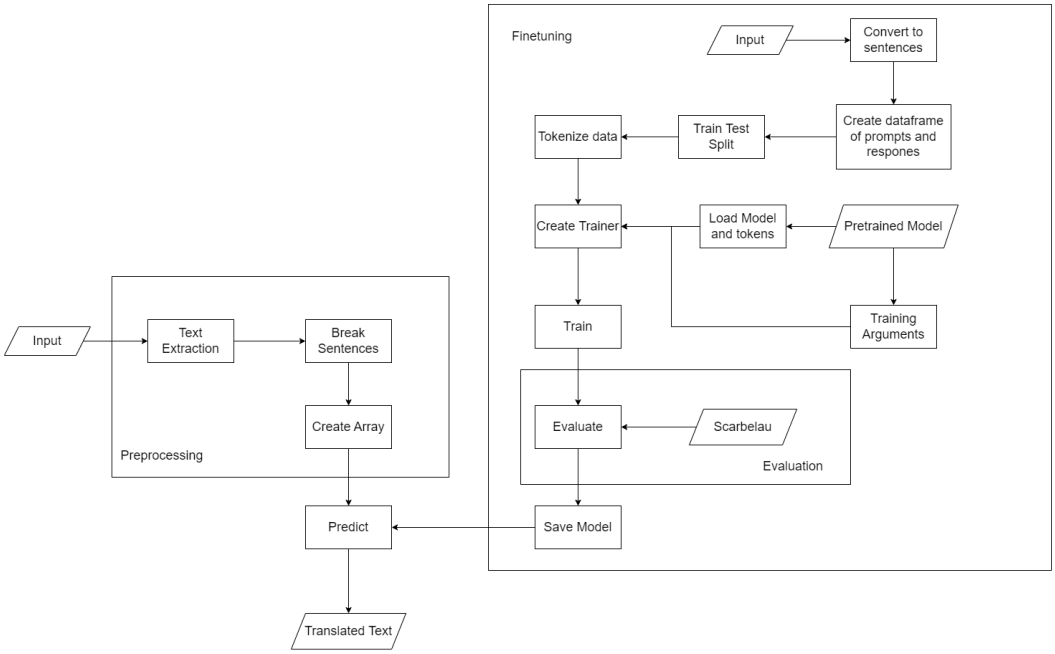

Figure: System Diagram.

Input The input to the system is the source text to be translated. This text can be in any language, but it is typically preprocessed to clean it up and normalize it.

Text Extraction The text extraction module extracts the text to be translated from the input. This may involve breaking up the input into sentences, identifying the source language, and detecting any special characters or formatting.

Preprocessing The preprocessing module performs a variety of tasks on the extracted text, such as tokenizing the text, removing stop words, and normalizing the case.

Pretrained Model The pretrained model is a neural network that has been trained on a large corpus of parallel text. The model is used to generate translations of the preprocessed text.

Finetuning The finetuning module is used to fine-tune the pretrained model on a specific domain or task. This can be done by providing the model with additional training data or by adjusting the model's hyperparameters.

Evaluation The evaluation module is used to evaluate the performance of the MT system on a held- out test set. This helps to ensure that the system is producing high-quality translations.

Output The output of the MT system is the translated text in the target language. This text can be used in a variety of applications, such as translation websites, machine translation APIs, and multilingual software.

4. CONCLUSIONS

In conclusion, The project *"Translation of Government Notices in Indian Languages using LLM"* provides an effective solution to overcome the linguistic barriers that citizens often face in accessing government information. By combining **OCR-based text extraction**, **fine-tuned Large Language Models**, and a **user-friendly web interface**, the system ensures that official notices are translated into clear, understandable, and inclusive formats. This approach not only enhances **citizen– government communication** but also empowers minority communities who may not be proficient in the regional language, thereby fostering **transparency, inclusivity, and democratic participation**. While the current implementation focuses on Marathi, the framework can be scaled to support multiple Indian languages, making it a powerful tool for nationwide deployment. In essence, the project demonstrates the practical potential of **AI-driven language technologies** in solving real- world socio-technical challenges, ultimately contributing to a more connected, informed, and equitable society.

5. REFERENCES

[1] Touvron, Hugo, et al. “Llama 2: Open Foundation and Fine-Tuned Chat Models.” arXiv.Org, 19 July 2023, arxiv.org/abs/2307.09288.
[2] S. P. Singh, A. Kumar, H. Darbari, L. Singh, A. Rastogi and S. Jain,”Machine translation using deep learning: An overview,” 2017 International Conference on Computer, Communications and Electronics (Comptelix), Jaipur, India, 2017, pp. 162-167, doi: 10.1109/COMPTELIX.2017.8003957.
[3] P. Salunkhe, A. D. Kadam, S. Joshi, S. Patil, D. Thakore and S. Jadhav, ”Hybrid machine translation for English to Marathi: A research evaluation in Machine Translation: (Hybrid translator),” 2016 International Confer ence on Electrical, Electronics, and Optimization Techniques (ICEEOT), Chennai, India, 2016, pp. 924-931, doi: 10.1109/ICEEOT.2016.7754822.
[4] Daron Acemoglu and Pascual Restrepo. Artificial intelligence, automa tion, and work. In The economics of artificial intelligence: An agenda, pages 197–236. University of Chicago Press, 2018.
[5] Joshua Ainslie, James Lee-Thorp, Michiel de Jong, Yury Zemlyanskiy, Federico Lebr´ on, and Sumit Sanghai. Gqa: Training generalized multi query transformer models from multi-head checkpoints, 2023.
[6] Ebtesam Almazrouei, Hamza Alobeidli, Abdulaziz Alshamsi, Alessan dro Cappelli, Ruxandra Cojocaru, Merouane Debbah, Etienne Goffinet, Daniel Heslow, Julien Launay, Quentin Malartic, Badreddine Noune, Bap tiste Pannier, and Guilherme Penedo. Falcon-40B: an open large language model with state-of-the-art performance. 2023.
[7] Jacob Austin, Augustus Odena, Maxwell Nye, Maarten Bosma, Henryk Michalewski, David Dohan, Ellen Jiang, Carrie Cai, Michael Terry, Quoc Le, and Charles Sutton. Program synthesis with large language models, 2021.
[8] Yuntao Bai, Andy Jones, Kamal Ndousse, Amanda Askell, Anna Chen, Nova DasSarma, Dawn Drain, Stanislav Fort, Deep Ganguli, Tom Henighan, et al. Training a helpful and harmless assistant with reinforce ment learning from human feedback. arXiv preprint arXiv:2204.05862, 2022a.
[9] Emily M Bender, Timnit Gebru, Angelina McMillan-Major, and Shmar garet Shmitchell. On the dangers of stochastic parrots: Can language mod els be too big? In Proceedings of the 2021 ACM conference on fairness, accountability, and transparency, pages 610–623, 2021b.
[10] Shaily Bhatt, Sunipa Dev, Partha Talukdar, Shachi Dave, and Vinodkumar Prabhakaran. Re-contextualizing fairness in nlp: The case of India, 2022