WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

MACHINE TRANSLATION FROM HINDI TO ENGLISH

Darshan Vadagave

MCA Student, Department of Computer Science, Rani Channamma University, Belagavi,

ABSTRACT :

The "Hindi to English Machine Translation" employs natural language processing (NLP) as a groundbreaking strategy to surmount linguistic obstacles and promote cross cultural dialogue. This technology stands out as a crucial instrument for fostering accessibility and understanding in a world that is increasingly interconnected, where linguistic diversity can pose challenges to effective communication. The goal of the innovator is to develop a sophisticated, context-sensitive translation system that not only translates Hindi into English but also captures the subtleties, nuances, and cultural sensitivities embedded in language. By combining state-of-the-art neural machine translation (NMT) models with linguistic knowledge, we strive to create a translation tool that goes beyond simple word-for-word translations, ensuring that the cultural context and idiomatic expressions are accurately conveyed.

KEYWORDS— Cross cultural, Hindi to English, Linguistic, Machine translation, Neural machine translation (NMT)

1 INTRODUCTION :

In an era of heightened interconnectedness, communication extends beyond the boundaries of geography. Language, which once posed a significant challenge, can now be effectively navigated through the use of machine translation tools. Within this framework, Hindi, a language spoken by millions, has emerged as a focal point for the advancement and enhancement of machine translation systems. This research paper offers a concise overview of the current state of machine translation tools aimed at facilitating conversions from Hindi to English. The significance of this topic lies in its capacity to empower individuals and enterprises operating across diverse regions. Whether for cross-cultural dialogue, content localization, or the acquisition of global knowledge, precise and efficient Hindi to English machine translation is crucial. Our investigation commences with an analysis of existing machine translation tools, evaluating their strengths and weaknesses, and determining their applicability in real-world scenarios. We will highlight opportunities for improvement and enhance the quality of Hindi-English translations. The field of machine translation has progressed remarkably, with Neural Machine Translation (NMT) leading this evolution. NMT models have shown an exceptional ability to grasp linguistic subtleties and context, thereby transforming translation accuracy. This article will explore how NMT technology has been leveraged to enhance translation quality for Hindi speakers.

2 LITERATURE SURVEY :

In the field of Natural Language Processing (NLP), Machine Translation (MT) acts as a bridge for communication between different languages. MT employs machine translation techniques to address language ambiguity while preserving the intended meaning. As MT systems evolved from rule-based approaches to corpus-based methods, the reliance on linguistic expertise diminished. This evolution has facilitated a wide array of NLP tasks, such as Named Entity Recognition, Speech Tagging, Chunking, Word Sense Disambiguation, and has also addressed the challenges of language diversity in Interlingua-based MT [1].

Analyzing machine translation systems that involve Indian languages and their relationships with international languages indicates that it is generally easier to translate between language pairs that exhibit minor structural differences, such as Hindi and English, compared to those with considerable structural differences, like Hindi and Punjabi[2].

The use of direct machine translation at the word level is relatively uncommon. In this process, there is no intermediary representation employed when converting words into the target language[5].

Direct machine translation represents a basic translation technique that translates each word from the source sentence into its corresponding word in the target sentence, implementing necessary syntactical modifications to uphold the sentence structure of the language. [16][18].

The NMT system, utilizing LSTM, effectively addresses both the long-term dependency issue and the sequence-to-sequence learning challenge associated with source and target phrases of varying lengths. This system demonstrates exceptional capabilities in context analysis, thereby improving translation predictions[3].

The process of Direct Machine Translation requires a separate translator for each specific source and target language pair. A multimodal machine translation framework was created for the Hindi Visual Genome dataset, which translates between English and Hindi. According to the author, the performance of a machine translation system is significantly improved when it leverages various input modalities[7].

3 PROPOSED METHODOLOGY :

The process of translating from Hindi to English through Neural Machine Translation (NMT) represents a cutting-edge methodology that utilizes advanced deep learning neural networks to ensure precise and contextually relevant translations

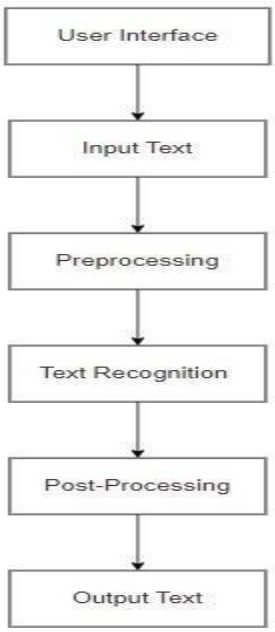
NMT techniques are applied in this process :

- Data Collection: Initiate the process by gathering a comprehensive bilingual dataset that includes aligned rulings in both Hindi and English. This dataset will serve as the foundational training material for the Neural Machine Translation (NMT) model.
- Data Preprocessing: Refine and prepare the dataset by eliminating extraneous elements, special characters, and any inconsistencies in formatting. Segment the text into words or subword units to effectively manage variations in language.
- NMT Model Selection: Opt for an NMT architecture, such as the transformer model, which has demonstrated significant efficacy in translation tasks. These models are specifically designed to capture long-range dependencies in language and excel in handling contextual nuances.
- Training: The specified NMT model should be trained using the preprocessed bilingual dataset. Throughout the training process, the model acquires the ability to associate Hindi sentences with their corresponding English translations by adjusting its parameters based on the input-output pairs.
- Attention Mechanism: NMT models often utilize attention mechanisms that enable the model to focus on particular segments of the source text while producing the translation. This attention mechanism aids the model in capturing context and enhancing the quality of the translation.
- Hyperparameter Tuning: Adjust the model's hyperparameters, such as learning rates, batch sizes, and architectural parameters, to maximize translation performance.
- Evaluation Metrics: Implement evaluation metrics such as BLEU (Bilingual Evaluation Understudy) to quantitatively measure the quality of translations. BLEU assesses the machine-generated translations against human references.
- Validation and Testing: Divide the dataset into training, validation, and testing subsets. The validation set is used to monitor the model's performance during training, while the testing set is employed to evaluate its generalization capabilities.
- Integrate the trained neural machine translation (NMT) model into user-friendly applications or interfaces, enabling users to input text in Hindi and receive English translations in an accessible manner.
- The application of NMT methodologies in translating from Hindi to English demonstrates remarkable translation quality, positioning it as an essential resource for overcoming language barriers across multiple fields.

3.1 SYSTEM ARCHITECTURE

This architecture illustrates that users input Hindi text through the user interface. The preprocessing service subsequently receives this input, where it is cleaned and normalized. Following this, the text recognition component identifies the prepared Hindi text and translates it into English. A post-processing service then refines the translated material. Ultimately, the system generates an improved final output in English, as depicted in Figure 1 below.

Figure1: General architecture of converting Hindi text to English text



The above architecture diagram each module is explained in details

- User Interface: The front-end server is responsible for hosting the user interface and facilitating user interactions.
- Input Text: It oversees the management of user input, ensuring that it is accurately captured by the system.
- Preprocessing: This stage involves the cleaning and normalization of the raw Hindi text to prepare it for subsequent processing.
- Text Recognition: This process identifies Hindi text through optical character recognition (OCR) and subsequently translates it into English.
- Post Processing: This phase improves the recognized text, which may involve enhancing formatting, rectifying errors, or fine-tuning the translation.
- Output Text: This step produces the final translated text in English.

4 RESULT :

Input text : The input for a machine translation system transitioning from Hindi to English comprises text or sentences presented in the input Hindi Text (figure 2).

Figure 2: Input Hindi Text

Output text : The output generated by the machine translation system is the English translation derived from the input illustrated in figure 2.

Figure 3: Output English Text

5 CONCLUSIONS AND FUTURE WORK

The initiative focused on Hindi to English machine translation represents a significant endeavor that illustrates the role of technology in facilitating communication across different languages. While there is still a need for further enhancements to improve accuracy, this tool possesses considerable potential to foster effective communication and cultural exchange between speakers of Hindi and English in various settings. We are optimistic about its prospective impact and contribution to enhancing global connectivity.

Future efforts for the project will prioritize improving translation accuracy, refining the user experience, and addressing potential biases and ethical considerations. Our Hindi to English machine translation technology will be continuously developed and expanded to better meet user needs and promote effective communication between languages.

ACKNOWLEDGEMENT

I am grateful to Dr. Shivanand Gornale, Professor, Department of Computer Science, Rani Channamma University, Belagavi for his valuable guidance for completion of this work.

REFERENCES :

1. Shachi Dave, Jignashu Parikh, and Pushpak Bhattacharyya, "Interlinguabased English Hindi Machine Translation and Language Divergence," Machine Translation, vol. 16(4), pp.251–304 (2001).

2. V Goyal, G S Lehal. "Advances in Machine Translation Systems". Language In India, Vol. 9, No. 11, 2009, pp. 138-150.

3. Anoop Kunchukuttan, Pratik Mehta, and Pushpak Bhattacharyya, "The IIT bombay english-hindi parallel corpus," In proceedings of the eleventh international conference on language resources and evaluation (LREC 2018), May 7-12, 2018, Miyazaki, Japan, isbn: 979-10-95546-00-9, (2018).

4. Deepak Khemani, "A First Course in Artificial Intelligence", 2013, pages 702-703

5. M. D. Okpor, "Machine Translation Approaches: Issues and Challenges", IJCSI International Journal of Computer Science Issues, Vol. 11, Issue 5, No 2, September 2014

6. Wang, D., & Xiong, D. (2021). "Efficient Object-Level Visual Context Modeling for Multimodal Machine Translation: Masking Irrelevant Objects Helps Grounding." In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 35, No. 4, pp. 2720-2728).

7. Wagadiya N, Ravarta P English-Hindi Translation system with Scarce resources. International journal of innovative research and development.

8. ManishaRoy,MadhuraWalivadekar,PoojaKadam "Sentence Validation Using Natural Language Processing". Proceedings of IRF International Conference, 23rd February 2014, Pune, India. ISBN: 978-93-82702-61-0

9. Shibli Syeed Ashrafi , Md. Humayun Kabir , Md. Musfique Anwar, A.K.M .Noman, Gulshan, "English to Bangla Machine Translation System Using Context - Free Grammars" ,IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 3, No 2, May 2013.

10. Vishal Goyal, Gurpreet Singh Lehal "Hindi To Punjabi Machine Translation System". Proceedings of the ACLHLT 2011 System Demonstrations, pages 1- 6,Portland,Oregon USA.21 June 2011.Association for Computational Linguistics.

11. Ankita Agarwal, Pramila, Shashi Pal Singh, Ajai Kumar, Hemant Darbari "Morphological Analyzer for Hindi – A Rule Based Implementation" ,International Journal of Advanced Computer Research (ISSN (print): 2249-7277 ISSN (online): 2277-7970) Volume-4 Number-1 Issue-14 March-2014.

12. David Peter S, Latha R. Nair" Machine Translation System for Indian Languages",International Journal of Computer Applications (0975 – 8887) Volume 39– No.1, February 2012.

13. Sudip Naskar, Sivaji Bandyopadhyay "Use of Machine Translation in India:Current Status", Jadavpur University,Kolkata.

14. Holger Schwenk, "Continuous space translation models for phrasebased statistical machine translation," In Proceedings of COLING 2012: Posters. The Coling 2012 Organizing Committee, Mumbai, India, pp. 1071–1080 (2012).

15. Sainik Kumar Mahata, Dipankar Das, and Sivaji Bandyopadhyay, "Mtil2017: Machine translation using recurrent neural network on statistical machine translation," Journal of Intelligent Systems pp. 1–7 (2018).

16. J. Nair, Amrutha Krishnan K, Deetha R, "An Efficient English to Hindi Machine Translation System Using Hybrid Mechanism", 2016 Intl. Conference on Advances in Computing, Communications and Informatics (ICACCI). IEEE. 2016, Sept, 21-24.

17. N. Wagadiya, P. Ravarta, "English-Hindi Transl ation system with Scarce resources,. International journal of innovative research and development.

18. L. R. Nair, David Peter S., "Machine Translation Systems for Indian Languages", International Journal of Computer Applications (0975 – 8887) , Volume 39– No.1, February 2012.