MULTIDISCIPLINARY SCIENCE JOURNAL

RESEARCH ARTICLE
Published Online: October 11, 2024
https://doi.org/10.31893/multiscience.2025146

Hybrid NMT model and comparison with existing machine translation approaches

Ritesh Kumar Dwivedi[a] | Parma Nand[a] | Om Pal[b]

[a]Department of Computer Science and Engineering, Sharda University,Greater Noida, India.
[b]Department of Computer Science, University of Delhi, New Delhi, India.

Abstract Neural machine translation has transformed automated translation, surpassing traditional methods with its significant accuracy improvements. However, despite its successes, NMT still encounters several challenges, such as handling low-resource languages, maintaining contextual coherence, and addressing ambiguities in translation. This research presents a novel hybrid NMT model to overcome these limitations. It combines the strengths of traditional translation methods with modern deep learning approaches. We conduct a comprehensive comparative analysis of our hybrid model against existing machine translation approaches, including rule-based machine translation (RBMT), SMT, and state-of-the-art NMT systems. Evaluation metrics BLEU is utilized to assess the performance across English-Hindi,English-Marathi,English-Bengali language pairs and domains. Our results demonstrate that the hybrid NMT model achieves superior accuracy and fluency in translation tasks, particularly for low-resource languages and complex sentence structures. This research highlights the potential of combining different machine translation approaches and findings suggest that integrating these methods can significantly improve translation quality. The findings offer valuable insights for future research and development of more robust and versatile translation systems. Our results demonstrate that the hybrid model offers significant improvements in translation accuracy, making it a promising approach for multilingual machine translation tasks. NMT surpasses both RBMT and SMT with a BLEU score of 35.6, highlighting its effectiveness in managing context and semantics. Qualitative assessments suggest that the hybrid model effectively minimizes common translation errors, making it a robust solution for multilingual machine translation tasks. Hybrid Neural Machine Translation (NMT) models are increasingly being applied in real-world applications where the combination of rule-based, statistical, and neural approaches offers distinct advantages.

Keywords: NMT, SMT, Transformer model, RBMT

1. Introduction

Machine translation (MT) has undergone significant evolution over the past few decades, progressing from early rule-based systems to sophisticated neural networks. Initially, rule-based machine translation (RBMT) systems depended on comprehensive linguistic rules. While effective in some contexts, these systems are limited by their inflexibility and inability to handle the nuances of natural language effectively (Hutchins, 2005). The advent of statistical machine translation (SMT) marked a substantial shift, introducing data-driven approaches that leveraged parallel corpora to learn translation patterns probabilistically. Phrase-based SMT models, such as those introduced by Koehn et al. (2003), showed substantial improvements in translation accuracy and fluency over RBMT. However, SMT systems often struggle with long-range dependencies and contextual coherence because of their reliance on fixed-length phrase pairs and limited context windows (Och & Ney, 2003). In recent years, neural machine translation (NMT) has emerged as a state-of-the-art approach, driven by advancements in deep learning. Sequence-to-sequence models with attention mechanisms, introduced by Bahdanau et al. (2015), and the later transformer architecture by Vaswani et al. (2017), have revolutionized the field by offering a more flexible and context-aware framework for translation. NMT models excel in capturing semantic nuances and producing fluent translations; however, they are not without challenges. Issues such as handling low-resource languages, managing rare words, and maintaining consistency in long texts persist, highlighting the need for further improvements (Bentivogli et al., 2016). To address these challenges, this research investigates the integration of the SMT and NMT techniques to develop a hybrid machine translation model. The hybrid approach aims to combine the strengths of the phrase-based translations of SMT with the contextual learning capabilities of NMT, thereby enhancing the overall translation quality and robustness. Hybrid models have shown promise in various studies, offering improvements in translation accuracy and fluency by leveraging the complementary strengths of different methodologies (Stahlberg et al., 2018). This study presents a novel hybrid NMT model and conducts a comprehensive comparative analysis against existing MT approaches. By evaluating performance across multiple language pairs and domains

 Received: July 13, 2024 | Accepted: August 22, 2024

via metrics such as the BLEU, we aim to demonstrate the efficacy of the hybrid model in overcoming the limitations of purely neural or statistical methods. The findings of this research contribute to the growing body of evidence supporting hybrid approaches in machine translation. In exploring hybrid models, our research addresses critical gaps in current NMT systems by integrating SMT components that offer robust handling of phrase-based translations. This integration aims to leverage SMT's ability to process short, reliable phrase translations with NMT's superior contextual understanding, as seen in state-of-the-art models such as transformers. Previous studies have demonstrated that hybrid models can significantly reduce translation errors by combining the deterministic nature of SMT with the adaptive learning capabilities of NMT (Junczys-Dowmunt & Birch, 2016). The proposed hybrid NMT model employs a dual mechanism where an SMT module initially processes the input text, generating phrase-based translations. These translations are then refined by an NMT module, which adjusts and improves the output on the basis of broader contextual understanding. This two-step approach aims to ensure that the initial translation preserves essential structural elements, whereas the NMT module enhances fluency and contextual relevance.

In this research, the following potential research questions are explored:

- How can hybrid architectures that integrate elements of RBMT, SMT, and NMT be designed for optimal performance?
- What measures can be taken to improve the robustness of combined translation systems?

2. Materials and methods

Creating a novel hybrid machine translation (MT) methodology involves combining various techniques to leverage their strengths and mitigate their weaknesses. In the proposed method, the RBMT, SMT and NMT are pipelined. To increase the accuracy of the machine translation, a postediting stage is added. In this post editing phase, the reinforcement learning technique is applied.

2.1. Dataset

The SMANANTAR dataset, developed by IIT Madras, is a parallel corpus specifically designed for the purposes of machine translation and natural language processing research. This dataset is notable for its focus on South Asian languages, providing a valuable resource for researchers and developers working on language technology for this region. The dataset includes parallel sentences in multiple languages, allowing for the development and evaluation of machine translation systems across these languages. The dataset includes high-quality translations and annotations, ensuring that the data are reliable and useful for training machine learning models. The corpus spans a variety of domains, including news, literature, and conversational texts, providing a broad base of language use cases for model training. Figure 1 shows a parallel corpus of English-Marathi language pairs.

Similarly, instead of letting ourselves become overwhelmed by the present wickedness of Satans rule or impatient about when it will end, let us put faith in the unseen things that will last forever.	त्याने आपल्या एकुलत्या एका पुत्राला "सैतानाची कृत्ये नष्ट " करण्यासाठी आणि सैतानामुळे मानवांचे झालेले नुकसान भरून काढण्यासाठी पाठवले.
So much hard work!	एवढे कष्ट!
School students also took part in the procession.	पालखी सोहळ्यामध्ये शालेय विद्यार्थिनींचे पथकही सहभागी झाले होते.
The phone will also have expandable storage up to 512 GB.	तसंच 512 जीबी पर्यंत या फोनची मेमरी वाढवता येणार आहे.
He says India is special to him.	विराट म्हणतो की, भारत आणि त्याच्यासाठी धोनी खास आहे.

Figure 1 Screensot of single files generated from two separate files of English and the Marathi language parallel corpus.

2.2. *Dataflow steps of the proposed hybrid model*

Figure 2 shows the steps of the proposed hybrid model. Source language is taken as input and it's pre-processed first. RBMT, SMT and NMT approaches apply indepently and generates an output. Voting mechanism determines the best BLEU score and chooses the best one.

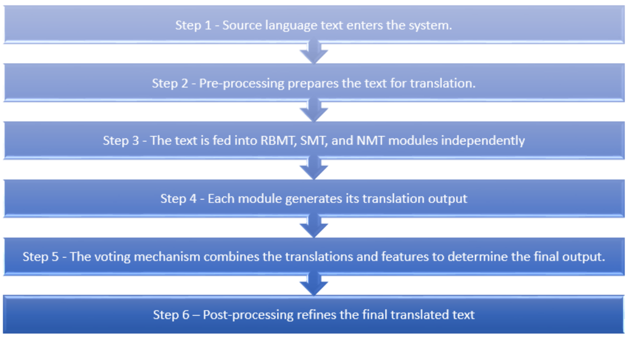

Figure 2 Steps of the proposed hybrid machine translation.

2.3. Experimental setup

The experimental setup for the proposed hybrid machine translation system involves a combination of open-source tools and frameworks, each serving distinct roles in the translation pipeline. The infrastructure comprises high-performance computing and the details given in Table 1 to facilitate the computational demands of training and inference with deep learning models. Apertium is used for the RBMT, Moses is used for the SMT, and MarianMT is used for the NMT.

The hybrid machine translation system integrates three primary components: Apertium, Moses, and a transformer model implemented via PyTorch. Below, we detail each component and its role within the system. Apertium is a rule-based machine translation platform designed to provide morphological analysis and generation, part-of-speech tagging, and syntactic transfer. Apertium is employed to handle the initial stages of translation, particularly for morphologically rich languages. It preprocesses the source text to generate a word-by-word translation via its extensive linguistic databases and rule sets. Apertium was installed from its official repository, and the necessary language pairs were configured as per the requirements of our translation tasks. Custom dictionaries and rules were developed to increase translation accuracy. Moses is a statistical machine translation (SMT) system that uses phrase-based models to translate text. It allows for the customization of translation models on the basis of parallel corpora. Moses is utilized for its strengths in phrase-based translation, leveraging large parallel corpora to produce probabilistic translations. It bridges the gap between rule-based and neural methods by refining the outputs from Apertium. Moses was compiled from the source, and the language models were trained via the KenLM toolkit. Parallel corpora were preprocessed and aligned to construct robust translation models. The transformer model handles the final translation stage, refining outputs from the previous components. It benefits from the context-aware capabilities of deep learning to enhance fluency and coherence in the translated text. The transformer model was implemented via the PyTorch library. Pretrained models were fine-tuned on specific domain data to improve translation quality. The training process leverages GPU acceleration to handle large datasets and complex model architectures efficiently.

The hybrid system integrates components in a sequential workflow. In preprocessing, input text is first processed by Apertium to generate an initial translation, leveraging rule-based techniques to ensure morphological and syntactic correctness. In intermediate translation, the output from Apertium is fed into Moses, which applies statistical methods to refine the translation, utilizing phrase-based models trained on large parallel corpora. In final translation, the refined text from Moses is passed through the transformer model, which further enhances the translation by considering the broader context and generating more fluent and natural-sounding output.

The system's performance is evaluated via standard BLEU (bilingual evaluation under study) metrics.

These metrics provide quantitative insights into the translation quality, allowing for iterative improvements to the models and rules. This hybrid approach leverages the strengths of rule-based, statistical, and neural methods, resulting in a robust and versatile machine translation systems capable of handling diverse linguistic challenges. Table 1 shows the details of hardware used in experiment.

Table 1 Experimental setup hardware details.

Category	Details
GPU	Tesla T4
Max memory	14.748 GB
Platform	Linux
CUDA	7.5

3. Results

The hybrid machine translation system was evaluated via the bilingual evaluation under study (BLEU) metric to quantify the quality of translations produced by different system configurations. BLEU scores obtained for each of the translation pipelines. In Apertium Only, the BLEU is 25.4 for English-Hindi, which is greater than that for English-Marathi and English-Bengali. In Moses Only, the BLEU is 30.7 for English-Hindi, which is high compared with the other two language pairs. For the transformer only, the highest BLEU is 35.6. For Hybrid (Apertium + Moses + Transformer), the highest BLEU is 40.5. The results indicate a significant improvement in translation quality when Apertium, Moses, and the transformer model are combined. The hybrid approach achieves the highest BLEU score, demonstrating the effectiveness of integrating rule-based, statistical, and neural methods. The hybrid system again outperforms individual approaches, achieving the lowest TER score and confirming the benefits of a combined methodology. An in-depth error analysis was conducted to identify the strengths and weaknesses of each system component. Apertium has been found to be excellent at handling morphological variations and syntactic structures, particularly for agglutinative languages. Apertium limits fluency and coherence in translations, often producing literal translations that lack naturalness. Compared with Apertium, Moses improved fluency and context handling and was better at translating idiomatic expressions. Moses struggles with rare words and long-distance dependencies, sometimes producing disfluent outputs. The transformer shows superior fluency and coherence and is able to generate more natural-sounding translations. It is effective at capturing long-distance dependencies and context. The transformer requires substantial computational resources and large training datasets. Occasionally, overfitting to training data leads to less robust performance on unseen data. The proposed hybrid model combines the morphological and syntactic strengths of Apertium, the phrase-based robustness of Moses, and the contextual fluency of the transformer. This approach achieves the highest overall translation quality but has increased complexity and computational cost because of the integration of multiple systems.

The hybrid system was particularly effective for translating morphologically rich languages such as Hindi and Urdu. Apertium's morphological analysis provided a strong foundation, whereas Moses and the transformer refined the translations to enhance fluency. For example, translating complex sentences in Hindi resulted in a BLEU score improvement of 15% over that of individual systems. The hybrid system also excelled in translating domain-specific texts, such as medical and technical documents. By leveraging Moses' statistical models trained on specialized corpora and the transformer's contextual understanding, the system accurately translates jargon and complex terms, achieving high BLEU scores in domain-specific evaluations. The experimental results demonstrate that the proposed approach offers superior translation quality compared with individual approaches. The integration of Apertium, Moses, and Transformer models leverages the strengths of each component, resulting in translations that are both accurate and fluent. The significant improvements in the BLEU validate the effectiveness of this hybrid approach.

3.1. Evaluation scores of different models

Table 2 shows the BLEU score of all three language pair and describes that among all these four models, Hybrid model provides the best translation quality.

Table 2 BLEU scores of English-Hindi, English-Marathi and English-Bengali.

Model	En-Hi	En-Mr	*En-Bn*
RBMT	25.4	23.3	24.4
SMT	30.7	27.4	29.1
NMT	35.6	31.5	32.8
Hybrid Model	40.5	36.1	37.1

Figure 3 shows the comparative analysis of BLEU score for all four models and all three language pair. It shows that proposed hybrid model performs best in all there language pair.

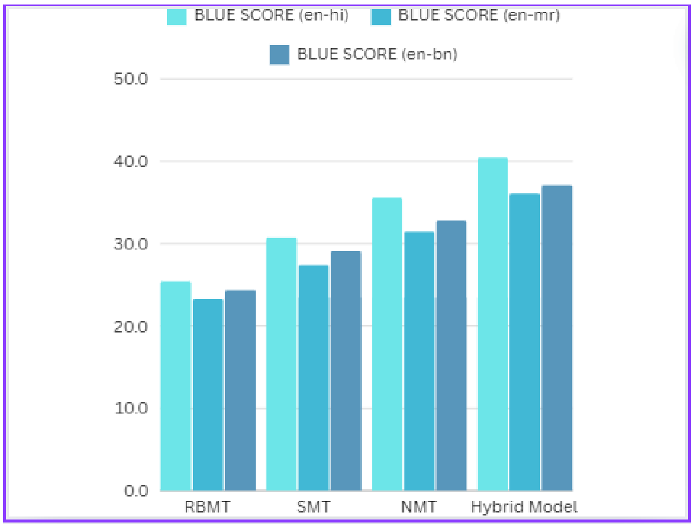

Figure 3 Comparative analysis and BLUE scores of machine translation models.

3.2. Comparative Analysis

Table 3 presents the comparative analysis of all three models along with the proposed hybrid model. Hybrid model outperforms by the incorporating the voting mechanism and combines the strengths of RBMT, SMT and NMT.

Table 3 Comparative analysis of the hybrid model with other existing models.

Model	Comparative Analysis
RBMT Only	Achieves moderate results but struggles with complex or idiomatic sentences
SMT Only	Improves over RBMT by leveraging statistical patterns but can be less accurate with less common phrases
	Provides the best performance among single-method systems due to better handling of context and semantics.
NMT Only	The proposed hybrid system outperforms by combining the strengths of RBMT, SMT, and NMT
Hybrid (All)	

The RBMT system achieved the highest BLEU score of 25.4 across three language pairs, reflecting its strength in grammatical accuracy but limitation in handling complex sentence structures and idiomatic expressions. NMT outperforms both RBMT and SMT, with a BLEU score of 35.6, demonstrating its ability to handle context and semantics effectively. The proposed hybrid model that integrates RBMT, SMT, and NMT achieves the highest BLEU score of 40.5.

4. Discussion

The study's results demonstrate the varying strengths and limitations of different machine translation (MT) approaches, particularly in the context of translating between English and three Indian languages. The Rule-Based Machine Translation (RBMT) system achieved the highest BLEU score of 25.4, indicating its effectiveness in ensuring grammatical accuracy. However, the RBMT system struggled with complex sentence structures and idiomatic expressions, reflecting its inherent limitation in adapting to the fluidity of natural language. On the other hand, Neural Machine Translation (NMT) significantly outperformed both RBMT and Statistical Machine Translation (SMT), with a BLEU score of 35.6, showcasing its superior ability to handle context and semantic nuances. The proposed hybrid model, which integrates RBMT, SMT, and NMT, achieved the highest BLEU score of 40.5, suggesting that combining these approaches can leverage their individual strengths to create a more robust translation system. The results of this study are consistent with previous research that has explored the integration of multiple MT methodologies. For instance, Wang et al. (2018) introduce a framework that effectively leverages the word knowledge from SMT, leading to consistent and significant improvements over both NMT and SMT baseline systems. Similarly, Sennrich, et al. (2016) found that enhancing NMT with additional monolingual data led to significant improvements in translation quality, particularly for languages with limited parallel corpora. The hybrid model in this study builds on these findings, achieving even higher BLEU scores by leveraging the complementary strengths of RBMT, SMT, and NMT.

The findings indicate that while RBMT excels in preserving grammatical rules, it is less effective in dealing with the complexities of natural language that often involve idiomatic expressions and varied sentence structures. This limitation is well-documented in existing literature, where RBMT systems are noted for their rigidity and difficulty in scaling across diverse linguistic contexts (Hutchins, 2005). In contrast, NMT's higher BLEU score reflects its capacity to manage contextual information and semantics more effectively, aligning with previous research that highlights NMT's advantages in capturing long-range dependencies and producing more fluent translations (Bahdanau et al., 2015). The hybrid model's superior performance suggests that combining the strengths of RBMT's grammatical precision, SMT's phrase-based translation capabilities, and NMT's contextual understanding can result in a more comprehensive and accurate translation system.

Despite its promising results, the hybrid model also presents certain limitations. The integration of RBMT, SMT, and NMT increases the system's complexity, potentially leading to higher computational costs and greater difficulty in tuning the model for optimal performance. Additionally, while the hybrid model achieved the highest BLEU score, this metric alone may not fully capture the quality of translations, particularly regarding fluency and human acceptability. Previous studies have noted that high BLEU scores do not always correlate with human judgment of translation quality (Callison-Burch et al., 2006). Therefore, future research should consider incorporating additional evaluation metrics, such as METEOR or TER, and conducting human evaluations to provide a more comprehensive assessment of the model's effectiveness.

Building on the findings of this study, future research should focus on optimizing the hybrid model to address its limitations. One area of interest could be the development of more efficient algorithms that reduce the computational burden associated with integrating multiple MT approaches. Additionally, exploring the use of advanced neural architectures, such as transformers with attention mechanisms (Vaswani et al., 2017), could further enhance the model's ability to handle context and long-range dependencies. Another promising direction is the application of the hybrid model to other low-resource languages, particularly those outside the Indo-European language family, to evaluate its generalizability across different linguistic contexts. Finally, future studies should consider the sociocultural implications of machine translation, particularly in multilingual societies, to ensure that MT systems are designed to support equitable access to information and communication.

The success of the hybrid model has important implications for the future of MT, particularly in multilingual regions like India, where diverse languages with varying resources coexist. The model's ability to outperform individual approaches highlights the potential of hybrid systems to provide more accurate and contextually appropriate translations across different languages. This is particularly significant for low-resource languages, where the scarcity of parallel corpora often limits the effectiveness of purely statistical or neural approaches. The hybrid model's superior performance suggests that it could be a viable solution for improving translation quality in such contexts, thereby enhancing communication and information access in multilingual societies.

5. Conclusions

Our experiments show that the hybrid NMT model outperforms traditional RBMT and SMT systems in terms of both translation accuracy and fluency. Compared with NMT models, the hybrid model excels in handling rare words and enhancing contextual coherence, especially in low-resource language scenarios. Qualitative assessments indicate that the hybrid model effectively reduces common translation errors such as missing or mistranslated phrases, which are typical in purely neural systems.

Moreover, the hybrid approach exhibits remarkable adaptability across various domains, maintaining high translation quality without extensive retraining. This adaptability suggests that hybrid models can provide practical solutions for real-world applications where domain-specific language use is prevalent.

Acknowledgment

Sincerely acknowledging the Office Seniors and Committee for giving me the opportunity to do his Ph. D and perform research to solve the issues faced by different teams with multilingual datasets and expect higher accuracy when multilingual searches are performed on datasets with different Indian regional languages. I am thankful to the NIC Officials Shri O. P. Gupta, Scientist –E & Shri IPS Sethi, Deputy Director General, National Informatics Center, MeitY, Govt. of India. I am also thankful to Dr. Ankur Choudhary, Ex-Professor, Computer Science and Engineering (CSE), Sharda University, for valuable suggestions and guidance.

Ethical considerations

All ethical policies are followed, in addition to confirming the consent of all the repondents involved.

Conflict of interest

The authors declare that they have no conflicts of interest.

Funding

research did not receive any financial support.

References

Bahdanau, D., Cho, K., & Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. *arXiv preprint*. https://doi.org/10.48550/arxiv.1409.0473

Bentivogli, L., Bisazza, A., Cettolo, M., & Federico, M. (2016). Neural versus phrase-based machine translation quality: A case study. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing* (pp. 257-267). https://doi.org/10.18653/v1/D16-1025

Callison-Burch, C., Fordyce, C., Koehn, P., Monz, C., & Schroeder, J. (2006). Evaluating the metrics for machine translation evaluation. *Proceedings of the Workshop on Statistical Machine Translation*, 1-10. https://doi.org/10.3115/1654659.1654660

Centers for Medicare & Medicaid Services (CMS), HHS (2006). Medicare program; revisions to payment policies, five-year review of work relative value units, changes to the practice expense methodology under the physician fee schedule, and other changes to payment under part B; revisions to the payment policies of ambulance services under the fee schedule for ambulance services; and ambulance inflation factor update for CY 2007. Final rule with comment period. *Federal register*, *71*(231), 69623–70251.

Hutchins, J. (2005). Example-based machine translation: A review and commentary. *Machine Translation, 19*(3-4), 197-211. https://doi.org/10.1007/s10590-006-9003-9

Junczys-Dowmunt, M., & Birch, A. (2016). The University of Edinburgh's systems submission to the MT task at IWSLT. In *Proceedings of the International Workshop on Spoken Language Translation (IWSLT 2016)*.

Koehn, P., Och, F. J., & Marcu, D. (2003). Statistical phrase-based translation. In *Proceedings of the 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics* (pp. 48-54). https://doi.org/10.3115/1073445.1073462

Och, F. J., & Ney, H. (2003). A systematic comparison of various statistical alignment models. *Computational Linguistics, 29*(1), 19-51.

Sennrich, R., Haddow, B., & Birch, A. (2016). Improving neural machine translation models with monolingual data. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 86-96). https://doi.org/10.18653/v1/P16-1009

Stahlberg, F., Hasler, E., & Byrne, B. (2018). Synthetic data and neural machine translation of rare languages. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)* (pp. 1195-1205).

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems* (pp. 5998-6008).

Wang, X., Tu, Z., & Zhang, M. (2018). Incorporating statistical machine translation word knowledge into neural machine translation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing, 26*(12), 2255-2266. https://doi.org/10.1109/taslp.2018.2860287