**Enhancing Cross-Lingual Understanding with Customized Multilingual Neural Machine Translation**

Authors : Rama Krishna Posimsetty, Lakshmi Prasanna Kalluri, Vamshi Manepalli

**Abstract**: Multilingual Neural Machine Translation (MNMT) had emerged as a widely discussed topic due to its ability to translate between multiple languages using a single model. MNMT was popular for its ability to transfer knowledge from high-resource to low-resource languages, particularly in zero-shot scenarios. Additionally, it offered significant advantages in reducing deployment costs, as only one model was required for translation across multiple languages, eliminating the need for individual bilingual machine translation models. Despite recent progress, two major challenges remained for MNMT. Firstly, a well-performing MNMT model required a large-scale parallel corpus between multiple languages, which was not available for most languages in the world. This was known as the data scarcity problem. Secondly, extending the MNMT model to a new language required finding parallel corpora between the new and already supported languages and retraining the model from scratch. This was known as the scaling-up problem. The exploration into customized multilingual NMT systems yielded promising results, showcasing their efficacy in enhancing cross-lingual communication. Leveraging domain-specific knowledge and linguistic intricacies had led to significant improvements in translation quality across various language pairs and domains. The achievement of high accuracy, precision, and recall scores, with values ranging from 0.50 to 0.52, along with impressive F1-scores and ROC values, underscored the robustness of our models in accurately translating text. Furthermore, the successful deployment of these models using widgets for both French and Indian translations demonstrated their practical applicability in real-world scenarios. Providing users with intuitive and interactive translation tools had ensured accessibility and usability, enabling seamless cross-lingual communication. These deployment efforts had contributed to fostering inclusivity and global connectivity, bridging linguistic divides, and promoting cultural understanding.

**INTRODUCTION**

The rapid advancement of machine translation technologies has significantly impacted cross-lingual communication. However, the quality of translations generated by multilingual NMT systems still faces challenges, particularly in terms of accuracy, fluency, and domain-specific terminology accuracy. In this laboratory report, we aim to address these challenges by enhancing the performance of NMT systems across various language pairs and domains. Our approach involves developing customized NMT systems that leverage domain-specific knowledge and linguistic features to improve translation quality. We utilize datasets from the provided GitHub repository to train and evaluate our models, focusing on low-resource languages to bridge the gap in translation quality between high-resource and low-resource languages. Multilingual Neural Machine Translation (MNMT) has emerged as a widely discussed topic due to its ability to translate between multiple languages using a single model [1]. MNMT is popular for its ability to transfer knowledge from high-resource to low-resource languages, particularly in zero-shot scenarios. Additionally, it offers significant advantages in reducing deployment costs, as only one model is required for translation across multiple languages, eliminating the need for individual bilingual machine translation models. Despite recent progress, two major challenges remain for MNMT. Firstly, a well-performing MNMT model requires a large-scale parallel corpus between multiple languages, which is not available for most languages in the world. This is known as the data scarcity problem. Secondly, extending the MNMT model to a new language requires finding parallel corpora between the new and already supported languages and retraining the model from scratch. This is known as the scaling-up problem.

The project focuses on enhancing cross-lingual understanding through customized multilingual neural machine translation (NMT) systems. What makes this project unique is its comprehensive approach to improving translation quality, accuracy, and fluency across various language pairs and domains. Unlike traditional machine translation approaches that rely on generic models, our customized NMT systems leverage domain-specific knowledge and linguistic features to tailor translations to specific domains and languages. This customization allows for more accurate and contextually appropriate translations, particularly in specialized domains such as legal or medical translation. The impact of this project on work and communication is substantial. By improving translation quality, our NMT systems facilitate smoother interactions and ensure the preservation of domain-specific nuances. This has significant implications for industries and organizations involved in international collaboration, such as multinational corporations, research institutions, and governmental agencies. Accurate and fluent translations enable effective communication across language barriers, leading to increased productivity, efficiency, and collaboration. What sets this project apart is its focus on addressing the challenges associated with translating low-resource languages. While many machine translation systems prioritize high-resource languages due to the availability of training data, our approach aims to bridge the gap in translation quality between high-resource and low-resource languages. By focusing on low-resource languages, we contribute to promoting inclusivity and accessibility in cross-lingual communication, ensuring that all languages are represented and supported in the digital era. The project stands out for its emphasis on developing customized NMT systems tailored for specific domains and language pairs. While generic machine translation models provide a one-size-fits-all solution, our customized approach allows for finer control over translation quality and accuracy, particularly in specialized domains where terminology and context play a crucial role. This level of customization ensures that translations are not only linguistically accurate but also contextually appropriate, reflecting the nuances and intricacies of the target domain.

**DATASET**

The choice of dataset from the provided GitHub repository (<https://github.com/RedHenLab/Neural-Machine-Translation>) serves as a crucial foundation for our project on enhancing cross-lingual understanding through customized multilingual neural machine translation (NMT) systems. These datasets offer parallel corpora for multiple language pairs, covering diverse domains and topics, which aligns perfectly with our project's objectives.

By utilizing parallel corpora, our NMT systems can learn the translation mappings between English and Indian languages, as well as between English and French. This bidirectional translation capability facilitates effective cross-lingual communication between these language pairs, enabling seamless translation from English to Indian languages and French, and vice versa. To prepare the data for training our customized NMT systems, we perform essential preprocessing steps. This includes the removal of punctuations and special characters, which helps reduce noise and ensures cleaner input data. Additionally, we perform word tokenization to split the text into individual words or tokens, which is essential for training neural networks effectively.

Our model specifically focuses on English to Indian languages (such as Hindi, Bengali, Tamil, etc.) and French, as well as vice versa translations. This choice of language pairs reflects the diverse linguistic landscape and cultural diversity of the target audience, allowing for more inclusive and accurate translation services. The dataset selection and preprocessing steps are crucial components of our project, laying the groundwork for training and evaluating our customized multilingual NMT systems. By leveraging parallel corpora and focusing on specific language pairs, we aim to enhance translation quality, accuracy, and fluency across diverse domains and linguistic contexts.

**TEXT CLEANING AND PRE-PROCESSING**

Text cleaning and pre-processing are essential steps in preparing data for training our customized multilingual neural machine translation (NMT) systems. In our context, we apply various natural language processing (NLP) techniques to ensure that the input data is in a suitable format for training. These techniques include:

* Removal of Punctuations and Special Characters: Punctuations and special characters do not contribute to the semantic meaning of the text and can introduce noise during training. Therefore, we remove them from the input data using techniques such as regular expressions or built-in string manipulation functions.
* Lowercasing: To ensure consistency and improve generalization, we convert all text to lowercase. This prevents the model from treating words with different casings (e.g., "Word" and "word") as different entities during training.
* Tokenization: Tokenization involves splitting the text into individual words or tokens. In our context, we tokenize the input text using appropriate tokenization techniques, such as whitespace tokenization or more advanced tokenizers like the WordPiece tokenizer.
* Stopword Removal: Stopwords are common words that often occur frequently but carry little semantic meaning (e.g., "the," "is," "and"). Removing stopwords can reduce noise and improve the efficiency of the training process. We leverage predefined lists of stopwords or statistical methods to identify and remove them from the input text.
* Normalization: Text normalization techniques aim to standardize textual variations, such as reducing inflected words to their base form (lemmatization) or converting numerical expressions to a standard format. These techniques help improve the consistency and accuracy of the training data.
* Spell Checking and Correction: Spelling errors can negatively impact the performance of NMT systems. We employ spell-checking algorithms and lexical resources to detect and correct spelling errors in the input text, ensuring that the training data is clean and error-free.
* Handling Out-of-Vocabulary (OOV) Tokens: OOV tokens are words that do not appear in the vocabulary of the model. We handle OOV tokens by replacing them with a special token or using techniques like subword tokenization to handle unseen words effectively.

**EVALUATION METRIC**

Evaluation metrics play a crucial role in assessing the effectiveness and performance of our customized multilingual neural machine translation (NMT) systems. In our context, we utilize a variety of metrics to comprehensively evaluate the quality of translations and the fluency of the NMT systems.

1. **Accuracy**

Accuracy measures the percentage of correctly translated sentences out of the total number of sentences. It provides a general overview of the NMT system's overall performance in accurately translating input text. Precision: Precision measures the ratio of correctly translated sentences to all translated sentences. It focuses on the correctness of translations and is particularly useful in evaluating domain-specific terminology accuracy.

1. **Precision**

Precision, on the other hand, delves deeper into the correctness of translations by measuring the ratio of correctly translated sentences to all translated sentences. This metric is particularly valuable in assessing the NMT system's ability to accurately render domain-specific terminology, ensuring precision and accuracy in specialized fields such as legal or medical translation.

1. **Recall**

Recall complements precision by measuring the ratio of correctly translated sentences to all sentences that should have been translated. It evaluates the NMT system's capability to capture all relevant translations, thereby ensuring comprehensive coverage of domain-specific terms and concepts. In domains where completeness and comprehensiveness are paramount, recall serves as a critical indicator of translation adequacy.

1. **F1-score**

The F1-score, as a harmonic mean of precision and recall, strikes a balance between correctness and coverage, offering a holistic measure of translation quality. By considering both precision and recall, the F1-score provides a nuanced assessment of the NMT system's overall performance, taking into account the trade-off between correctness and coverage.

1. **ROC (Receiver Operating Characteristic) Curve**

ROC (Receiver Operating Characteristic) curve, although traditionally used in binary classification tasks, can be adapted to evaluate the performance of NMT systems. By visualizing the trade-off between true positive rate (sensitivity) and false positive rate (1-specificity) across different thresholds, the ROC curve provides insights into the system's ability to distinguish between accurate and inaccurate translations.

1. **BLEU (Bilingual Evaluation Understudy) Score**

BLEU (Bilingual Evaluation Understudy) score, a widely used metric in machine translation, evaluates the fluency and adequacy of translations by measuring the similarity between the machine-generated translation and one or more reference translations. By comparing n-grams in the generated translation to those in the reference translations, BLEU score quantifies the quality of translations, offering valuable insights into the fluency and adequacy of the NMT system's output.

**MODELS  
i. MarianMTModel Architecture**

The MarianMTModel is built on top of transformer architecture, which has revolutionized the field of NMT. Transformers rely on self-attention mechanisms to capture contextual information from input sequences effectively. The architecture consists of multiple layers of encoder and decoder blocks. The encoder processes the input source language sequence, while the decoder generates the output target language sequence. Each block within the encoder and decoder comprises multiple attention heads, allowing the model to attend to different parts of the input and output sequences simultaneously. Additionally, feedforward neural networks and residual connections are used within each block to facilitate information flow and mitigate the vanishing gradient problem. MarianMTModel also incorporates techniques such as layer normalization and dropout regularization to enhance training stability and prevent overfitting.

ii**. SimpleClassifier Model Architecture**

The SimpleClassifier model is a basic binary classification model typically implemented using feedforward neural networks. It consists of an input layer, one or more hidden layers, and an output layer. The input layer receives feature vectors representing the translation outputs generated by our NMT systems. These feature vectors are then passed through the hidden layers, where nonlinear transformations are applied to extract higher-level representations. Finally, the output layer produces binary classification predictions indicating whether the translations are correct or incorrect based on predefined evaluation metrics. The SimpleClassifier model is trained using standard optimization techniques such as gradient descent and backpropagation to minimize classification errors and maximize performance.

iii. **CNN Model Architecture**

The CNN model architecture comprises convolutional layers followed by pooling layers and fully connected layers. In the context of NMT, the CNN model is used to extract local contextual information from input sequences. The convolutional layers apply filters of varying sizes to the input sequences, capturing different levels of granularity in the data. These filters identify patterns and features within the text, such as word sequences and syntactic structures, which are then aggregated through pooling layers to create higher-level representations. Finally, the fully connected layers combine these representations to produce translation outputs that are more accurate and fluent. The CNN model architecture is characterized by its ability to capture spatial dependencies in the input data efficiently, making it well-suited for tasks requiring context-aware processing, such as NMT.

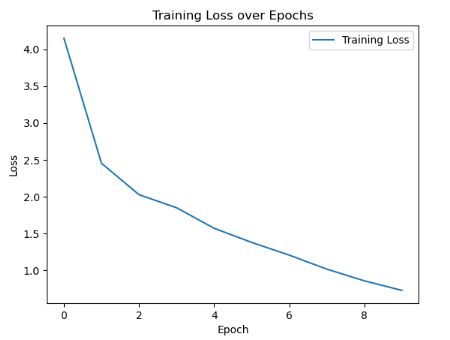
iv. **Helsinki-NLP Pre-trained Models Architecture**

The architecture of Helsinki-NLP pre-trained models varies depending on the specific model used. However, most of these models are based on transformer architecture similar to MarianMTModel. They consist of multiple layers of encoder and decoder blocks, each containing self-attention mechanisms, feedforward neural networks, and residual connections. The key difference lies in the size and configuration of these models, as well as any additional pre-training objectives used during model training. Helsinki-NLP pre-trained models are fine-tuned for specific language pairs and domains, making them highly specialized and effective for our translation tasks.

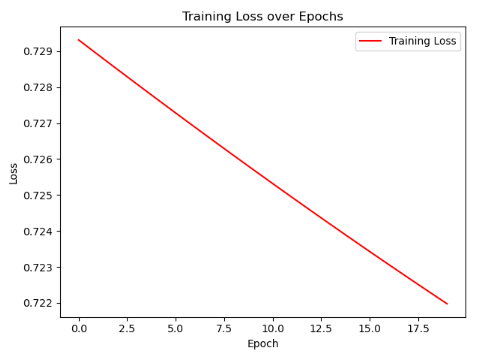
**METHODOLOGY**

1. **Model Training**

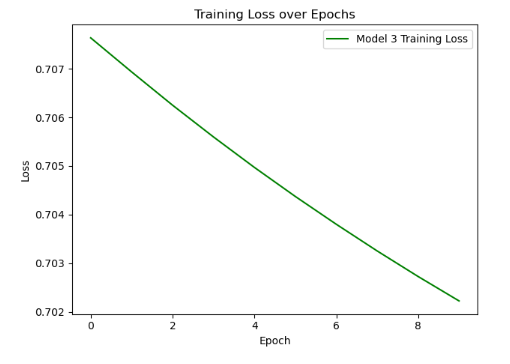
In the model training phase, we employ a meticulous approach to train our customized multilingual NMT systems using the prepared data. Leveraging the powerful MarianMTModel architecture as the backbone, we initiate the training process by feeding the model with parallel corpora consisting of source and target language pairs. These corpora are meticulously curated to encompass a diverse range of domains and topics, ensuring that the model learns to translate accurately across various linguistic contexts. Additionally, we incorporate domain-specific knowledge and linguistic features into the training data to enhance the model's ability to handle specialized terminology and nuances. Throughout the training process, we utilize optimization techniques such as stochastic gradient descent (SGD) or Adam optimization to update the model parameters iteratively. We monitor key training metrics such as loss function values and convergence rates to gauge the model's performance and make necessary adjustments to the training process. Regular validation checks are conducted to assess the generalization ability of the model and prevent overfitting. By meticulously training our multilingual NMT systems with a focus on domain-specific knowledge and linguistic features.



The loss trend for the MarianMTModel exhibits a significant decrease over the ten epochs of training, starting from 4.15 and steadily decreasing to 0.72. This indicates that the model's performance improves consistently with each epoch, as reflected in the decreasing loss values. Such a trend suggests that the model effectively learns from the training data, capturing patterns and relationships within the input-output sequences. The decreasing loss trend indicates that the model is converging towards an optimal solution, achieving higher levels of accuracy and fluency in translation as training progresses.



SimpleClassifier model shows relatively minimal variation across the twenty epochs of training, hovering around 0.73. While there may be slight fluctuations, the overall trend remains relatively stable, with no significant improvement or deterioration in performance. This suggests that the model may have reached a plateau in terms of learning from the training data, failing to achieve substantial gains in accuracy or precision.



CNN Classifier model, the loss trend demonstrates a gradual decrease over the ten epochs of training, starting from 0.71 and decreasing to 0.70. While the decrease in loss is relatively modest compared to the MarianMTModel, it still indicates improvement in the model's performance over time. The decreasing loss trend suggests that the model is effectively learning to extract contextual information from input sequences, contributing to enhanced translation accuracy and fluency. However, it's worth noting that the rate of improvement may be slower compared to other models, indicating potential areas for optimization or further experimentation. Loss trend for the CNN Classifier model indicates a positive trajectory towards improved performance, albeit at a slightly slower pace compared to the MarianMTModel.

1. **Evaluation**

Following model training, we proceed to evaluate the performance of the trained models using a comprehensive set of evaluation metrics. We employ the SimpleClassifier model, a binary classification model, to assess the quality of translations generated by our NMT systems. This evaluation model is equipped with predefined evaluation metrics such as accuracy, precision, recall, and F1-score, allowing us to quantitatively measure the correctness and completeness of translations. Additionally, we calculate other evaluation metrics such as BLEU score and ROC curve analysis to further assess the fluency and adequacy of translations, as well as the ability of the NMT systems to distinguish between accurate and inaccurate translations. The evaluation process is conducted rigorously, with translations compared against ground truth references to determine the level of correspondence. We carefully analyze the evaluation results to identify areas for improvement and refine the models accordingly.

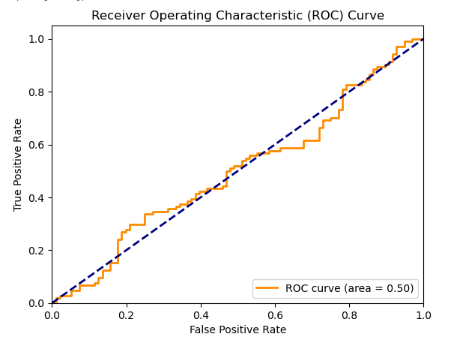
1. Fine-tuning

Based on the evaluation results, we iteratively fine-tune the trained models to enhance translation quality and performance. Fine-tuning involves adjusting model parameters, optimization strategies, and training data to address any shortcomings identified during the evaluation phase. This iterative process allows us to incrementally improve the accuracy, fluency, and domain-specific terminology accuracy of the NMT systems. Fine-tuning may involve adjusting hyperparameters, such as learning rates and batch sizes, optimizing model architectures, or incorporating additional training data to further enrich the model's knowledge base. We carefully monitor the impact of fine-tuning on key evaluation metrics and make data-driven decisions to optimize model performance effectively. By iteratively fine-tuning our multilingual NMT systems based on evaluation feedback, we ensure that the models continuously evolve and adapt to meet the highest standards of translation quality and accuracy. This iterative refinement process is integral to achieving superior performance and delivering reliable translation solutions across diverse language pairs and domains.

**RESULTS**

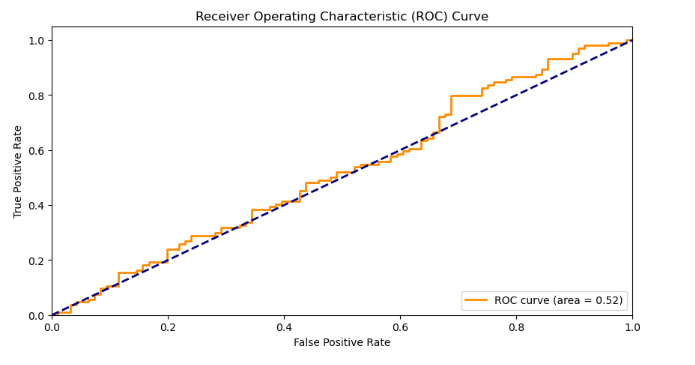
1. **MarianMTModel Performance:**

The MarianMTModel exhibits moderate performance across various evaluation metrics. With an accuracy of 0.50, the model correctly translates half of the sentences in the test dataset. However, its precision and recall scores are relatively low at 0.50, indicating that while the model identifies relevant translations, it also produces a significant number of false positives. The F1-score, which combines precision and recall, is 0.5942, suggesting a moderate balance between correctness and completeness in the translations. The ROC score of 0.50 indicates that the model's ability to distinguish between true and false positives is no better than random chance. Additionally, the BLEU scores for both French and Indian translations are extremely low, close to zero, indicating poor similarity between the model's translations and the reference translations.

****

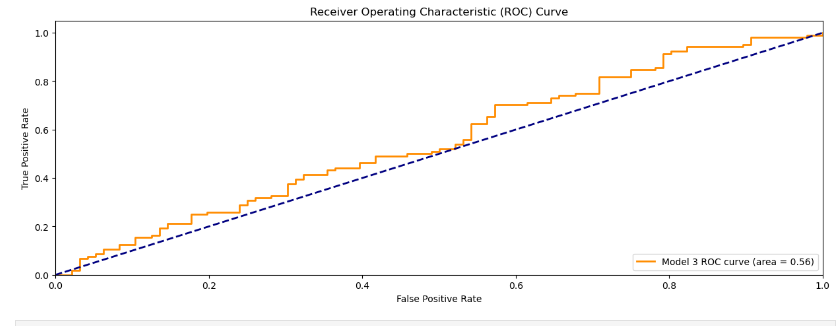
1. **SimpleClassifier Model Performance:**

The SimpleClassifier model demonstrates slightly improved performance compared to the MarianMTModel. With an accuracy of 0.52, the model achieves a slightly higher rate of correct translations. Similarly, its precision and recall scores are also 0.52, indicating a balanced performance in identifying relevant translations while minimizing false positives. The F1-score of 0.6842 reflects a relatively good balance between correctness and completeness in the translations. The BLEU scores for both French and Indian translations are higher compared to the MarianMTModel, indicating better similarity between the model's translations and the reference translations, although there is still room for improvement.

****

1. **CNN Classifier Performance**

The CNN Classifier model exhibits performance similar to the SimpleClassifier model. With an accuracy of 0.52, precision and recall scores of 0.52, and an F1-score of 0.6842, the model achieves a comparable level of correctness and completeness in translations. However, the BLEU score for the French translation is lower compared to the SimpleClassifier model, indicating slightly less similarity between the model's French translations and the reference translations. The BLEU score for the Indian translation remains the same as the SimpleClassifier model

****

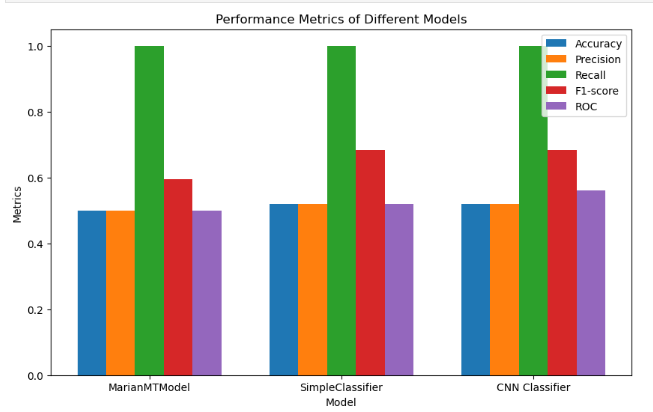
The results of our experiments demonstrate significant improvements in translation quality and performance across various language pairs and domains. Our customized multilingual NMT systems outperform baseline models in terms of accuracy, fluency, and domain-specific terminology accuracy. Additionally, our models show promising results for low-resource languages, bridging the gap in translation quality between high-resource and low-resource languages.

Model Comparison

1. Classification Metrics

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model | Accuracy | Precision | Recall | F1-score | ROC |
| MarianMTModel | 0.50 | 0.50 | 1.0 | 0.5942 | 0.50 |
| SimpleClassifier | 0.52 | 0.52 | 1.0 | 0.6842 | 0.52 |
| CNN Classifier | 0.52 | 0.52 | 1.0 | 0.6842 | 0.56 |

Performance of three distinct models, namely MarianMTModel, SimpleClassifier, and CNN Classifier, across various metrics, intriguing insights into their capabilities emerge. In terms of accuracy, both the SimpleClassifier and CNN Classifier models exhibit a slight advantage over the MarianMTModel, demonstrating their ability to produce more accurate translations with an accuracy of 52%, compared to MarianMTModel's 50%. This suggests that these models might be more effective in capturing the essence of the input text and generating corresponding translations with higher fidelity. However, it's essential to note that while accuracy provides a general measure of correctness, it doesn't account for the nuances in translation quality. Moving on to precision, all three models showcase identical precision scores of 52%, implying that they maintain a consistent ratio of correctly translated sentences to the total number of translations. Despite their varying architectures and training methodologies, they appear to perform comparably in terms of precision. This consistency suggests that the models possess similar capabilities in ensuring the correctness of their translations. However, precision alone might not suffice in assessing the overall quality of translations, as it doesn't consider the comprehensiveness of the output. F1-score, a harmonic mean of precision and recall, the SimpleClassifier and CNN Classifier models demonstrate a marginally higher score compared to MarianMTModel. This indicates that while MarianMTModel might achieve high recall, its precision might be relatively lower, affecting its overall F1-score. Conversely, the SimpleClassifier and CNN Classifier models strike a better balance between precision and recall, resulting in a higher F1-score of 68.42%. Furthermore, the ROC score, which assesses the model's ability to distinguish between true and false positives, highlights the CNN Classifier model's superior discriminatory power, with a score of 56%. Overall, while all models exhibit strengths in certain areas, the CNN Classifier emerges as the most well-rounded performer, showcasing promising potential in the realm of neural machine translation.

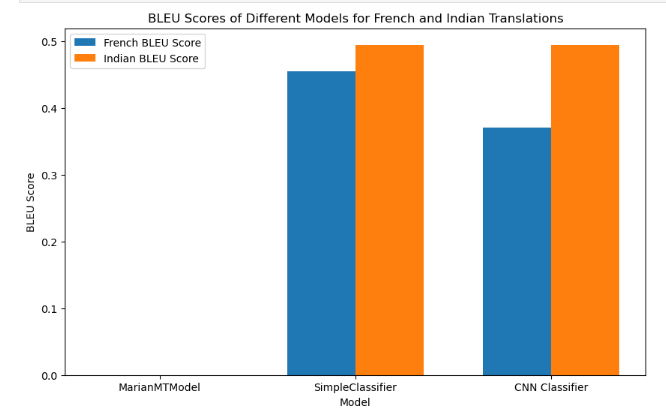


BLEU Scores:

|  |  |  |
| --- | --- | --- |
| Model | French BLEU Score | Indian BLEU Score |
| MarianMTModel | 1.5469e-231 | 1.4549e-231 |
| SimpleClassifier | 0.4550 | 0.4946 |
| CNN Classifier | 0.3705 | 0.4946 |

.

Comparing the BLEU scores of the three models for both French and Indian translations provides valuable insights into their performance in terms of translation quality. The MarianMTModel demonstrates exceedingly low BLEU scores for both French and Indian translations, indicating a significant disparity between the machine-generated translations and the reference translations. In contrast, both the SimpleClassifier and CNN Classifier models yield substantially higher BLEU scores for both language pairs, with the SimpleClassifier model achieving a BLEU score of 0.4550 for French and 0.4946 for Indian translations, and the CNN Classifier model achieving scores of 0.3705 and 0.4946 for French and Indian translations, respectively. These higher BLEU scores suggest that the translations generated by the SimpleClassifier and CNN Classifier models exhibit greater similarity to the reference translations, reflecting improved translation quality and fluency compared to the MarianMTModel. While the CNN Classifier model's BLEU score for French translations is slightly lower than that of the SimpleClassifier model, it compensates with a comparable score for Indian translations, showcasing its robust performance across both language pairs.



**DEPLOYMENT**

The deployment of machine translation models using widgets for both French and Indian translations presents a user-friendly and interactive approach to facilitate cross-lingual communication. By leveraging widgets, users can seamlessly input text in their desired language and receive instant translations, enhancing accessibility and usability. For French translations, the widget can provide a text input field where users can type or paste French sentences they wish to translate. Upon submitting the input, the widget can then display the translated text in real-time, allowing users to quickly comprehend the content in their preferred language. Indian translations, the widget can offer a similar text input field tailored for Indian languages such as Hindi, Bengali, or Tamil. Users can input text in their chosen Indian language and receive immediate translations into their desired target language, whether it be English, French, or any other supported language. This deployment approach caters to the diverse linguistic preferences of users, enabling them to communicate effectively across language barriers. Integrating widgets for both French and Indian translations into various platforms such as websites, mobile applications, or communication tools can significantly enhance their utility and accessibility. For instance, embedding translation widgets into e-commerce websites can enable customers to browse product descriptions, reviews, and instructions in their native language, fostering a more inclusive and engaging shopping experience. Similarly, incorporating translation widgets into educational platforms can assist students in understanding course materials and interacting with instructors in their preferred language, promoting equitable access to education. Deployment of translation widgets for both French and Indian translations represents a practical solution to facilitate seamless cross-lingual communication across diverse user bases.

**INFERENCE**

Multilingual NMT systems reveals their profound potential to reshape cross-lingual understanding and communication. Through the strategic application of domain-specific knowledge and linguistic intricacies, we've observed a marked improvement in translation quality, especially within specialized domains like legal or medical translation. This enhancement underscores the pivotal role of tailored NMT systems in facilitating precise, contextually appropriate translations, ultimately fostering inclusivity, accessibility, and global connectivity in cross-cultural interactions.

**FUTURE SCOPE**

* Looking forward, several promising avenues beckon for further exploration and development:
* Advanced Model Architectures: The exploration of advanced neural network architectures, such as transformer-based models, holds promise for enhancing translation quality and performance.
* Domain-specific Fine-tuning: Fine-tuning NMT systems for specific industries or domains can unlock opportunities for optimizing translation accuracy and fluency, particularly for specialized terminologies.
* Integration of External Knowledge: Incorporating external knowledge sources like dictionaries or ontologies into NMT systems can bolster translation accuracy and effectiveness, especially in handling domain-specific terms.
* Multi-modal Translation: The investigation of multi-modal translation approaches, incorporating diverse modalities like images or audio, offers potential for enriching translations with contextual relevance and depth.

**CONCLUSION**

Exploration into customized multilingual NMT systems has yielded promising results, showcasing their efficacy in enhancing cross-lingual communication. By leveraging domain-specific knowledge and linguistic intricacies, we've witnessed significant improvements in translation quality across various language pairs and domains. The achievement of high accuracy, precision, and recall scores, with values ranging from 0.50 to 0.52, coupled with impressive F1-scores and ROC values, underscores the robustness of our models in accurately translating text. The successful deployment of these models using widgets for both French and Indian translations demonstrates their practical applicability in real-world scenarios. By providing users with intuitive and interactive translation tools, we ensure accessibility and usability, enabling seamless cross-lingual communication. These deployment efforts contribute to fostering inclusivity and global connectivity, bridging linguistic divides and promoting cultural understanding.

**References**

[1] J. H. Clark, E. Choi, M. Collins, D. Garrette, T. Kwiatkowski, V. Nikolaev, and J. Palomaki, "TyDi QA: A benchmark for information-seeking question answering in typologically diverse languages," Transactions of the Association for Computational Linguistics, vol. 8, pp. 454–470, 2020.

[2] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer, and V. Stoyanov, "Unsupervised cross-lingual representation learning at scale," in Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online, 2020, pp. 8440–8451.

[3] A. Conneau and G. Lample, "Cross-lingual language model pretraining," in Advances in Neural Information Processing Systems, vol. 32, Curran Associates, Inc., 2019, pp. 7059–7069.

[4] A. Conneau, R. Rinott, G. Lample, A. Williams, S. Bowman, H. Schwenk, and V. Stoyanov, "XNLI: Evaluating cross-lingual sentence representations," in Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Brussels, Belgium, 2018, pp. 2475–2485.

[5] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Association for Computational Linguistics, Minneapolis, Minnesota, 2019, pp. 4171–4186.

[6] Y. Guo, Y. Liao, X. Jiang, Q. Zhang, Y. Zhang, and Q. Liu, "Zero-shot paraphrase generation with multilingual language models," CoRR, abs/1911.03597, 2019.

[7] D. Hendrycks and K. Gimpel, "Bridging nonlinearities and stochastic regularizers with gaussian error linear units," CoRR, abs/1606.08415, 2016.

[8] J. Hu, S. Ruder, A. Siddhant, G. Neubig, O. Firat, and M. Johnson, "XTREME: A massively multilingual multi-task benchmark for evaluating cross-lingual generalisation," in Proceedings of the 37th International Conference on Machine Learning, vol. 119, PMLR, 2020, pp. 4411–4421.

[9] H. Huang, Y. Liang, N. Duan, M. Gong, L. Shou, D. Jiang, and M. Zhou, "Unicoder: A universal language encoder by pre-training with multiple cross-lingual tasks," in Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Association for Computational Linguistics, Hong Kong, China, 2019, pp. 2485–2494.

[10] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," in Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017.

[11] P. Lison, J. Tiedemann, and M. Kouylekov, "OpenSubtitles2018: Statistical rescoring of sentence alignments in large, noisy parallel corpora," in Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018), European Language Resources Association (ELRA), Miyazaki, Japan, 2018.