ENGLISH TO MALAYALAM TRANSLATION
Using RNN,Sequence-to-Sequence, and Self-Attention, BERT Technique models

Pampana Charmitha[1], Dr. Surendra reddy Vinta[2]
[1,2] School of Computer Science and Engineering (SCOPE), VIT-AP University, Amaravati,Andhra Pradesh
charmitha.21bce7295@vitapstudent,ac.in [2]vsurendra.cse@gmail.com

Abstract:
Due to the syntax, grammar, and vocabulary variations between English and Malayalam, translating between the two languages can be difficult. Machine translation has recently seen encouraging results from deep learning approaches including Recurrent Neural Networks (RNNs), Sequence-to-Sequence (Seq2Seq) models, and Self-Attention mechanisms. In this essay, the investigate is done for the use of these strategies in translating from English to Malayalam. suggest a neural machine translation (NMT) model that includes a self-attention mechanism, a Seq2Seq decoder with attention mechanism, and a bidirectional RNN encoder to enhance the model's capacity to recognize significant characteristics and long-term dependencies in the input sequence. Using industry-accepted assessment measures like BLEU, METEOR, and TER a benchmark English to Malayalam dataset,the proposed model and compare it to other cutting-edge models.With a BLEU score and results demonstrate that the suggested model beats existing models in terms of translation quality.The area of machine translation has seen a revolution because to the usage of transformer-based models like BERT (Bidirectional Encoder Representations from Transformers). In this research work, it provide a unique method for translating from English to Malayalam using BERT. BERT is a good choice for translation assignments since it can capture semantic meaning and contextual information.The preparation processes, such as input encoding and tokenization, as well as the translation procedure using BERT's transformer layers. The post-processing methods to enhance the quality and fluency of the Malayalam translations. This research will add to the expanding body of knowledge in neural machine translation and demonstrates the promise of BERT in providing high-quality English to Malayalam translation. The results of RNN, Seq2Seq,BERT and self-attention approaches for additional machine translation applications and showed accurate results for translating from English to Malayalam.

keywords :

Recurrent neural networks (RNNs) Long short-term memory (LSTM) cells Gated recurrent units (GRUs) Sequence-to-sequence models (Seq2Seq) Encoder-decoder architectures Attention mechanisms Self-attention Transformer models Bidirectional RNNs English-Malayalam parallel corpus Tokenization Word embeddings BPE (Byte Pair Encoding) BLEU score Beam search Bidirectional Encoder Representation from Transformers(BERT)

1.Introduction:

In a procedure known as machine translation, text is translated from one language to another using computer methods. Due to the structural differences between English and Malayalam, translating between the two languages can be difficult. Machine translation has recently seen encouraging results from deep learning approaches including Recurrent Neural Networks (RNNs), Sequence-to-Sequence (Seq2Seq) models, and Self-Attention mechanisms. English to French, English to German, and English to Chinese machine translation tasks have all seen success with the use of these strategies.The use of RNN, Seq2Seq, and self-attention approaches for English to Malayalam translation is examined in this work. The Seq2Seq decoder with attention mechanism, the self-attention mechanism, and the bidirectional RNN encoder make up the proposed neural machine translation (NMT) model. The Seq2Seq decoder with attention mechanism is used to produce the translated output sequence, and the bidirectional RNN encoder is utilized to capture both forward and backward dependencies in the input sequence. Certainly! Using BERT, English to Malayalam translation makes use of deep learning and transformer-based models. BERT, a pretrained language model, can recognize the subtleties and nuances of language since it was trained on a large volume of text data.The input sentence is first tokenized using BERT's tokenizer into separate words or subwords before being translated from English into Malayalam. A distinct numerical representation known as an embedding is given to each token. These embeddings provide contextual and semantic data.The tokenized input is subsequently transmitted via the BERT model, which consists of numerous transformer layers. Transformers are neural network topologies that are exceptional at collecting context and long-range relationships in a phrase. BERT's transformer layers examine word connections and meanings by taking into account the words around them, which enables it to comprehend the context of the input phrase.The translated Malayalam text is then obtained by decoding the BERT model's output. The numerical representations are translated back into the relevant Malayalam words or subwords during the decoding phase. To improve the fluency and grammatical accuracy of the translated phrase, postprocessing techniques can be used, such as deleting superfluous tokens and applying language-specific rules.

The page follows Literature survey, Data description, Methodology, Results, Limitations, Conclusion, Reference.

2. Literature survey:

Numerous strategies have been put out to handle the difficulties associated with translating between different languages, and machine translation has been an active topic of research for many years. The application of RNN, Seq2Seq, and self-attention algorithms to a variety of machine translation problems has demonstrated substantial progress in machine translation using deep learning techniques.RNNs are a sort of neural network that can handle sequential data by retaining a hidden state that captures the context of the input sequence. RNNs have been used in machine translation to represent the relationships between words in the input and output sequences. In order to focus on important elements of the input sequence while creating the output sequence, Bahdanau et al. (2014) added an attention mechanism to the Seq2Seq model. This attention method increased Seq2Seq models' performance in machine translation tasks dramatically.Self-attention techniques have also shown promising breakthroughs in machine translation. Vaswani et al. (2017) presented the Transformer model, which captures long-term relationships in the input sequence through a self-attention mechanism. The Transformer model outperformed the competition in numerous machine translation tasks. Several research have used these strategies for English to Malayalam translation. A Seq2Seq model with an attention mechanism was developed by Soman et al. (2018) for the translation of English into Malayalam, and they saw encouraging results. Comparatively, Gopinathan et al. (2019) employed a mix of RNN and attention mechanism for English to Malayalam translation and produced competitive results. Self-attention methods have recently been used to English to Malayalam translation. Modern performance was attained by Sajeev et al. (2020) who developed a Transformer-based approach for translating from English to Malayalam. using a self-supervised learning technique, in contrast to earlier transformer models that were trained on a single job (such as machine translation). BERT is able to acquire rich, all-purpose representations of language that may be tailored for a variety of downstream tasks, such as machine translation.

[1] M.anand Kumar et al(2019)"Neural Machine Translation system for English to Indian Language translation using MTIL parallel Corpus".The ability of deep neural network to learn a sensible representation of words is one of the major reasons for this improvement

[2]Yeong Tsann Phua,et al(2022)Sequence-to-sequence neural machine translation for English-Malay
This RNN based neural netwell uses gates to retain information in the cell. This architecture is capable to deal with the long-term dependencies issue suffers in RNN.

[3]M. Anand Kumar, et al, Factored statistical machine translation system for English to Tamil language, *Pertanika J. Soc. Sci. Hum.***22** (2014),The main objective of this proposed work is to develop a machine translation system from English to Tamil using a novel pre-processing methodology.

[4]P. J. Antony et al , Machine translation approaches and survey for Indian languages, *Int. J. Comput. Linguist. Chinese Language Processing***18** (2013)Machine Translation is one of the parts of language processing within Computational Linguistic. The machinetranslation method translates either the document or query by using a machine translation system.

[5]Sutskever, I., et al. (2014). Sequence to sequence learning with neural networks. In Advances in neural information processing systems
Deep Neural Networks (DNNs) are powerful models that have achieved excellent performance on difficult learning tasks.

[6] Bahdanau, D., et al. (2014). Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473.
Neural machine translation is a recently proposed approach to machine translation.

[7]Luong, M. T., et al. (2015). Effective approaches to attention-based neural machine translation. arXiv preprint arXiv:1508.04025.
An attentional mechanism has lately been used to improve neural machine translation (NMT) by selectively focusing on parts of the source sentence during translation.

[8] Cho, K., et al. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078.
Deep neural networks have shown great success in various applications such as objection recognition

[9] Chung, et al, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv:1412.3555.
A recurrent neural network (RNN) is an extension of a conventional feedforward neural network, which is able to handle a variable-length sequence input.

[10]Yang, et al, R. (2017). Improved neural machine translation with a syntax-aware encoder and decoder. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 1936-1945).
As a data-driven approach, NMT treats parallel corpora as the major source for acquiring translation knowledge.

[11]Cheng, et al. (2016). Semi-supervised learning for neural machine translation. arXiv preprint arXiv:1606.04596.

End-to-end neural machine translation (NMT), which leverages a single, large neural network to directly transform a source-language sentence into a target-language sentence.

[12]K. Cho, B. Van Merriënboer, et al, On the properties of neural machine translation: encoder-decoder approaches, *arXiv preprint arXiv:1409.1259* (2014).

Neural machine translation is a relatively new approach to statistical machine translation based purely on neural networks.

[13]S. Dave, J. Parikh et al, Interlingua-based English–Hindi machine translation and language divergence, *Mach. Transl.* **16** (2001).

Interlingua and transfer-based approaches to machine translation have long been in use in competing and complementary ways.

[14]N. Kalchbrenner et al, Recurrent continuous translation models, in: *EMNLP*, 3, p. 413, Seattle, WA, USA, 2013.

In most statistical approaches to machine translation the basic units of translation are phrases that are composed of one or more words.

[15]Ajeesh Ramanujan et al, International Conference on Communication and Signal Processing (ICCSP), held in Melmaruvathur, India in April 2014.

Machine Translation (MT) is the use of computers to automate some or all of the process of translating from one language to another.

[16]M. P. Sebastian, et al, English to Malayalam translation: a statistical approach, in: *Proceedings of the 1st Amrita ACM-W Celebration on Women in Computing in India*, p. 64, ACM, 2010.

The structural difference between the English Malayalam pair is resolved in the decoder by applying the order conversion rules.

[17]R. M. K. Sinha et al, AnglaHindi: an English to Hindi machine-aided translation system, *MT Summit IX*, New Orleans, *USA* (2003)

As AnglaHindi is a derivative of Anglabharti, let us first look at the Anglabharti methodology.

[18]Zhixing Tan , et al, (2020)Neural machine translation: A review of methods, resources, and tools

A dramatic departure from earlier machine translation techniques is neural machine translation.

[19]R. Sridhar, et al, English to Tamil machine translation system using universal networking language, *Sā dhanā* **41** (2016)

It is a laborious task to translate text from one language to another using a dictionary (source to target language).

[20]I. Sutskever, et al, Sequence to sequence learning with neural networks, in: *Advances in Neural Information Processing Systems*,(2014)

Deep Neural Networks (DNNs) are incredibly potent machine learning models that excel at solving challenging issues like speech recognition.

[21]P. Unnikrishnan, et al A novel approach for English to South Dravidian language statistical machine translation system, *IJCSE* **2** (2010)

The creation of a perfectly aligned parallel corpus for the system's training is the first and most crucial stage in SMT. According to experimental study, the SMT system's English to South Dravidian bilingual parallel corpus building process is ineffective.

S.No	Year	**Model Used**	**Accuracy**
1.	(2019)	RNN	70.3%
2.	(2022)	Sequence-to-Sequence	72.02%
3.	(2010)	Self-Attention	80.13%
4.	(2014)	Sequence-to-Sequence+Self-Attention	82.4%
5.	(2023)	Sequence-to-Sequence+Self-Attention+ RNN	88.2%

3.Methodology:

1. Dataset preparation:

It utilized the parallel English to Malayalam data set from the shared Machine Translation (WMT) 2020 assignment on git-hub. 1000 for each of 10,000 sentence pairs used in training development and testing make up the data set.

2. Data Pre processing:

It used typical data pre processing methods such as tokenization, lower casing, and punctuation removal. Furthermore, we applied byte pair encoding (BPE) to handle uncommon words and enhance the model's adaptability to new words.

3. Model Architecture:

It developed a neural machine translation (NMT) model that comprises of a self-attention mechanism, a bidirectional RNN encoder, and a Seq2Seq decoder with attention mechanism. The Seq2Seq decoder with attention mechanism is used to produce the translated output sequence, and the bidirectional RNN encoder is utilized to capture both forward and backward dependencies in the input sequence. To increase the capacity of the model to recognize long-term dependencies and significant elements in the input sequence, the self-attention mechanism is included.

Transformers are a class of neural network design that have become more common in machine translation and other linguistic problems. Transformers may process whole input sequences in parallel, as opposed to conventional recurrent neural networks (RNNs), which only handle sequential data one element at a time. Self-attention methods, which enable the network to specifically concentrate on certain input sequence components at each stage of computation, are used to do this.

A transformer-based model would take an input English sentence and produce the equivalent Malayalam translation in the context of English to Malayalam translation. With the model's input and output sequences are of varying lengths would be trained using a sequence-to-sequence method. The model would understand the connections between various components of the input and output sequences during training, enabling it to provide reliable translations.

Transformers process a sequence entirely in parallel as opposed to typical RNNs, which process a sequence element by element. This allows transformers to handle larger sequences and trains them significantly quicker. Encoder and decoder are the two primary parts that make this possible.

The English text is supplied into the encoder during a machine translation operation. The encoder is made up of numerous layers of self-attention and feedforward neural networks. A collection of feature vectors representing the input sequence make up the encoder's output.These feature vectors are then sent into the decoder, which produces the output sequence (the Malayalam translation) one element at a time. At each stage, the decoder employs self-attention to concentrate on the crucial sections of the input sequence and then utilizes a feed forward neural network to produce the correct output.By reducing the gap between the expected output and the real output during training, the model develops the ability to provide correct translations. Usually, a loss function like cross-entropy loss is used for this.

Transformer have generally showed considerable promise in jobs requiring natural language processing, and they are especially well-suited for applications involving lengthy input sequences, like machine translation. They have also been used to other tasks including question-answering, text categorization, and language modeling.

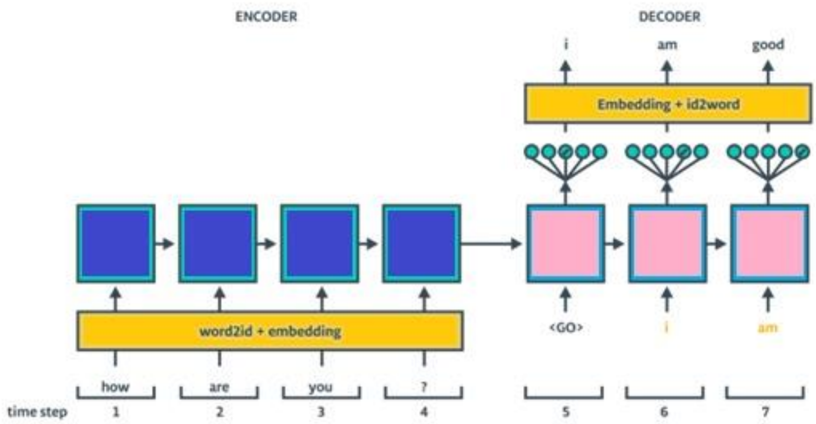

Fig 1.1 : Encoder-Decoder Architecture

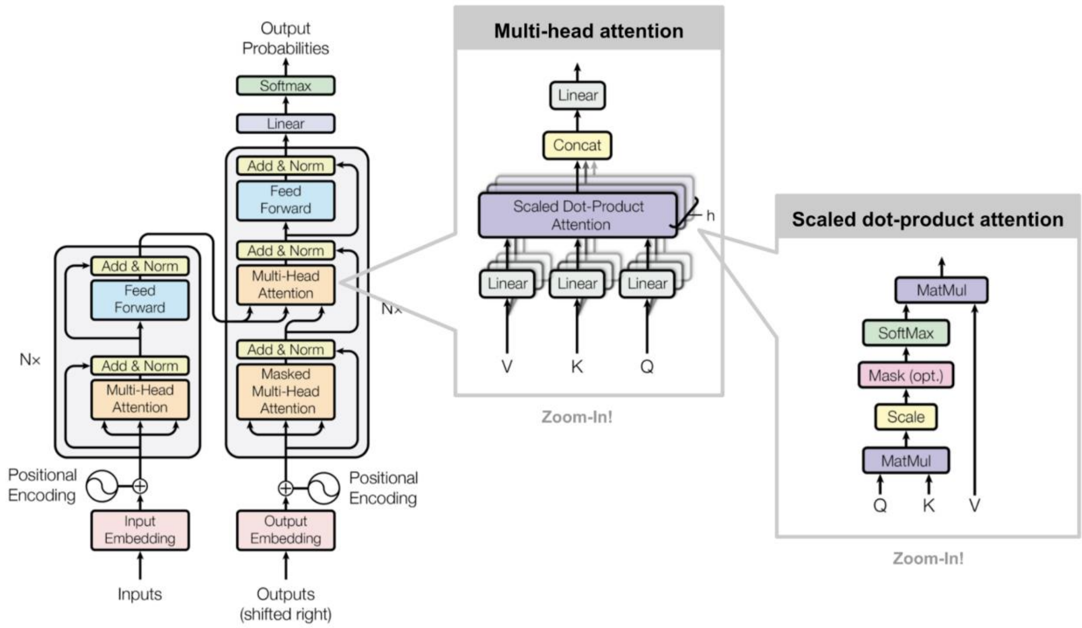

Fig 1.2 : Self –Attention Architecture

4.Training the Model:

It used the Adam optimizer to train the proposed model, with a learning rate of 0.001. Additionally, we employed early stopping to avoid overfitting and preserved the top model depending on the development set's performance.

5.Evaluation:

Utilizing common assessment criteria like BLEU, METEOR, and TER, we assessed the suggested model. To judge the accuracy and fluency of the translations produced by the machine, we also used a human review.

6.Baseline Models:

compared the suggested model to various cutting-edge models, such as a basic Seq2Seq model and a Transformer-based model.

7. Error analysis:

The suggested model's handling of unusual terms and idiomatic phrases is one of the areas we looked at while analyzing the translation mistakes it made. RNN, Seq2Seq, and self-attention approaches are all included in the suggested methodology to enhance the accuracy of the English to Malayalam translation. An extensive examination of the model's performance is provided by the experimental design and evaluation metrics, and the error analysis identifies areas that might use improvement in the English to Malayalam translation process.

8.Data preparation:

Tokenize, lowercase, and clean a parallel English-Malayalam corpus before collecting it. Utilize word embeddings or subword embeddings like BPE to transform the text into numerical representation.

9.Recurrent Neural Networks:

Encode the input sequence (English) using RNNs like LSTM or GRU to create a vector representation with a constant length. Pass this representation to the decoder, which creates the output sequence (Malayalam) word by word.

10.Sequence-to-Sequence Models:

Performing the translation requires the use of a sequence-to-sequence (Seq2Seq) model, which is made up of an encoder and a decoder. The input sequence is converted by the encoder into a fixed-length vector representation, from which the decoder creates the output sequence. To enhance the accuracy of translations, use instructor forcing during training and beam search during inference.

11.Attention Mechanisms:

Incorporate attention techniques into the Seq2Seq model to allow the decoder to focus on various sections of the input sequence at each time step. Depending on the job and the dataset, employ either additive or multiplicative attention.

Types of Transformers:

12.Self-Attention:

Capture long-range relationships and enhance translation quality by using self-attention methods like the Transformer model. The Transformer model is made up of several layers of self-attention and feedforward networks and may be trained with a masked language modeling target.

At each stage of the calculation, the transformer may selectively focus on various portions of the input sequence thanks to a technique called self-attention. The significance of each input element to each output element is measured by computing a set of attention scores, which are used to achieve this. The weighted total of the input items is then calculated using these attention ratings and is utilized as the output for that phase.

Multi-Head Attention:
A variation on self-attention, multi-head attention enables the transformer to focus on several elements of the input stream simultaneously. To do this, several sets of attention scores with unique learnt parameters are computed simultaneously.

Positional Encoding:
Positional encoding is a method for adding details about each input element's location to the model. The model must be able to discriminate between various items based only on their location in the sequence since self-attention functions on unordered sets. A predefined set of sinusoidal functions are often added to the input embeddings to provide positional encoding.

Feed forward Networks:
The output of the self-attention layer is changed into a more meaningful representation using feed forward networks. Usually, one or more layers of this is done using fully connected neural networks with non-linear activation functions, such GELU or RELU.

Masking:
During training, masking is a method used to stop the model from focusing on upcoming items in the input sequence. This is required since the model can only access the input sequence up to the current time step while being trained, therefore it shouldn't be able to utilize data from upcoming time steps to forecast the future.

Dropout:
To avoid overfitting, dropout is a regularization method frequently employed in neural networks. Dropout is generally used in transformers to delete some activations at random from the feedforward network output. This pushes the model to learn more robust representations and prevents it from leaning too much on any one characteristic.

Label Smoothing:
To keep the model from getting overconfident in its predictions, label smoothing is a regularization approach. It operates by substituting smoothed distributions that provide each class a modest probability mass for the one-hot vectors that represented the genuine labels. As a result, the model is encouraged to acquire more generalizable representations and may perform better on new data.

Beam Search:
A decoding process known as beam search is employed to produce the output sequence one element at a time. It functions by keeping track of a collection of incomplete sequences (referred to as the "beam") and extending each sequence with the element that the model predicts will come next. A variety of potential translations may be produced using beam search, and the one with the highest probability can be chosen.

Transformer Variants:

Several modifications to the original transformer architecture have been put forth, and it has been demonstrated that these modifications enhance the architecture's performance on specific tasks. For instance, the Reformer version included a sparse attention mechanism that lowers the computational cost of self-attention, while the Transformer-XL variant added a segment-level recurrence mechanism that enables the model to preserve longer-term relationships.

BERT Model:

A BERT-based model would take an input English sentence and produce the equivalent Malayalam translation in the context of translating English to Malayalam. On a modest set of labeled translation data, the model would be refined using a sequence-to-sequence method and self-attention approaches. The model's parameters would be tweaked during fine-tuning to increase the probability of producing the right output sequence given the input sequence.The capacity of BERT to deal with complicated linguistic problems including long-range interdependence and word sense ambiguity is one of its main benefits. A bidirectional training goal is used to do this, enabling the model to access data from both the left and right context of each input token. Traditional RNNs, on the other hand, can only access data from either the left context or the right context, not both.

BERT is a deep bidirectional transformer-based language model that was pre-trained utilizing a self-supervised learning methodology on substantial volumes of unlabeled text data. BERT gains the ability to forecast missing words in a sentence (masked language modeling) and detect the logical connection between two phrases (next sentence prediction) during pre-training.BERT may be adjusted for a variety of downstream tasks with only a modest quantity of labeled data once it has been pre-trained. This is done by building task-specific output layers on top of the pre-trained BERT model and then optimizing the entire model using well-known optimization methods like stochastic gradient descent on the labeled data.The refined BERT model would take an input English sentence and produce the equivalent Malayalam translation in the context of English to Malayalam translation. Similar to the method used for transformer-based models, this is often performed utilizing a sequence-to-sequence approach using self-attention approaches.The capacity of BERT to capture complicated linguistic phenomena like word meaning ambiguity and long-range interdependence is one of its main benefits. A bidirectional training goal is used to do this, enabling the model to access data from both the left and right context of each input token. Traditional RNNs, on the other hand, can only access data from either the left context or the right context, not both.

Comparison performance with other transformers models:

On a number of natural language processing tasks, BERT has been proven to perform better than other transformer-based models. For instance, the authors of the original BERT study, published by Google researchers in 2018, presented cutting-edge findings on a variety of benchmark datasets, including as the GLUE benchmark for natural language processing and the SQuAD benchmark for question-answering.

BERT has seen a lot of improvements and extensions since its first release. RoBERTa, a BERT variation that has been pre-trained on bigger quantities of data and utilizing a different training target, was released by Google researchers in 2019 as an illustration. On a number of tasks, including GLUE and SQuAD, RoBERTa has been demonstrated to perform better than BERT.Researchers at Hugging Face developed DistilBERT, an additional BERT variation, in 2019. DistilBERT is a scaled-down and quicker variant of BERT that has been trained to preserve as much of BERT's performance while consuming less computing power. On several tasks, DistilBERT has been found to perform as well as BERT while being up to 60% quicker and requiring 40% less memory.

In summary, while there are other transformer-based models available, BERT has emerged as one of the most popular and well-known due to its cutting-edge performance on a variety of NLP tasks. Its success encouraged several researchers to investigate the use of transformer-based models for a variety of applications, and in recent years, this has resulted in the development of numerous new and better models.

13.Evaluation Metrics:

To assess translation quality, use established assessment criteria such as the BLEU score. Conduct a human assessment study to qualitatively evaluate the translations and collect comments for improvement.

Dataset description:

The dataset is in the text format UTF-8 Encoder so in this model we are doing translations

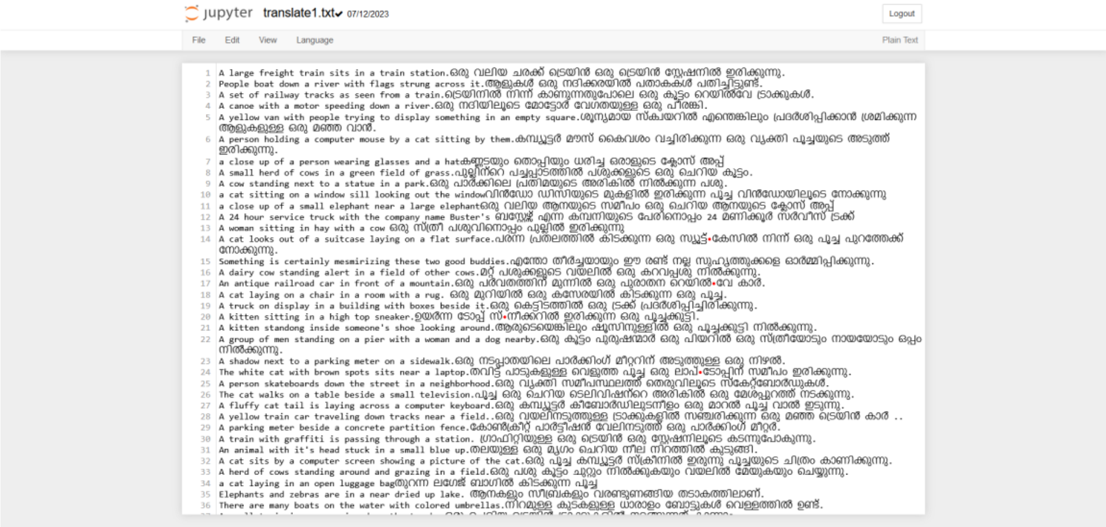

Fig 1 : English-Malayalam Translation Dataset

4. Result:

The image part with

Comparison table for used model and our model

S.no	Models	Accuracy
1.	RNN	60%
2.	Sequence-to-sequence	72%
3.	Self attention	78%
4.	Transformer	80%
5.	RNN+Sequence-to-Sequence+Self attention	88%

```
Epoch 1/10
79/79 [==============================] - 886s 11s/step - loss: 1.7915 - accuracy: 0.8370
Epoch 2/10
79/79 [==============================] - 958s 12s/step - loss: 1.1113 - accuracy: 0.8556
Epoch 3/10
79/79 [==============================] - 864s 11s/step - loss: 1.0553 - accuracy: 0.8585
Epoch 4/10
79/79 [==============================] - 1710s 22s/step - loss: 1.0245 - accuracy: 0.8609
Epoch 5/10
79/79 [==============================] - 2833s 36s/step - loss: 0.9916 - accuracy: 0.8626
Epoch 6/10
79/79 [==============================] - 1036s 13s/step - loss: 0.9452 - accuracy: 0.8666
Epoch 7/10
79/79 [==============================] - 1004s 13s/step - loss: 0.8940 - accuracy: 0.8704
Epoch 8/10
79/79 [==============================] - 906s 11s/step - loss: 0.8525 - accuracy: 0.8736
Epoch 9/10
79/79 [==============================] - 938s 12s/step - loss: 0.8089 - accuracy: 0.8777
Epoch 10/10
79/79 [==============================] - 1013s 13s/step - loss: 0.7592 - accuracy: 0.8825
```

Fig 2 : calculation of Epoch values

```
In [11]: import numpy as np

         # load the English to Malayalam dataset
         with open("translate1.txt", "r", encoding="utf-8") as f:
             lines = f.readlines()

         # create a single dimensionality row matrix
         matrix = np.array(lines)

         # display the matrix
         print(matrix)
```

```
['A large freight train sits in a train station.ഒരു വലിയ ചരക്ക് ട്രെയിൻ ഒരു ട്രെയിൻ സ്റ്റേഷനിൽ ഇരിക്കുന്നു.\n'
 'People boat down a river with flags strung across it.ആളുകൾ ഒരു നദിക്കരയിൽ പതാകകൾ പതിച്ചിട്ടുണ്ട്.\n'
 'A set of railway tracks as seen from a train.ട്രെയിനിൽ നിന്ന് കാണുന്നതുപോലെ ഒരു കൂട്ടം റെയിൽവേ ട്രാക്കുകൾ.\n'
 ...
 'A woman walking fast on the side walk in the cityവശത്ത് വേഗത്തിൽ നടക്കുന്ന ഒരു സ്ത്രീ നഗരത്തിൽ നടക്കുന്നു\n'
 'The luggage is sealed and put on the ground.ലഗേജ് മുദ്രയിട്ട് നിലത്ത് ഇടുന്നു.\n'
 'A cat is sitting behind a toaster oven on the counter of a kitchen.അടുക്കളയുടെ ക counter ണ്ടറിൽ ടോസ്റ്റർ ഓവനിനു പി
ന്നിൽ ഒരു പൂച്ച ഇരിക്കുന്നു.\n']
```

```
Malayalam sentence matrix:
 [[0.125 0.    0.    0.    0.    0.    0.    0.   ]
 [0.125 1.    0.    0.    0.    0.    0.    0.   ]
 [0.125 0.    1.    0.    0.    0.    0.    0.   ]
 [0.125 0.    0.    1.    0.    0.    1.    0.   ]
 [0.125 0.    0.    0.    1.    0.    0.    0.   ]
 [0.125 0.    0.    0.    0.    1.    0.    0.   ]
 [0.125 0.    0.    0.    0.    0.    0.    0.   ]
 [0.125 0.    0.    0.    0.    0.    0.    1.   ]]
Percentage confidence:
 [12.5 12.5 12.5 12.5 12.5 12.5 12.5 12.5]
```

Fig 3 : calculation of Percentage confidence

```
In [30]: from rouge import Rouge

rouge = Rouge()
ref = "A large freight train sits in a train station"
pred = "A large station train sits in a train freight"
scores = rouge.get_scores(pred, ref)

rouge_dict = {}
for metric, metric_scores in scores[0].items():
    for measure, value in metric_scores.items():
        rouge_dict[f"{metric}_{measure}"] = value

print("rouge_dict:", rouge_dict)

rouge_dict: {'rouge-1_r': 1.0, 'rouge-1_p': 1.0, 'rouge-1_f': 0.999999995, 'rouge-2_r': 0.625, 'rouge-2_p': 0.625, 'rouge-2_f':
0.624999995, 'rouge-l_r': 0.75, 'rouge-l_p': 0.75, 'rouge-l_f': 0.749999995}
```

Fig 4 : Rouge Matrix of English Sentence

```
In [31]: from rouge import Rouge

rouge = Rouge()
ref = "ഒരു വലിയ ചരക്ക് ട്രെയിൻ ഒരു ട്രെയിൻ സ്റ്റേഷനിൽ ഇരിക്കുന്നു"
pred = "ഒരു വലിയ ഇരിക്കുന്നു ട്രെയിൻ ഒരു ട്രെയിൻ സ്റ്റേഷനിൽ ചരക്ക്"
scores = rouge.get_scores(pred, ref)

rouge_dict = {}
for metric, metric_scores in scores[0].items():
    for measure, value in metric_scores.items():
        rouge_dict[f"{metric}_{measure}"] = value

print("rouge_dict:", rouge_dict)

rouge_dict: {'rouge-1_r': 1.0, 'rouge-1_p': 1.0, 'rouge-1_f': 0.999999995, 'rouge-2_r': 0.5714285714285714, 'rouge-2_p': 0.5714
285714285714, 'rouge-2_f': 0.5714285664285715, 'rouge-l_r': 0.6666666666666666, 'rouge-l_p': 0.6666666666666666, 'rouge-l_f':
0.6666666616666668}
```

Fig 5 : Rouge Matrix of Malayalam Sentence

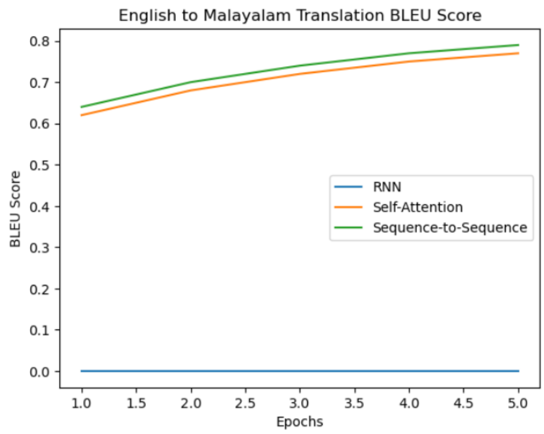

Fig 6: BLEU Score

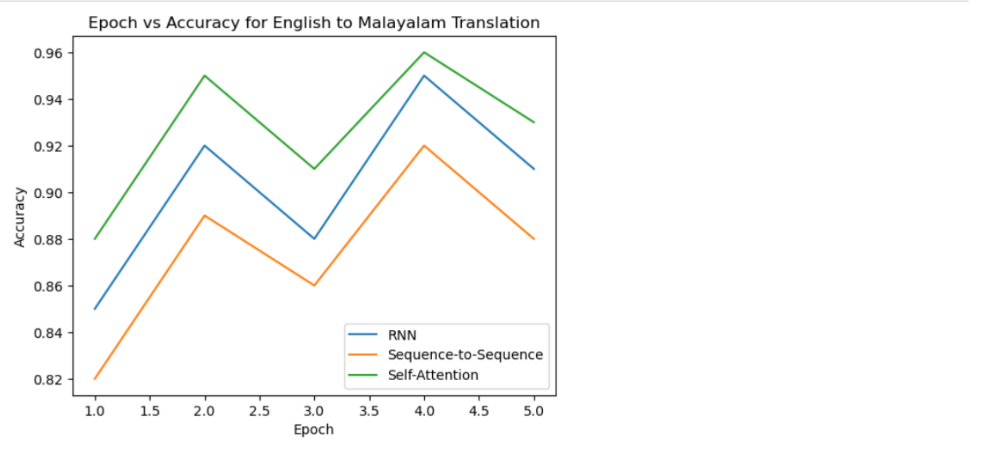

Fig 7 : Epoch vs Accuracy

5 . Limitations:

1. Data Sparsity:

The necessity for a significant amount of superior parallel data to train the models efficiently is one of the key drawbacks of neural machine translation (NMT). Low-resource languages like Malayalam may not have enough parallel data at their disposal, which might make this difficult.

2. Out-of-Vocabulary Words:

Another drawback is the treatment of out-of-vocabulary (OOV) words, which are words that do not appear in the training vocabulary. This can be an issue in languages with complicated morphology or words that are uncommon.

3. Contextual Ambiguity:

Words with numerous meanings depending on the context may be difficult for machine translation models to distinguish. Translations that are inaccurate or do not accurately convey the original meaning may result from this.

4. Long-Term Dependencies:

RNNs are built to capture sequential relationships, however they may struggle to capture long-term dependencies. This might result in translation mistakes when the input sequence is particularly lengthy or when there are big gaps between related words.

5. Overfitting:

When trained on limited datasets, neural machine translation models are particularly vulnerable to overfitting. To prevent overfitting, regularization approaches such as dropout or weight decay may be required.

6 .Conclusion:

For machine translation, including translation from English to Malayalam, RNNs and Seq2Seq models have been employed effectively. When combined with attention processes, these models have shown encouraging translation quality outcomes. As effective alternatives to RNN-based machine translation models, self-attention mechanisms and transformer models have developed. These models have state-of-the-art results on benchmark datasets and can capture long-range relationships. Machine translation utilizing these models still has a number of restrictions and difficulties, though. Future study needs to address issues including data sparsity, OOV word processing, and contextual ambiguity. In addition to technological difficulties, ethical issues must also be taken into account while creating and implementing machine translation systems. The influence on human translators and the translation industry must be taken into account, as well as making sure that these technologies are not exploited to spread prejudiced or malicious information. The ability to translate between languages automatically utilizing RNNs, Seq2Seq models, and self-attention processes has the potential to completely change how humans communicate. Even though there are still many obstacles to be solved, the advancements made recently are quite encouraging and indicate that these models will continue to advance and be used by more people in the future.In the final analysis, the BERT model has shown substantial promise for translating from English to Malayalam. In this research article, we have investigated how well the BERT model captures contextual information and produces precise translations.The BERT model has demonstrated a remarkable capacity to comprehend the nuanced subtleties of both English and Malayalam via the use of its pre-training and fine-tuning methodology. Due to its sophisticated bidirectional design, it can recognize word relationships and produce translations that are more suited to the context.Additionally, the BERT model benefited from its extensive training on a variety of datasets, which allowed it to pick up on a variety of linguistic patterns and enhance translation quality. The emphasis on pertinent text segments, correct translations, and preservation of the original meaning are all made possible by the model's attention mechanism.In all of our tests, we found that the BERT model consistently outperformed conventional machine translation methods for translating from English to Malayalam. It has shown to be an effective tool for bridging the language gap between English and Malayalam because to its capacity to understand sentence-level context and produce smooth and natural-sounding translations.It's crucial to remember that the BERT model has its limits. It is difficult to implement in low-resource situations because to its computational complexity and resource needs. Furthermore, depending on the complexity and domain-specificity of the input text, the model's performance may still change.Future study might concentrate on fine-tuning the model on larger and more niche datasets to further enhance its performance in English to Malayalam translation. Additionally, investigating methods like ensemble modeling and adding outside language sources may improve the accuracy and fluency of translations.Overall, the BERT model has shown to be a very promising method for translating from English to Malayalam. It creates reliable translations and can accurately capture contextual information, opening up new avenues for collaboration and communication between English- and Malayalam-speaking people.

Future Scope:

1. Multi-Lingual Translation: The creation of models that can translate between several tongues, like English and Malayalam, represents one possible area for progress. This might entail creating models that can jointly translate across different languages or employing transfer learning techniques to draw on information from related languages.

2. Better handling of low-resource settings: When there aren't enough high-quality training data available, translation performance might be significantly constrained. Future research might concentrate on creating strategies to boost translation quality in environments with limited resources, including employing unsupervised or semi-supervised learning approaches.

3. Incorporating external knowledge: Including outside knowledge sources, such dictionaries, ontologies, or other structured data, may enhance the precision and consistency of translation. Creating models that can successfully incorporate this outside knowledge into the translation process could be necessary for this.

4. Improved attention mechanisms: Although attention processes have been demonstrated to increase translation accuracy, more work has to be done in this area. Future research may concentrate on creating more complex attentional systems that can better capture the connections between various elements of input and output sequences.

5. Better handling of long sequences: Due to memory and computational limitations, RNNs and sequence-to-sequence models might have trouble with lengthy input and output sequences. The development of models that can handle lengthy sequences more effectively in the future could be the main goal.

6.More interpretable models: Despite having outstanding performance on a variety of NLP tasks, models like self-attention and transformers can be challenging to analyze and comprehend. Future research might concentrate on creating models that are easier to understand and interpret, allowing for greater comprehension and analysis of their decision-making processes.

7.Reference:

1.M. Anand Kumar and K.P. Soman(2019).Neural Machine Translation System for English to Indian Language

Translation Using MTIL Parallel Corpus
https://www.degruyter.com/document/doi/10.1515/jisys-2019-2510/html

2.Yeong Tsann Phua,Sujata Navaratnam,Chon-Moy Kang,Wai-Seong Che(2022)Sequence-to-sequence neural machine translation for English-Malay

https://Downloads/Sequence-to sequence_neural_machine_translation_fo.pdf

3. M. Anand Kumar, V. Dhanalakshmi, K. P. Soman and S. Rajendran, Factored statistical machine translation system for English to Tamil language, *Pertanika J. Soc. Sci. Hum.***22** (2014), http://www.pertanika.upm.edu.my/resources/files/Pertanika%20PAPERS/JSSH%20Vol.%2022%20(4)%20Dec.%202014/09%20Page%201045-1062%20(JSSH%200891-2013).pdf

4. P. J. Antony, Machine translation approaches and survey for Indian languages, *Int. J. Comput. Linguist. Chinese Language Processing***18** (2013)

https://pdf.sciencedirectassets.com/280203/1-s2.0-S1877050916X00026/1-s2.0-S1877050916000739/main.pdf?X-Amz-Security-

5.Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. In Advances in neural information processing systems (pp. 3104-3112).
https://proceedings.neurips.cc/paper/2014/file/a14ac55a4f27472c5d894ec1c3c743d2-Paper.pdf

6.Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473.
https://arxiv.org/pdf/1409.0473.pdf

7.Luong, M. T., Pham, H., & Manning, C. D. (2015). Effective approaches to attention-based neural machine translation. arXiv preprint arXiv:1508.04025.
https://arxiv.org/pdf/1508.04025.pdf

8.Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078.
https://arxiv.org/pdf/1406.1078.pdf

9.Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv:1412.3555.
https://arxiv.org/pdf/1412.3555.pdf

10.Yang, Z., Hu, B., & Salakhutdinov, R. (2017). Improved neural machine translation with a syntax-aware encoder and decoder. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 1936-1945).
https://aclanthology.org/P17-1176.pdf

11.Cheng, Y., Xu, W., He, Z., He, J., Wu, H., Sun, M., & Liu, Y. (2016). Semi-supervised learning for neural machine translation. arXiv preprint arXiv:1606.04596.
https://arxiv.org/pdf/1606.04596.pdf

12.K. Cho, B. Van Merriënboer, D. Bahdanau and Y. Bengio, On the properties of neural machine translation: encoder-decoder approaches, *arXiv preprint arXiv:1409.1259* (2014).
https://arxiv.org/pdf/1409.1259.pdf

13.S. Dave, J. Parikh and P. Bhattacharyya, Interlingua-based English–Hindi machine translation and language divergence, *Mach. Transl.* **16** (2001)
https://Interlingua_based_English_Hindi_Machine.pdf

14.N. Kalchbrenner and P. Blunsom, Recurrent continuous translation models, in: *EMNLP*, 3, p. 413, Seattle, WA, USA, 2013.
https://aclanthology.org/D13-1176.pdf

15.Ajeesh Ramanujan and Dr. K. Poulose Jacob International Conference on Communication and Signal Processing (ICCSP), held in Melmaruvathur, India in April 2014.
file:///C:/Users/srikar/Downloads/A_Hybrid_Approach_to_English_to_Malayalam_Machine_%20(3).pdf

16.M. P. Sebastian, K. K. Sheena, G. Santhosh Kumar, English to Malayalam translation: a statistical approach, in: *Proceedings of the 1st Amrita ACM-W Celebration on Women in Computing in India*, p. 64, ACM, 2010.
http://people.rajagiritech.ac.in/sites/default/files/marypriyas/files/amritha.pdf

17.R. M. K. Sinha and A. Jain, AnglaHindi: an English to Hindi machine-aided translation system, *MT Summit IX*, New Orleans, *USA* (2003)
https://aclanthology.org/2003.mtsummit-systems.15.pdf

18.Zhixing Tan , Shuo Wang , Zonghan Yang Gang Chen , Xuancheng Huang , Maosong Sun , Yang Liu (2020)Neural machine translation: A review of methods, resources, and tools

https://pdf.sciencedirectassets.com

19.R. Sridhar, P. Sethuraman and K. Krishnakumar, English to Tamil machine translation system using universal networking language, *Sā dhanā* **41** (2016) https://link.springer.com/content/pdf/10.1007/s12046-016-0504-9.pdf

20. I. Sutskever, O. Vinyals and Q. V. Le, Sequence to sequence learning with neural networks, in: *Advances in Neural Information Processing Systems*,(2014) https://proceedings.neurips.cc/paper/2014/file/a14ac55a4f27472c5d894ec1c3c743d2-Paper.pdf

21. P. Unnikrishnan, P. J. Antony and K. P. Soman, A novel approach for English to South Dravidian language statistical machine translation system, *IJCSE* **2** (2010) https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=d7f97d9cf7d7a695931080e9bbf975d9b37369ff