MALAYALAM TO ENGLISH BIDIRECTIONAL NEURAL MACHINE TRANSLATOR

*Anupam Shankar.A, *Athira S, *Gokul V Ram, *Nimisha P, **Chitra S Nair
*UG Student, **Assistant Professor, Department of Computer Science and Engineering, NSS College Of Engineering, Palakkad

Abstract

Neural Machine Translation (NMT) is one of the most promising approaches to machine translation because of its superior performance. The issues with all the existing systems such as Sequence-to-Sequence simple RNN, RNN with LSTM, RNN with GRU etc., is that they fail to give accuracy in translation. This work proposes a system that facilitates the translation of official documents from Malayalam to English implemented using Bidirectional Neural Machine Translation with Attention-based learning. Bidirectional neural machine translation takes into account both the history and future information at the same time. Hence the proposed system is expected to produce more semantically correct sentences and thereby improves the overall accuracy of the translation.

Introduction

- There is no single best translation for a given sequence of text in a source language to another language. This is because the human languages are naturally ambiguous and flexible. As a result, machine translation has become one of the most difficult tasks in artificial intelligence.
- Initially, translation was performed with the help of rule-based systems. Later, the rule-based systems were replaced with statistical methods. It is considered to be effective, yet its focus on data driven approaches results in the ignorance of key syntax distinctions called linguists.
- Then came the Neural Machine Translation that predicts the likelihood of a sequence of words with the help of artificial neural networks.
- All the existing NMT methods generate the target language sequence from left to right token by token. This won't make the full use of the target-side future contexts and thus leads to incorrect translations.
- Hence, a machine translation system is proposed which uses Bidirectional NMT with attention mechanism for translation.

Problem Definition

Requirement for an efficient machine translational system is increasing. Hence we propose a machine translation system which uses Bidirectional NMT for translation. The proposed system is expected to produce semantically correct sentences than the existing ones.

Design

- <u>Tokenization</u>
  Tokenization or segmentation covers processes such as separating punctuation from words, or processes such as applying morphological knowledge. NMT only requires a limited-size vocabulary for computational cost and enough examples to estimate word embedding. Separating punctuation and splitting tokens into words or sub words is found to be helpful to reduce vocabulary and increase the number of examples of each word, thereby improving the translation quality.

- <u>Bidirectional Encoder Network</u>
  It transforms the source sentence, one vector per input symbol, into a list of vectors. Input sequences are fed into the model with one word for every time step. The encoder is bidirectional. Let (X, Y) be a source and target sentence pair, X = $x_1,x_2,\ldots,x_m$ be the sequence of M symbols in the source sentence and Y=$y_1,y_2,\ldots,y_n$ be the sequence of N symbols in the target sentence. Then the encoder is simply a function of the following form:

$$x_1,x_2,\ldots,x_m = EncoderRNN(x_1,x_2,\ldots,x_m)$$

So the conditional probability of the sequence P(Y|X) is:

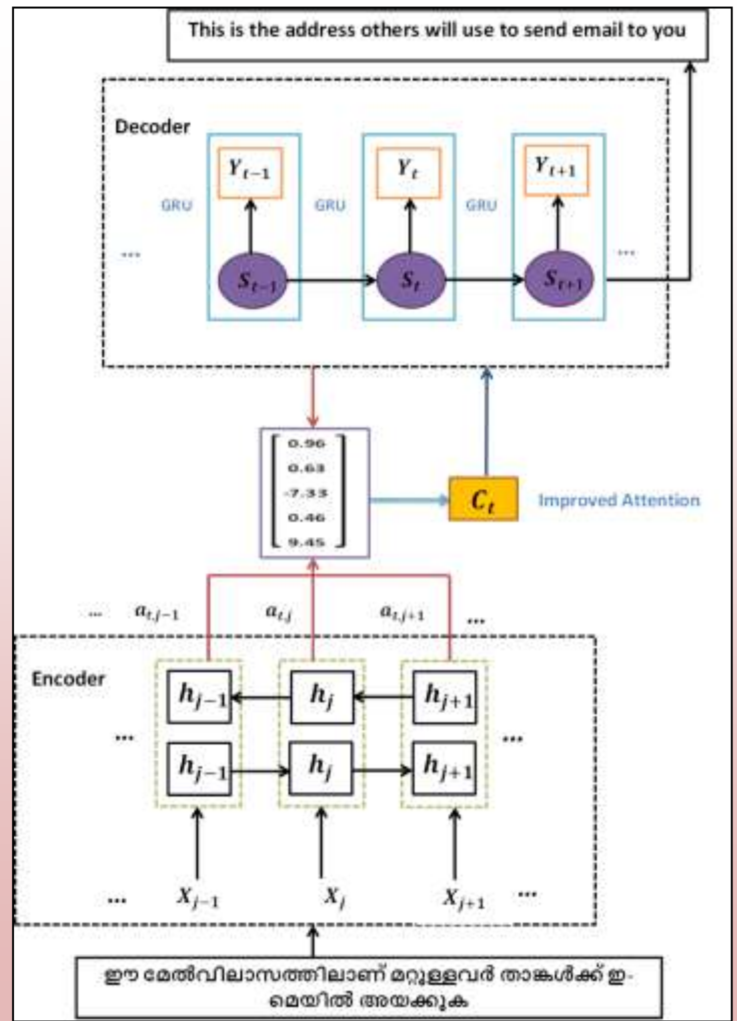
$$P(Y|X)=P(Y|x_1,x_2,\ldots,x_m)=\Pi P(y_1,y_2,\ldots,y_n;x_1,x_2,\ldots,x_m) \quad i\text{-}1,2,\ldots n$$

- <u>Attention Mechanism</u>
  It allows the decoder to concentrate on different regions of the source sentence over the course of decoding. Let y_{i-1} be the decoder RNN output from the previous decoding time step. Attention context a_i for the current time step is computed according to the following formulae.

$$s_t = \text{AttentionFunction}(y_{i-1},x_t)\ \forall t,\ 1 \leq t \leq M$$
$$p_t = \forall t,\ 1 \leq t \leq M$$
$$a_t = \Sigma p_t.\ \forall t,\ 1 \leq t \leq M$$

- <u>Decoder</u>
  The decoder must transform the learnt internal representation of the input sequence into the correct output sequence. The decoder RNN generates a hidden state y_i for the next symbol to be predicted, which is then fed to the SoftMax layer to generate a probability distribution over candidate output symbols. The outputs are calculated using the hidden state at the current time step together with the respective weight W(S). To find the sequence Y that maximizes a score function S(Y, X) a trained model beam search is used.

Architecture



Result and Analysis

- <u>Comparison with other Models</u>
  The following table shows the translation obtained from the three different models.

Input(Mal)	Output(English)		
	Simple RNN	Simple RNN with word embedding	RNN with bidirectional encoder
താങ്കളുടെ സങ്കേത പദം	The the..	The the the..	Your Password
അജ്ഞാതം	The the..	The the the…	Unknown

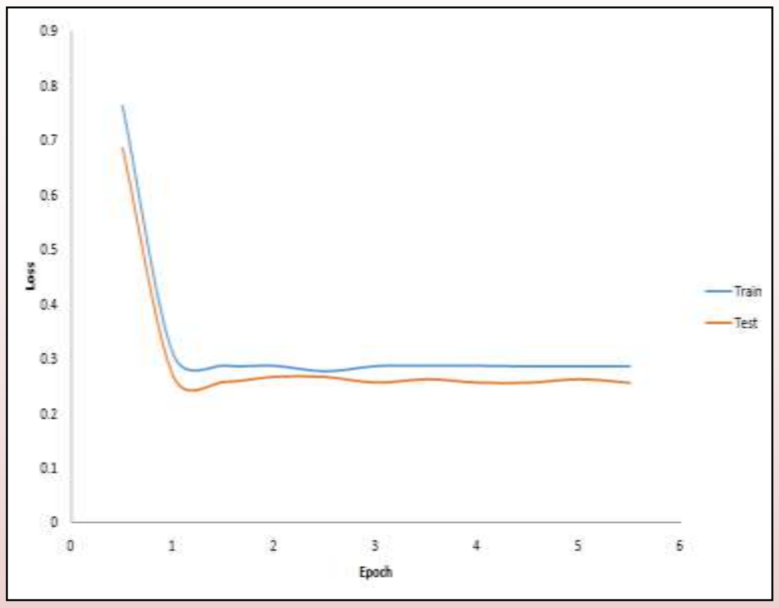
Learning curve for simpler models.

- ➢ For Simple RNN, training loss was decreasing and continued to decrease at the end of the plot. This signifies underfitting. This indicates that the model is capable of further learning and possible further improvements.
- ➢ The embedded RNN learning curve showed a flat line or noisy values of relatively high loss, indicating that the model was unable to learn the training dataset at all.

- <u>Performance Evaluation</u>

Learning curve for Bidirectional Encoder-Decoder model

Initially, the training and validation loss was considerably less than the simpler models. But still there was a large gap between the two graphs. This is because continued training of a good fit will likely lead to an overfit. Then the no. of epochs was reduced (by trial and error) and the following graph was obtained



From the graph, it can be observed that the plot of training loss decreases to a point of stability. Also, the plot of validation loss decreases to a point of stability and has a small gap with the training loss. So, it can be concluded that the proposed model was better when compared to simpler models.

Bleu Score for translation

Input	Output	BLEU score
താങ്കളുടെ സങ്കേത പദം താങ്കൾ മറന്നുവോ	Your Password if the	0.898
അജ്ഞാതം	Unknown	0.920
അച്ചടിയന്ത്ര സവിശേഷതകൾ	User Properties	0.854

The proposed model shows very high bleu rating and hence the system is showing very good results.

Conclusion and Future Work

- All of the existing methods fail to give proper translation of input source sentences. The proposed system, from the results obtained, proves to perform better than any of these methods.
- Official and professional language can be efficiently translated from Malayalam to English by using bidirectional neural machine translation with attention based learning. The system achieved improved accuracy in translation by producing more semantically correct sentences.
- The improvements that can be made in future are listed below.
  - ➢ The system can be extended to a general purpose translator.
  - ➢ It can also be extended to a general purpose translator focusing on both official and common sentences.

References

1. Johnson, M., Schuster, M., Le, Q.V., Krikun, M., Wu, Y., Chen, Z., Thorat, N., Viégas, F., Wattenberg, M., Corrado, G. and Hughes, M., 2017. Google's multilingual neural machine translation system: Enabling zero-shot translation. Transactions of the Association for Computational Linguistics, 5, pp.339-351.

2. Wu, S., Zhang, D., Zhang, Z., Yang, N., Li, M. and Zhou, M., 2018. Dependency-to-dependency neural machine translation. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 26(11), pp.2132-2141.

3. Li, Q., Wong, D.F., Chao, L.S., Zhu, M., Xiao, T., Zhu, J. and Zhang, M., 2018. Linguistic knowledge-aware neural machine translation. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 26(12), pp.2341-2354.

4. Sutskever, I., Vinyals, O. and Le, Q.V., 2014. Sequence to sequence learning with neural networks. In Advances in neural information processing systems (pp. 3104-3112).