Neural Network Based Machine Translation Systems for Low Resource Languages: A Review

Sreedeepa H. S.[1*], Sumam Mary Idicula[2]

[1]Dept. of Computer Science, Cochin University of Science and Technology, Cochin-22, India

[2]Dept. of AI & Data Science, MITS, Kochi, India

*Corresponding author's e-mail: sreedeepa@cusat.ac.in
doi: https://doi.org/10.21467/proceedings.160.43

ABSTRACT

Machine translation of documents into regional languages has important role nowadays. Deep neural networks are used in neural machine translation (NMT), which is the process of converting a set of words from one source language to some other. It is a neural network-based, fully automated translation technique. Instead of just translating a word on its own, NMT takes into account the context in which a word is used to produce more accurate translation. Instead of starting with a set of established rules, the neural network in neural machine translation is in charge of encoding and decoding the source text. MT has several advantages as compared to the traditional translation techniques and approaches. Critical analysis of different approaches used for machine translation of low resource languages were done here. Deep learning based machine translation systems, transformer learning, transfer learning techniques are some of them. After the study it is concluded that nowadays NMTs developed by taking the advantages of Deep neural networks and transfer learning approaches. Gives better accuracy than other systems. Though it is a tedious task to convert one or multiple languages to another language with 100% of accuracy as manual translation, the machine translation systems developed with these techniques can score a remarkable accuracy. As there is a lack of large parallel corpora for most of the Indian languages, the translation process become more tedious. The role of transfer learning comes in this point. Transfer learning can improve translation of low resource languages, as it can use prior knowledge in translation of a separate language pair in machine translation. This is a work done for developing a translation system for low resource language pair like Sanskrit and Malayalam. There're very less research works done in Sanskrit and Malayalam machine translation.

Keywords: Neural machine translation, Deep neural networks, transfer learning

1 Introduction

Languages are the medium of communication, and way to express the ideas of human. Given that there are more than 5000 different languages spoken worldwide, it is challenging for one person to learn and comprehend every single one of them. So, the system for translating one language to other language. Machine translation is the process of translating text from one language into another language using artificial intelligence (AI). In Neural Machine Translation it uses neural networks for automated machine translation. NMTs can provide more accurate translation results as they are considering the context of a word in addition to the individual word meaning for translation. In Neural machine translation and its extension Deep neural machine translation huge neural network layers predicts each words chance of appearance. NMT models use deep learning and representation learning. Without using a big library of rules, a trained deep neural network translates the phrases using history and prior experiences. Machine learning uses feature learning or representation learning techniques for finding the characteristics required for choosing relevant features and grouping of similar items from a given data set. Basically, MT replaces the words in a language with corresponding words in the other language. MT systems which perform translation just for a particular pair of languages are called bilingual systems. They are of two types unidirectional and bidirectional. multilingual machine translation systems are used to translate multiple pairs of languages. The

Proceedings DOI: 10.21467/proceedings.160; Series: AIJR Proceedings; ISSN: 2582-3922; ISBN: 978-81-965621-9-9

multilingual translation systems normally act bidirectional. The various approaches of machine translation approaches are- direct, transfer, interlingua based and Neural Machine Translation (NMT). It's the major classification of MT systems.

2 Neural Machine Translation

Neural machine translation (NMT) requires number of artificial neural network for translation and occurrence of words in a sentence [1]. Both the neural machine translation and its extension deep neural machine translation are work with the help of numerous neural networks. Deep neural machine translation processes large number of neural network layers rather than single layer. NMT models use both the deep learning and representation learning. Machine learning uses feature learning and representation learning techniques for finding the patterns needed for feature detection and classification of data from a data set or corpus.

The NMT systems with encoder-decoder architecture can obtain translation quality for many language pairs. The encoder network aligns the source sentence to corresponding context vectors and the decoder network, the language model, generates target words with the help of attention mechanism. The decoder calculates weight with context vectors and generates the information regarding word alignment. But these data-driven NMT systems are not sufficient for the low-resource languages. Using deep learning, neural machine translation had got more power and now boasts the most effective algorithm for performing MT tasks. Fig. 1 represents the block diagram of NMT. The performance of NMTs largely depends upon size of the parallel corpus. As the dimension of data set or corpus is directly proportional to the accuracy of output generated, the increase in the extent of the data set increases the accuracy of the output.

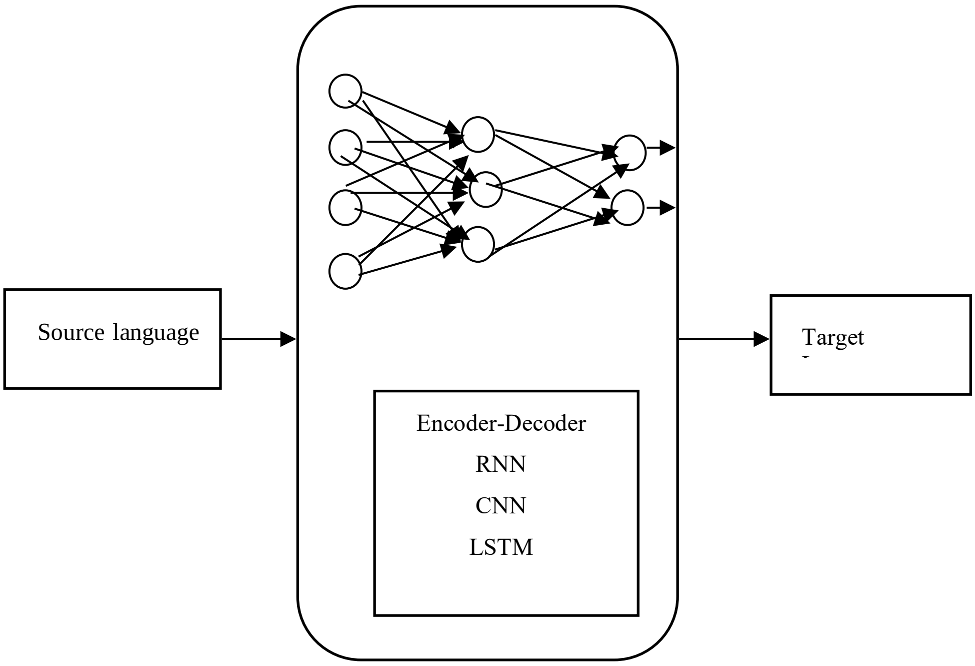

Figure 1: *Neural machine translation*

Neural machine translation uses encoder-decoder architecture with enormous neural networks to perform machine translation. The architecture of NMT is shown in the fig. 1. Neural networks given in the Fig. 1 can work with very large datasets [2]. Its two parts are one encoder network and a decoder network. Both the encoder and decoder are neural networks. It's possible to use Multilayer Perceptron neural network models for machine translation. In multilayer perceptron models it can only work with fixed length sequences. That means it can accept input sequence with fixed length and generates only the output of same length. Recurrent neural networks (RNN) are used for encoder-decoder architecture to work with variable length input and output sentences or sequences.

2.1 What Sets Neural Machine Translation Apart from Other Technologies?

TABLE. 1 gives the major difference between NMT and classical and Statistical machine translation [3]-[5].

Table 1: *How Does NMT Differ from Others*

Classical MT	Statistical MT	Neural MT
The input language is translated into target language with the help of certain rules.	Uses the statistical models for translating sentences in the given corpus.	predicts the likelihood of a sequence, such as an ordered group of words or a complete phrase, using a massive artificial neural network.
Need both the expertise to develop the rules, and large number of rules and exceptions.	Needs specialised systems to select target sentences with high probability	There is no need of pipeline of specialised systems in neural machine translation systems.
Uses parallel corpus and rule base to generate target sentence.	First splits the given input sequence into a set of words and phrases. then it replaces those to a tokens in the required output language.	Neural networks consider whole input sentence at each step and then generates the output sentence.

2.2 Deep Learning

When deep neural network (DNN) is used for MT, word alignment, Reordering and Structure prediction, Language modelling joint translation, etc. are the different stages included after the preprocessing of data. These stages can be implemented using different DNNs like recursive neural networks (RNN), recursive auto encoders (RAE), long short-term memory (LSTM) and gated recurrent units (GRU).
The architecture of encoder-decoder model is shown fig. 2 [6]. The LSTM/GRU cell acts as the encoder.

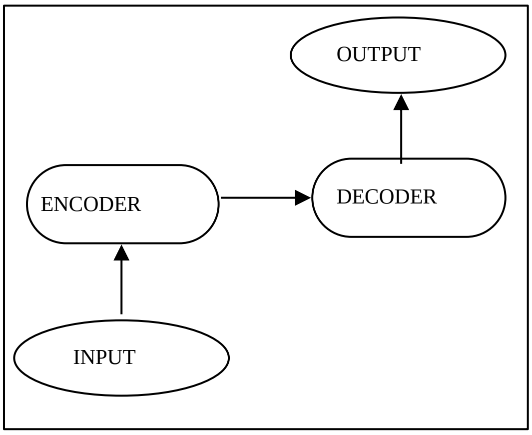

Figure 2: *Architecture of encoder-decoder model*

An encoder generates the information about the input sequence as internal state vectors. Only the internal states of encoder are considered while processing and all the output states of encoder will be neglected.
The encoder-decoder architecture implemented using RNN with attention is the current trend in machine translation. The Google Neural Machine Translation system also follows the same model, even if the system not efficiently works for low-resource languages. The accuracy of translation is very poor compared to human translation in the case of low-resource Indic languages like Sanskrit and Malayalam.
There are mainly three disadvantages for NMTs. They have slow training and inference speed, ineffective in processing of rare words, and in some situations, it completely fails to translate the whole source sentence. The sequence to sequence recurrent neural network used in encoder-decoder model can't process long sequences. It's one of the major drawback of such NMTs. That is the encoder model is poor in memorising long sequences and converting it into a fixed- length vector. So, the decoder can get only one information that is the last encoder output. So normally it generates the output by considering only some relevant parts of the sequence and couldn't consider entire sequence. The new approach called attention mechanism is introduced to handle this scenario. The attention mechanism can predict the next word by concentrating on a number of most important parts of the sequence without looking on the whole sequence. The global attention and local attention are the types of attention mechanisms used for this.

2.3 Transfer Learning

I. Zoph *et al.*, (2018) [7] applied transfer learning in machine translation and they state that with the help of prior knowledge in translation of a separate language pair it's possible to improve the translation of a low-resource language. Transfer learning method can increase the BLEU scores of low-resource languages. A high-resource language pair has to train initially, then use the resulting language model, the parent model, to initialize and train the low-resource language pair, the child model. Actually, Transfer learning gains the knowledge from a learned task and that learned knowledge is used in new related task. So, it needs less training data and can work effectively with low resource languages.

II. According to Sinno *et al.*, (2021) [8] transfer learning is categorised into inductive, transductive and unsupervised transfer learning. They done this classification after considering the input and output domains of various problems. The inductive transfer learning differs in the source and target task. There will be no issues matter even if the source and target domains are same or different.

III. Gong-Xu Luo *et al.*, (2018) [9] concluded that using data-driven NMT systems for low-resource languages are inadequate. Transfer learning will be an optimisation technique for both saving time and getting more accurate results.

IV. According to Alham *et al.*, (2020) [10] the technique which is suitable for the machine translation of low- resource languages is transfer learning. By feeding the model with the identical words for both the source and the target, it is feasible to duplicate the sequence model (or auto-encoder) model. Cross-lingual transfer learning depends on the internal layers of the network. If the vectors are successfully mapped, both the embeddings and the inner layers include transferrable information without considering parent language or corpus used. It is feasible to initialise models with a pre-trained model. This characteristic makes transfer learning a most suitable initialization for the child model. As a result, the child model will start training with gradients that are more stable.

V. Toan Q. Nguyen et al. [11] uses BPE for representing the words in a sentences with Transfer Learning approach for the translation of Low-Resource, Related Languages. This helps to transfer similar features from parent to child model. Depending on the domain, the application, and the data availability, multiple transfer learning methodologies and approaches may be used.

Table 2: Different Transfer Learning Approaches

Inductive	Transductive	Unsupervised
Target domain contains labelled data.	Source domain contains marked data	The source and target domain doesn't contain any labelled data.
Source and target tasks will learn simultaneously	Uses Domain adaptation, sample selection biase/covarience shift	The target domain is transferred based on unsupervised tasks
It can transfer instances feature representation, Parameters and Relational-Knowledge.	Learn and Transfer domain features.	Parameter transfer
Uses self-taught and multitask learning	Classified into subcategories with different feature spaces	It deals with different task with same source and target.

Table 2 shows different approaches used for transfer learning and their relationship. In transfer learning we can use one of the following approaches to solve a problem after selecting a pertained model to use knowledge from the parent model. In first approach after selecting a pre-trained model as base model freeze some layers of it and train the remaining layers on a new dataset in order to solve the new task. The second method is to create a new model and include certain features from the parent model's layers in child model. The both approaches selects some of the learned features and used to train the remaining or new layers of the model. The only characteristics that are same for both tasks should be removed from the pre-trained model, and the remainder of the model should be trained and fine-tuned to fit the new dataset.

The embeddings and alignment can be transferred from the parent without fine-tuning and it will align the input diagonally and copy most of the tokens. It is possible to initialise models with a pre-trained model without considering the parent language or corpus. Domain Adaptation is a subcategory of TL that applies the most suitable model to an expected output dataset of a different application. This is done after getting sufficient knowledge from a source data set of a different but similar task.

Develop model, pre-trained model and Feature Extraction approach are the three major TL approaches [12]. In the develop model strategy, a source model that is trained for one task is reused for another purpose after the model has been correctly tuned.

The feature extraction technique uses deep learning to identify task's most crucial characteristics in order to determine the optimal representation of the task. These are best suitable for developing a low-resource NMTs. The proposed work combines these develop model approach and feature extraction approach to develop a TL based NMTs for low resource languages in domain adaptation category. As the proposed system uses TL the BLUE score of the system can be improved upto some points. The proposed system trying to develop a Sanskrit- English MTs as target model by selecting a Sanskrit-Hindi MTs as pre-trained model. The noticeable advantages of using Transfer learning are an optimisation to saving time or getting better performance.

The process of transferring information from one target domain to another is known as transfer learning.

"Source and target setting of the selected task," "nature of the source and the target domain," and the "order in which tasks are learned" are three categories of classification of transfer learning. There are three criteria on which transfer learning can be categorised. According to the three aspects, transfer learning is divided into four, they are as follows:

1. **Domain Adaptation:** With this approach, the best performing model is applied to a target dataset for a different but related task after learning from a source dataset for a task.
2. **Cross-lingual learning:** This approach looks at a source language and uses the best model to translate it into a target language that is linguistically similar.
3. **Multitask Learning:** In this method learn a number of connected tasks simultaneously. By taking use of similarities and variations between tasks, this aids in teaching the model using the representation of current tasks.
4. **Sequential Transfer Learning:** Using source and target tasks, we successively learn a number of tasks. When learning multiple activities consecutively, the source and target tasks may be surprisingly unlike from one another. However, it is crucial to make sure that the activities that immediately follow each other are sufficient for learning more accurate representation of the target task.

The following three methods for putting transfer learning into practise are the most frequently employed ones:

1. **Develop Model Approach:** Here a source model is trained to be proficient for one task and then used for another task after being correctly tuned.
2. **Pre-trained Model Approach:** In this method, select a model that is generally accessible, such as ULMFiT or BERT, then fine-tune it to perform a general task, such as next sentence prediction.
3. **Feature Extraction Approach:** This method is also called as representation learning. Here deep learning is used to identify the task's key characteristics in order to determine the optimal representation of the task.

The major advantages of transfer learning are:

1. **Improve efficiency of model:** By capturing the semantic and syntactic relationships between words in a genuine language, it enhances model performance.
2. **Reduced Training Time:** The model may be trained for a downstream job in less time.
3. **Fewer Training Data for Fine Tuning:** It minimises the amount of data required to fine-tune the model so that it can process information provided in regional languages.
4. **Less Experimentation Time:** The pre-trained architecture cuts down on the amount of time needed to conduct experiments.
5. Less complex code

Zero-shot learning is a different form of transfer learning, and it can learn from datasets with non-labelled examples. In order to learn unseen data, it will make efficient adjustments in training stage. Zero-shot learning is much suited for applications like machine translation, where no labels in the target language. Transfer learning is classified into categories like inductive transfer learning, transductive transfer learning and unsupervised transfer learning. This classification is done based on various cases of source and target domains and tasks. In case of inductive transfer learning the source, task is different from target task. The difference in domain also not a major issue.

The Transformer model is a novel encoder-decoder that analyses voice patterns using self-awareness. This supports parallel processing and performs at a significantly faster rate than any other model.

3 Conclusion

In Neural Machine Translation, large neural networks are used to translate the text automatically. In addition to the individual words NMTs provide better translation results by considering usage of words in different situations. Both the NMTs and its's extended version, deep neural networks use a large neural network. Instead of using a single neural network layer, deep neural machine translation uses numerous layers. NMT models use deep learning and representation learning techniques to find out representations required for feature extraction from datasets. Without employing a big library of rules, a trained deep neural network interprets the phrases using history and prior experiences. Even if the deep neural networks with attention mechanisms works efficiently for the machine translation of most of the languages it's not the case with low resource languages.

As the accuracy of translation systems are directly proportional to the size of the corpus, the data used for training and the models used for the low resource languages facing a lots of problems with both NMTs and deep neural networks with attention mechanism. After this study on different NMT based systems reveals that the transfer learning with combined develop model with inductive transfer learning and feature extraction approach is best suited model for the machine translation task of low-resource languages. Machines are replacing humans in different tasks and the technology is growing faster and improves with time. Even if machine translation can't replace human translation because it's that much complicated and tedious task for implementing. Also, the human can perform translations more faster and accurately.

4 Publisher's Note

AIJR remains neutral with regard to jurisdictional claims in institutional affiliations.

How to Cite

Sreedeepa & Idicula (2023). Neural Network Based Machine Translation Systems for Low Resource Languages: A Review. *AIJR Proceedings*, 330-336. https://doi.org/10.21467/proceedings.160.43

References

[1] Subhashree Satpathy Smita Prava Mishra Ajit Kumar Nayak , "Analysis of Learning Approaches for Machine Translation Systems" , 2019.

[2] M.D. Okpur , "Machine Translation Approaches : Issues and Challenges" , 2014

[3] V. Laximi and H. Kaur , "A Survey of Machine Translation Approaches, International Journal of Science, Engineering and Technology Research" , 2013

[4] Sneha Tripathi1 and Juran Krishna Sarkhel , "Approaches to Machine Translation" , 2019.

[5] P.J. Anthony, "Machine Translation Approaches and Survey for Indian Languages, Computational Linguistics and Chinese Language Processing" , 2013

[6] S. Tripathi and J.K. Sarkhel , "Approaches to Machine Translation, Annals of Library and Information Studies"2010

[7] Barret Zoph, Deniz Yuret, Jonathan May, and Kevin Knight. 2016. Transfer learning for low-resource neural machine translation. In Proc. EMNLP, pages 1568–1575.

[8] Pan, S. J. and Yang, Q. "A survey on transfer learning." , 2010. Zhang, L., Zuo, W., and Zhang, D. "LSDT: Latent sparse domain transfer learning for visual adaptation." , 2016. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[9] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[10] Toan Q. Nguyen and David Chiang. 2017. Transfer learning across low-resource, related languages for neural machine translation. pages 296–301.

[11] Barret Zoph, Deniz Yuret, Jonathan May, and Kevin Knight. 2016. Transfer learning for low-resource neural machine translation. In Proc. EMNLP, pages 1568–1575.

[12] Ying Wei, Yu Zhang, Junzhou Huang, Qiang Yang. "Transfer Learning via Learning to Transfer" , 2018.