Proceedings of the International Conference on Applied Artificial Intelligence and Computing (ICAAIC 2022)
IEEE Xplore Part Number: CFP22BC3-ART; ISBN: 978-1-6654-9710-7

2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC) | 978-1-6654-9710-7/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICAAIC53929.2022.9793214

Natural Language Processing based Machine Translation for Hindi-English using GRU and Attention

Jaskirat Singh
UG Scholar, Dept. of Computer Science SRM Institute of Science and Technology Tamil Nadu ,India
js1759@srmist.edu.in

Sansriti Sharma
UG Scholar, Dept. of Computer Science SRM Institute of Science and Technology Tamil Nadu ,India
ss9736@srmist.edu.in

Briskilal J
Assistant Professor Dept. of Computing Technologies
SRM Institute of science and technology Tamil Nadu ,India
briskilj@srmist.edu.in

***Abstract*— Machine translation, abbreviated as MT, is basically an extension of language computation that studies the concept of translating text or speech from one language to another. On the most rudimentary level, machine translation basically conducts mechanical reinstatement of words from one particular language into another, however, this alone rarely yields a good translation because it needs to recognize the closest match between the complete sentence and the target language. Not every term in one language has an equivalent word in another, and many words have many meanings. Unlike our everyday normal phrase-based translation systems, which are usually made up of many small and tiny sub-components that can be tweaked individually, NMT i.e., neural machine translation seeks to design and train a simple, massive, end-to-end neural network system that analyses a sentence and provides an accurate translation for the respective language. Neural Machine translation is a radical newer technique in the field of language translation and localization that trains neural models using deep neural networks and artificial intelligence. The main advantage of this approach is that you can train a single system directly with source and target text. This eliminates the need for special system pipelines required for statistical machine learning. This means you can perform training and translation on an end-to-end model at once.**

Keywords— Translation, Machine, Words, Text, Localization, Subfield, Target, Meanings, Phrases, Software.

I. INTRODUCTION

Natural Language Processing, also known as NLP, is a mixture of CS (computer science), natural languages, information technology, and expert systems. Humans communicate in words, but as we know computers operate in the language of numbers since for computers, the language in which humans communicate is a higher-level language which is hard for them to understand and process. These numbers can in actual sense act as a bridge between the different languages of our world. With the help of proper and deep natural language processing we can create translation systems which would lead to open and effective communication. On a very basic level, dividing paragraphs and sentences into linguistic units so that computers can understand the words used in our language can be said to be the first step towards processing natural language. These units are then converted to numbers and then back to words, but to words which are in different languages to complete the translation process. Specifically for the English – Hindi translation, availability for data is very highly stringent which led us to create our own dataset for the project. As a general point of view, translation is not just the literal replacement of a word, it is a process of interpreting and analyzing the word or phrase and how each affect 'the other'. Best way to define machine translation is “use of proper words in proper places”. Globalization brings people of different backgrounds, cultures and languages together. To overcome the language barrier, translators have stepped in. This particular paper deals with NMT of English – Hindi languages. NMT is a type of MT methodology that applies a neural model to anticipate the potential series of words, often in the form of an entire sentence. Unlike statistical machine translation, which consumes more memory and time, NMT, a neural machine translation, continuously trains those parts to maximize performance.

II. LITERATURE SURVEY

A. Machine Translation

Machine translation, also abbreviated as MT, is a process of automatically translating sentences from one particular language to another particular language using any natural language processing and machine learning technology. With increasing change in business environment, companies are using machine translation more than ever to supplement and speed up their translation process. Main advantage of MT is its speed and accuracy. When properly designed, MT can provide sufficient speed conversion for real-time enterprise communication. With the right server settings (often deployed in the cloud), MT arrays can translate hundreds of thousands of words per second. Using MT in combination with human intervention/checkers rounds can improve the turnaround time by about 35% or more, depending on the size of the project. Before the rise of neural machine translation, MT was still in rudimentary stage, producing translations that varied significantly in quality and sometimes humorously went bad or unreadable. Modern machine translation engines have changed all of the above stated problems and are now functioning as an integral tool in the translation process. It can be used "as is" in less important applications or combined with manual post-editing to speed up traditional translation workflows. From 1950s to 1980s knowledge driven approaches i.e., rule-based machine translations (RBMT), were used, which consisted of Direct MT, Transfer based MT and Interlingua MT. In simple terms RBMT consisted of rules which originated from human knowledge about language. As time went on, more and more people started taking interest in MT and hence new type of approaches of MT were originated i.e., example-based machine translation (EBMT), statistical machine translation (SMT) and neural machine translation (NMT). These approached were in general termed as data driven approaches because they basically deal with figuring out a way to find the translations based on prior present dataset.

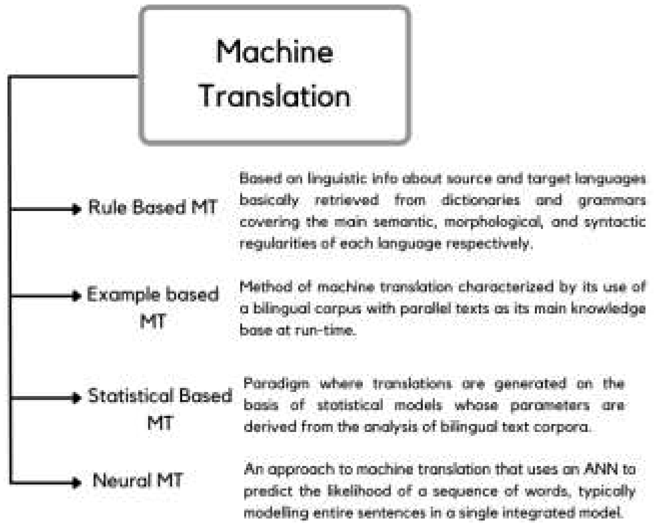

Figure 2.2.1: Machine Translation Approaches

B. Neural Machine Translation

A MT approach that uses artificial neural networks to anticipate the probability of a sequence of words is what Neural machine translation (NMT) basically is. It is simply modelling entire sentences into a single end-to-end integrated unit. NMT differs from a language-based statistical approach that uses individually developed subcomponents. Neural Machine translations main approach is the use of vector-based representations of words and internal states ("embedded", "continuous spatial representation"). The structure of the NMT model is cleaner and more comprehensive than the phrase-based model. There is only a single direct model which translates each entity i.e., word/sentence at a time instead of separate language models, translation models, and rearrangement models etc. But this sequence anticipation requires the entire source data and the whole target sequence data have already been generated. The NMT model uses deep learning and machine learning models in order to get the right translations. Word sequence modelling was mostly first performed using an RNN. A Bi-RNN, called an encoder, is used by the ANN to encode the output of a second RNN, called a decoder, which is used to anticipate the words in the target language. With iterative ANN, it is difficult to encode big/longer inputs into a single vector. This issue can be overcome for by using an attention mechanism that lets in the decoder to cognizance on specific components of the enter as it produces each phrase of output. Neural machine translation works perfectly fine until it faces sentences with longer structure. In technical terms basic encoder-decoder model faces an issue of information bottleneck. When the model has to compress all the information of the source language into a vector of fixed lengths, then this problem arises because we cannot be sure of what the length on the sentence that will be given as input will be. Since the issues of information bottleneck exist, the translation model is not able to cope up with sentences with long corpus. Attention is here for our rescue. The basic idea/concept of the attention architecture is that on each and every step of the decoder, we can simply put a connection to the encoder just to focus on a part of the source sentence. In technical terms we get a weighed sun of the values at the encoder level and then use them for our favor in the decoder part.

C. Sequence to Sequence Models

Many machine learning responsibilities contain transforming one sequence into another. Machine translation interprets textual content in a single language into textual content in another. In deep learning, such responsibilities are generally modeled using an intersequence model, additionally referred to as the Seq2Seq model. Seq2seq is a family of machine learning strategies utilized in language processing. Applications encompass speech translations, image captions, communication models, and textual content summaries. Seq2seq interprets one sequence into another. This is carried out the usage of recurrent neural networks or more commonly LSTMs or GRUs to keep away from vanishing gradient problems. A GRU is basically made up of 2 main gates i.e. a Reset gate and an update gate. The ability to use GRU's gate and memory cell is an efficient solution for vanishing gradient problems. This is because the gate from the GRU can be easily set to 0. Since we are using the sigmoid function, this absolute value of 0, which we have assumed for the sake of simplicity here, is very close to 0 in reality. And it is excellent to maintain this near 0, it is like 0.000001, and even smaller. This is why the GRU is not suffering from vanishing gradients problem. This smaller value of Gate that can be rounded to 0 helps the GRU set the actual memory value corresponding to the previous memory value, and during all time steps of our deep neural network set. In this way, GRU guarantees to be the best selection for dealing with long-term dependencies in case of basic RNN algorithms.

D. Attention based Model

Attention mechanisms are increasingly being used to enhance the performance of Neural MT (NMT) by selection that specialize in sentence subparts throughout translation. the eye mechanism is an element of the neural network. At every decoder step, confirm that supply portion is a lot of important. With this setting, the encoder doesn't need to compress the complete source into one vector, it provides a illustration of all source tokens.

E. Software Requirements Specification

- Processor: Intel(R) Core (TM) i5-1035G1 CPU @ 1.00GHz 1.19 GHz
- OS: Windows® 10
- RAM: 8.00 GB (7.77 GB usable)
- System type: x64-based processor
- Python Version: 3.7

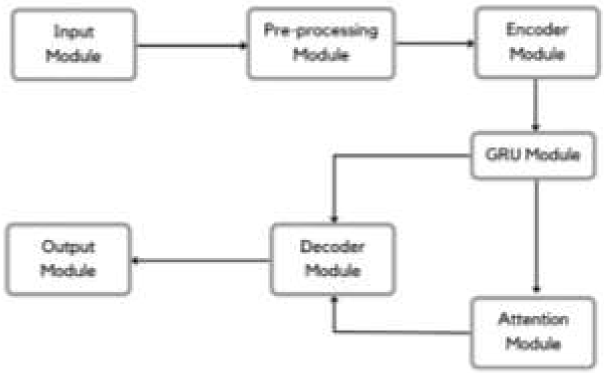

Figure 3.1: Basic Block Architecture

III. SYSTEM ARCHITECTURE AND DESIGN

A. Encoder - Decoder Model

The Encoder - Decoder RNN design is usually a good and commonplace approach for each neural MT (NMT) and intersequence prediction (seq2seq). The advantage of this approach is that you simply can directly train one end-to-end model victimization supply and target sentences, and you'll be able to handle input and output sequences of variable-length text. The Seq2seq architecture contains two long-term memory`s (LSTMs). One for the encoder and one for the decoder. GRUs (Gated Recurrent Units) are used instead of LSTMs in both encoders and decoders because GRUs require less processing power and provide almost the same results as LSTMs.

1) Encoder related steps:

a) Each word/entity with in the input sentence is embedded (encoding is being done) and defined in a different space with the dimension of the embedding. This step ensures that similar words are placed side by side in this space.
b) The embedded statement is then sent to the GRU. The final/last hidden state of the encoder GRU becomes the first/initial hidden state of the decoder GRU system. This final GRU hiding state (of encoder) contains the source set encoding. This encoding of the input sentence can also be provided by combining all the hidden states of the encoder.

2) Decoder related steps:

a) Like the encoder, there is also an embedded layer for the sequence in the target/final language. Every word in is defined by an embedded space. In embedded spaces, words with meanings that are almost same are in close proximity.
b) It also uses the current decoder hiding state and encoder output to get the weighted sum of the encoder outputs. This is done by using the attention layer in the system.
c) Finally, we put together the results that we obtained from the prior two steps. This final tensor is sent to the GRU layer of the decoder.
d) Output of this GRU layer is sent to a high-density layer that gives out the probability of all words that will appear. Word with a high probability would simply indicate that the system would considers that word to be the next word in the sentence.

B. Gated Recurrent Unit

The intent of using the GRU (Gated Recurrent Unit) is to solve the vanishing gradient downsides related to standard continual neural networks. GRU may also be thought-about as a variant of LSTM. This is often a result of their similarity in structure. To unravel the vanishing gradient problem of normal RNNs, GRUs uses update gates and reset gates. Basically, these are 2 gates confirm the data to pass to the output. What makes them special is that they'll be trained to retain previous information while not flashing information over time or deleting information that's not relevant to the predictions.GRU has 2 gates: Update gate and reset gate. The update gate is responsible for determining the amount of previous information (previous time step) that needs to be propagated along the following states. It's an important unit. It is similar to the output gate of repeating units of LSTM. The reset gate is used by the model to determine how much of the past information should be ignored. The expression is the same as for the update gate.

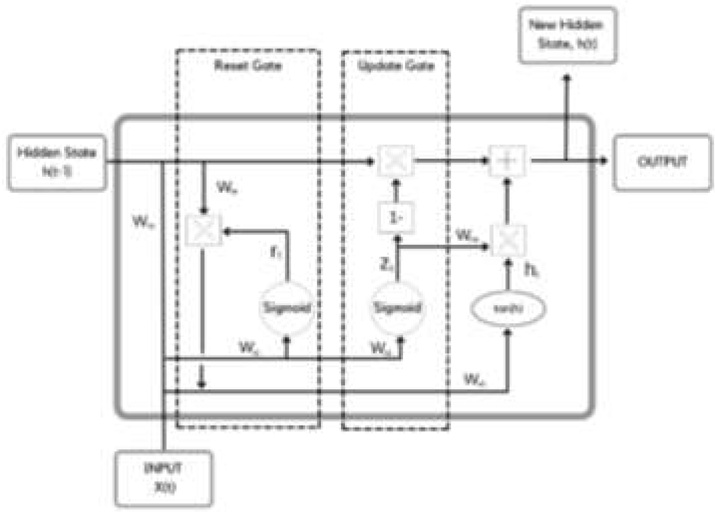

Figure 3.2.1: GRU Structure

There are differences in their weight and gate usage.

C. Attention

Attention has been planned as a resolution to the constraints of the Encoder-Decoder model, that encodes the input sequence into associated fixed set length vector and decodes every output time step from it. This issue is believed to be a tangle once secret writing long sequence. Attention has been proposed as each an alignment and translation method. Alignment may be an artificial intelligence problem that identifies which a part of the input sequence is related to each word within the output. Translation, on the opposite hand, is that the method of victimization relevant info to pick out the suitable output.

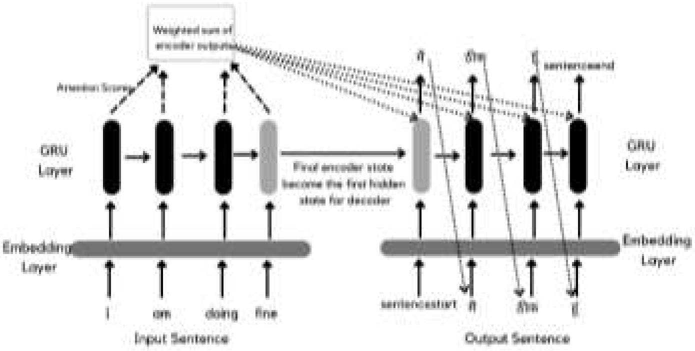

Figure 3.3.1: Encoder – Decoder Model having GRU with Attention

IV. METHODOLOGY

A. *Dataset*

The dataset for this project has been partially been fetched from online resources and partially been hand made. Due to unavailability of appropriate dataset, we have made our own novel dataset for this project. The dataset consists of around 4500 entities. Each entity consists of an English sentence and a translated sentence in Hindi language. After the dataset is formed, data pre-processing takes place where in the data is formatted in a proper useable form.

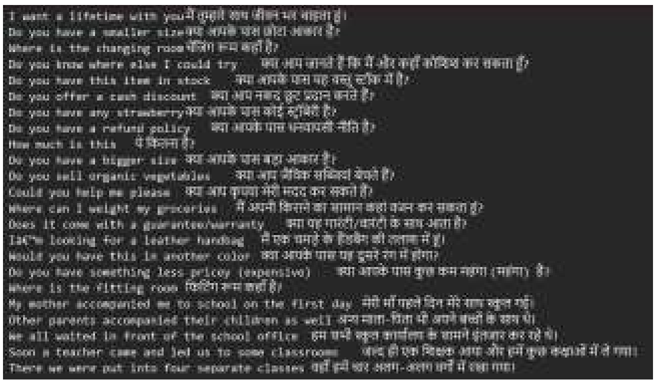
I want a lifetime with you मैं तुम्हारे साथ जीवन भर चाहता हूं।
Do you have a smaller size क्या आपके पास छोटा आकार है?
Where is the changing room चेंजिंग रूम कहाँ है?
Do you know where else I could try क्या आप जानते हैं कि मैं और कहाँ कोशिश कर सकता हूँ?
Do you have this item in stock क्या आपके पास यह वस्तु स्टॉक में है?
Do you offer a cash discount क्या आप नकद छूट प्रदान करते हैं?
Do you have any strawberry क्या आपके पास कोई स्ट्रॉबेरी है?
Do you have a refund policy क्या आपके पास धनवापसी नीति है?
How much is this ये कितना है?
Do you have a bigger size क्या आपके पास बड़ा आकार है?
Do you sell organic vegetables क्या आप जैविक सब्जियां बेचते हैं?
Could you help me please क्या आप कृपया मेरी मदद कर सकते हैं?
Where can I weight my groceries मैं अपनी किराने का सामान कहां वजन कर सकता हूं?
Does it come with a guarantee/warranty क्या यह गारंटी/वारंटी के साथ आता है?
Iâ€™m looking for a leather handbag मैं एक चमड़े के हैंडबैग की तलाश में हूं।
Would you have this in another color क्या आपके पास यह दूसरे रंग में होगा?
Do you have something less pricey (expensive) क्या आपके पास कुछ कम महंगा (महंगा) है।
Where is the fitting room फिटिंग रूम कहाँ है?
My mother accompanied me to school on the first day मेरी माँ पहले दिन मेरे साथ स्कूल गईं।
Other parents accompanied their children as well अन्य माता-पिता भी अपने बच्चों के साथ थे।
We all waited in front of the school office हम सभी स्कूल कार्यालय के सामने इंतजार कर रहे थे।
Soon a teacher came and led us to some classrooms जल्द ही एक शिक्षक आया और हमें कुछ कक्षाओं में ले गया।
There we were put into four separate classes वहाँ हमें चार अलग-अलग वर्गों में रखा गया।

Figure 4.1: Dataset

B. *Training of Neural Machine Translation*

The goal is to translate English -Hindi. There are basically 3 main models to select from Sequence-to-Sequence model (also known as Seq-Seq Model), Bidirectional-Long Short-Term Memory (Bi LSTM) and Gated Recurrent Unit (GRUs). Out of all the models GRUs comes up to be the most appropriate model since it has the less training time than the LSTM model and is more accurate than the Sequence-to-Sequence model. The translation model trains on our in-house made dataset and takes around 144 hours just to train the model.

V. RESULTS AND DISCUSSION

From the start itself GRU model performs better than the LSTM models. Currently the accuracy figure given by the model is very high indicating presence of overfitting, but this issue will be resolved over time by feeding the model more and more data entities. For now, the model has been trained on a dataset made of around 4500 entities. For the model to function at its peak and for getting a reasonable accuracy figure larger dataset and advanced hardware is required for faster processing. Currently the accuracy figures are shown the figure below:

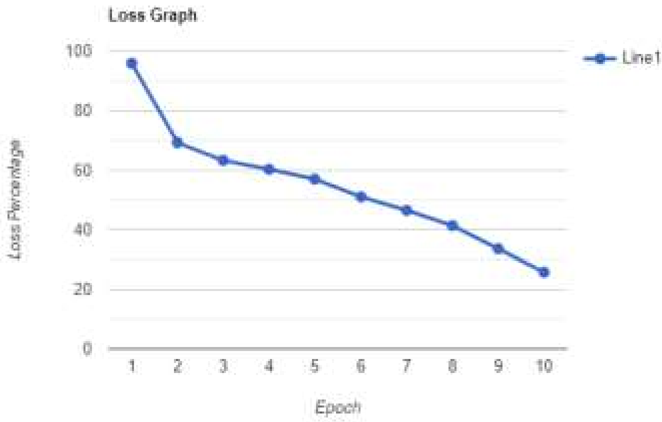

Figure 5.1: Loss V/S Epoch Number Graph

The above stated data is extracted from the NMT model we built for this paper. The above lose values have been taken from the results we get while training and testing the data. First the data comes into the input module where the input is taken in the form of strings of data. Then pre-processing takes place and after that encoding the model. Once the inputs have been encoded after that training and testing of the model starts. After all this process ends, we get an end-to-end trained model for Hindi English translation.

VI. CONCLUSION AND FUTURE ENHANCEMENT

The GRU with attention approach identifies the information in an input most pertinent to accomplishing a task and gives better performance especially in machine translation processes. The dataset made for the project has wide range of entities i.e., it consists of sentences with different lengths and even covers a good amount of vocabulary too. As far as the neural machine translation model is concerned, the current accuracy figures are around 75 percent. With increase in size of the dataset the accuracy figures will increase too. Currently the dataset is restricted to a certain range because the hardware required to train the model was not available. This was one of the biggest issues because in the training part massive amount of time is taken for the model just to train on the dataset. For future enhancement the quality of dataset can be improved and the resources to train the NMT model can be modified so that faster model training the testing can take place.

REFERENCES

[1] Shalu, P. and Meera, M., Neural Machine Translation For English to Hindi using GRU (May 22, 2021). Available at SSRN: https://ssrn.com/abstract=3851323 or http://dx.doi.org/10.2139/ssrn.3851323

[2] Ilya Sutskever, Oriol Vinyals, Quoc V. Le, Sequence to Sequence Learning with Neural Networks, Available at : arXiv:1409.3215

[3] Rahul Dey, Fathi M. Salem. Gate-Variants of Gated Recurrent Unit (GRU) Neural Networks Available at : arXiv:1701.05923

[4] Benková, Lucia & Benko, Ľubomír. (2020). Neural Machine Translation as a Novel Approach to Machine Translation.

[5] Biao Zhang, Deyi Xiong, Jinsong Su. A GRU-Gated Attention Model for Neural Machine Translation https://doi.org/10.48550/arXiv.1704.08430

[6] S. Saini and V. Sahula, "Neural Machine Translation for English to Hindi," 2018 Fourth International Conference on Information Retrieval and Knowledge Management (CAMP), 2018, pp. 1-6, doi: 10.1109/INFRKM.2018.8464781.

[7] Minh-Thang Luong, Hieu Pham, Christopher D. Manning , Effective Approaches to Attention-based Neural Machine Translation , https://doi.org/10.48550/arXiv.1508.04025%20Focus%20to%20learn%20more

[8] Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V. Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun, Yuan Cao, Qin Gao, Klaus Macherey, Jeff Klingner, Apurva Shah, Melvin Johnson, Xiaobing Liu, Łukasz Kaiser, Stephan Gouws, Yoshikiyo Kato, Taku Kudo, Hideto Kazawa, Keith Stevens, George Kurian, Nishant Patil, Wei Wang, Cliff Young, Jason Smith, Jason Riesa, Alex Rudnick, Oriol Vinyals, Greg Corrado, Macduff Hughes, Jeffrey Dean , Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation.

[9] Qiuxiang He, Guoping Huang, Qu Cui, Li Li, and Lemao Liu. 2021. Fast and Accurate Neural Machine Translation with Translation Memory. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), pages 3170–3180, Online. Association for Computational Linguistics.

[10] Deng Cai, Yan Wang, Huayang Li, Wai Lam, and Lemao Liu. 2021. Neural Machine Translation with Monolingual Translation Memory. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 7307–7318, Online. Association for Computational Linguistics.

[11] Rui Wang, Xu Tan, Renqian Luo, Tao Qin, Tie-Yan Liu, "A Survey on Low-Resource Neural Machine Translation", https://arxiv.org/abs/2107.04239

[12] Wei Wang," Translation Mechanism of Neural Machine Algorithm for Online English Resources", https://www.hindawi.com/journals/complexity/2021/5564705/

[13] Guo, Hangcheng & Liu, Wenbin & He, Yanqing & Wu, Zhenfeng & Pan, You & Lan, Tian & Xu, Hongjiao. (2021). ISTIC's Neural Machine Translation System for CCMT' 2021. 10.1007/978-981-16-7512-6_9.

[14] Zhixing Tan, Shuo Wang, Zonghan Yang, Gang Chen, Xuancheng Huang, Maosong Sun, Yang Liu, " Neural Machine Translation: A Review of Methods, Resources, and Tools" https://arxiv.org/abs/2012.15515

[15] Himanshu Choudhary, Shivansh Rao, and Rajesh Rohilla. 2020. Neural Machine Translation for Low-Resourced Indian Languages. In Proceedings of the 12th Language Resources and Evaluation Conference, pages 3610–3615, Marseille, France. European Language Resources Association.