Neural Machine Translation: English to Hindi

Conference Paper · December 2019

DOI: 10.1109/CICT48419.2019.9066238

4 authors, including:

Sahinur Rahman Laskar
University of Petroleum & Energy Studies

Dr. Partha Pakray
National Institute of Technology, Silchar

Sivaji Bandyopadhyay

Neural Machine Translation: English to Hindi

Sahinur Rahman Laskar
Department of Computer Science and Engineering
National Institute of Technology Silchar
Assam, India
sahinurlaskar.nits@gmail.com

Abinash Dutta
Department of Computer Science and Engineering
National Institute of Technology Silchar
Assam, India
abinashdtt45@gmail.com

Partha Pakray
Department of Computer Science and Engineering
National Institute of Technology Silchar
Assam, India
parthapakray@gmail.com

Sivaji Bandyopadhyay
Department of Computer Science and Engineering
National Institute of Technology Silchar
Assam, India
sivaji.cse.ju@gmail.com

***Abstract*—Machine Translation (MT) attempts to minimize the communication gap among people from various linguistic backgrounds. Automatic translation between pair of different natural languages is the task of MT mechanism, wherein Neural Machine Translation (NMT) attract attention because it offers reasonable translation accuracy in case of the context analysis and fluent translation. In this paper, two different NMT systems are carried out, namely, NMT-1 relies on the Long Short Term Memory (LSTM) based attention model and NMT-2 depends on the transformer model in the context of English to Hindi translation. System results are evaluated using Bilingual Evaluation Understudy (BLEU) metric. The average BLEU scores of NMT-1 system are 35.89 (Test-Set-1), 19.91 (Test-Set-2) and NMT-2 system are 34.42 (Test-Set-1), 24.74 (Test-Set-2) respectively. The results show better performance than existing NMT systems.**

***Index Terms*—Machine Translation (MT), Neural Machine Translation (NMT), BLEU score, Attention Mechanism**

I. Introduction

MT acts as a bridge for cross-language communication in Natural Language Processing (NLP). MT handles language perplexity problems using automatic translation in between two languages while preserving its meaning. MT systems developed in a progression of systems from rule-based to corpus-based approach, which wiped out requirement of linguistic expertise, ceaseless list of NLP tasks namely Named Entity Recognition, Parts-of-Speech tagging, Chunking, Word Sense Disambiguation and language diversity problem for Interlingua-based MT [1]. Corpora-based MT systems are broadly classified into Example-Based Machine Translation (EBMT), Statistical Machine Translation (SMT) and NMT. The scope of EBMT is quite limited because, in spite of the large corpus, everything that one wants to translate cannot be covered by examples. Hence SMT came into existence which relies upon Bayesian inference. SMT predicts translation probabilities of phrase pairs in corresponding source-target languages. By increasing the size of the dataset, the probability of a certain pair of phrases can be enhanced. However, the inability to achieve context information, different trained components, system complexity are the weak points of SMT, which led to the development of the NMT system [2], [3]. The NMT system handles sequence to sequence learning problem for variable length source and target sentences and also, handles long term dependency problem using LSTM. The NMT system improves translation prediction and excellent context-analyzing properties. In [11], the NMT system is trained using Nematus [15], in which encoder-decoder RNN used conditional gated recurrent units (GRU) with attention mechanism and obtained BLEU score 12.23 in English to Hindi translation. Similarly, in [12], sequence to sequence RNN used LSTM with attention mechanism for the same translation and acquired BLEU score 23.25, which is top scorer in MTIL-2017 [1]. But both the NMT systems predicted translations did not closely analyze the effect of different n-gram BLEU scores.

The main purpose of current work is to investigate English to Hindi translation using two different NMT systems and analyze on the grounds of different BLEU scores to improve the quality of the existing MT output [11], [12].

The rest of the paper is structured as follows: Effect of bi-gram model is discussed in Section II. Section III, concisely outlines related works. Section IV, elaborates system description. Section V, gives a brief description of corpus and experimental setup. Section VI, details results of automatic as well as manual translators and comparative analysis of results acquired from various systems. Section VII, discussion of current work. Lastly, Section VIII, ends the paper with remarks for future scope.

II. Effect of Bi-gram model on Indian Languages

In n-gram probabilistic language model, bi-gram model predicts occurrence of a word given only its previous word. Given a document written in L language having sequence of words $w_1, w_2, w_3 .... w_n$, the bi-gram model consider that w_2 depends only on w_1, w_3 depends only on w_2 and so on.

[1]https://nlp.amrita.edu/mtil_cen/#results

Mathematically,

$$P(w_1, w_2, w_3....w_n) \approx P(w_n|w_{n-1})$$

In [13], investigated the relationship among various Indian languages using interpolated distance bi-gram models. The bi-gram model was able to capture the interrelation between various Indian languages. They realized the relationship by presenting phylogenetic trees of different languages, where Hindi, Punjabi, Gujarati, and Marathi were in one group and Bengali, Assamese, and Oriya were in another group. Hence, the concept of bi-gram supports well for grouping similar Indian languages. In the current work, it is needed to calculate the BLEU score of predicted translation of Indian language, Hindi w.r.t. its reference sentence. The BLEU scores computed in terms of uni-gram (BLEU-1), bi-gram (BLEU-2) and tri-gram (BLEU-3) (refer to Section VI-A). The reason behind to calculate BLEU score up to tri-gram (BLEU-3) in the context of Hindi language is that the BLEU score decreases to very low when it crosses BLEU-3. Hence, in this work BLEU score restricted upto tri-gram (BLEU-3). The BLEU scores (refer to Section VIII) have been observed closely in the predicted translation of Hindi language to examine the effect of the bi-gram model.

III. Related works

Neural network based translation approach has been used to overcome limitations of SMT such as the issue of accuracy and context analyzing ability. In [4], [5], used Feed-forward neural network to score the phrase pairs by considering fixed size phrases. But in a real-world scenario, the input and output phrase length of the translation may vary. Hence, Recurrent Neural Network (RNN) came into existence to handle variable-length phrases [2]. In [6], proposed an approach using the RNN encoder-decoder system and pointed out that RNN based NMT works fine for short sentences while SMT works fine for long sentences. But, it is difficult to handle complex context-dependent cases by RNN. Hence, RNN adopted LSTM, which was able to learn long-term features for encoding and decoding [7]. Apart from the importance of LSTM, other aspects that improve the effectiveness of the NMT system like the requirement of stacked RNNs, test-time decoding using beam search, input feeding using attention mechanism [8]. In practice, it is difficult to correctly implement all these requirements. Hence, OpenNMT [2], an open-source NMT toolkit was developed [9]. Similarly, Marian [3], a research-friendly toolkit based on transformer model was developed, which was able to achieve high training and translation speed [10]. Both differ in their prioritization of features.

IV. System Description

The primary steps of system operations are data preprocessing, system training and system testing and same have been described in the following subsections.

[2]http://opennmt.net/.

[3]https://marian-nmt.github.io/.

A. Data Preprocessing

Corpus of data has been collected from various sources as mentioned in Section V-A. The key function preprocessing step is to tokenization of source and target sentences and creates a dictionary, which indexes the words present in the training process. The dictionary lists out all unique words ($< unk >$). The validation data set borrowed as specified in Section V-A to examine the convergence of the trained system.

B. System Training

After the dataset is pre-processed, the source and target documents are fed into a sequence-to-sequence RNN having attention mechanism and the system is trained for translation prediction. Multiple Graphics Processing Units (GPUs) are used to boost the performance of training. The system gets trained over a fixed number of iterations, which is known as an epoch. The encoder and decoder constitute main units of the NMT system architecture. The encoder module is responsible for encoding the entire input sequence into a fixed-size context vector or summary vector that contains the necessary features of the input sentence through the input time steps. The unidirectional sequencer acts as an encoder, which consists of a two-layer network of LSTM components contain 500 nodes in each layer. There are various contextual dependencies in a sentence that selectively needs to be remembered or forgotten. This is achieved by the gating mechanism in the LSTM units. Likewise, a decoder consists of a two-layer network of LSTM components having 500 nodes in each layer which is responsible for stepping through the output time steps while reading from the context vector.

a) NMT system with attention mechanism: The loophole in the basic encoder-decoder NMT system is that it compresses the source sentence into a fixed length vector. Hence, if the length of the sentence is too long, then there is a loss of valuable information. The encoder fails to encode all necessary information into the context vector. An attention mechanism is implemented to deal with this issue, which allows the decoder to focus on different parts of the source sequence at different decoding steps [16]. [8] extended the attention model that merges global, accompanying to all source words and local, only pay attention to a part of source words. The main idea of the attention mechanism is to represent each word with a fixed length vector called annotation vectors, in place of a fixed sentence vector. The context vector is calculated using the convex combination of the annotation vectors of the source sentence, are called attention weights. The attention weights compute the importance of the source word corresponding to the target word generation using a alignment model as shown in Equation (1).

$$e_{i,j} = align(z_{i-1}, h_j), \quad \forall j \in 1,2,....T, \quad \forall i \in 1,2,....T^{'} \tag{1}$$

Where $e_{i,j}$ is the alignment score of j-th source word corresponding to i-th target word, z_{i-1} is decoders last state, h_j is

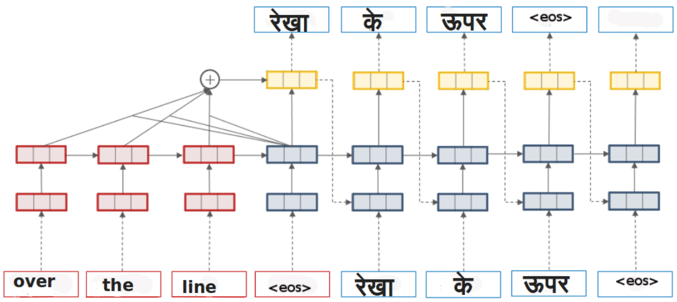

Fig. 1. NMT System Architecture.

the annotation vector of j-th source word and T and T' are the lengths of source and target sentences. Afterward, alignment scores are converted into probabilistic measures, are called attention weights ($\alpha_{i,j}$) as given in the Equation (2).

$$\alpha_{i,j} = \frac{exp(e_{i,j})}{\sum_{j'} exp(e_{j',i})} \quad (2)$$

Finally, the context vector is computed using the annotation vector (c_i) as shown in Equation (3).

$$c_i = \sum_{j=1}^{T} \alpha_{i,j} \times h_j \quad (3)$$

After context vector is computed then calculate decoders next state as a non-linear function (in case of LSTM) of context vector c_i, previous target word μ_{i-1} and decoders last state z_{i-1} using Equation (4).

$$z_i = f(c_i, \mu_{i-1}, z_{i-1}) \quad (4)$$

Figure 1 demonstrates the NMT system, where attention mechanism and input feeding are used to convert source sentence "over the line" into the target sentence " रेखा के ऊपर " [8]. Here, $< eos >$ marks the end of a sentence.

C. System Testing

System training is followed by the system testing/translation process where beam search, an optimized heuristic best first search technique is used to search the finest translations.

V. Experimental Design

This section includes a brief description of corpora and experimental setup deployed to build the NMT systems.

A. Corpus description

The NMT system have been trained using English and Hindi parallel source-target sentence pairs. The parallel corpus contains 14,92,827 number of instances, consists of data sets from a variety of existing sources. Prior system training, corpus data were preprocessed due to the presence of certain constraints as discussed in Section IV-A. The English-Hindi Parallel Corpus of IIT Bombay [11] and the MTIL-2017 [12], contain test corpus 2,507 (Test-Set-1) and 562 (Test-Set-2) English sentences used in the translation process and validate using 4000 instances. The details of the parallel corpus used in training, testing and validation are summarized in Table I.

B. Experimental Setup

Our NMT systems adopted concept of OpenNMT in NMT-1 and Marain NMT in NMT-2 system respectively.

1. Both NMT-1 and NMT-2 system were trained using English-Hindi parallel training corpora as shown in Section V-A to build the train models. The trained models were tested using two different test set as mentioned in Section V-A.

2. The NMT-1 and NMT-2 system were re-trained using the same parallel corpus and saved the trained model by the increase in the number of epochs, acquired at 13, 14 in case of NMT-1 and NMT-2 system respectively. Each of the 13 and 14 models, were tested using Test-Set-1 and Test-Set-2 and predicted results evaluated using BLEU score.

3. The predicted translations of both NMT-1 and NMT-2 system, were validated using sample sentences as mentioned in Section V-A to evaluate the convergence of the training process.

The results of the above experimental setup, have been analyzed in the Section VI respectively.

VI. Results and Analysis

The system results acquired from NMT-1, NMT-2 and translations of Google and Bing systems, have been evaluated using BLEU score [14]. BLEU is a widely accepted standard algorithm for evaluating the quality of the machine-translated text, which uses modified n-gram precision to compare system predicted translation against reference translation. The successive sections present system results and comparative analysis.

A. Results

In each configuration, the BLEU score for the translation are different and Table II, Table III demonstrates the BLEU scores on Test-Set-1 as well as Test-Set-2 respectively, as mentioned in Section V-A for each different configuration.

B. Analysis

Comparison among, our NMT-1 and NMT-2 system, and manual translators namely Google and Bing have been carried out to visualize the effectiveness of various systems. Figure 2 and 3 show the comparison of different systems. The comparative results have pointed out that our NMT-1 and NMT-2 system provide higher BLEU-1 score than the Top scorer in MTIL-2017 [12] as well as other NMT system used in [11]. But, Google and Bing translators outperforms our NMT systems.

To further analyze the performance of our NMT systems, the sample sentences from the predicted translation of MTIL-2017, have taken to examine best, average and worst performance from different perspectives. In the case of a short, medium and long sentence as shown in Example 1, 2 and 3 of Table IV, V, VI, our NMT-1 and NMT-2 systems provide perfect prediction like Google and Bing systems for the given test sentence. The average performance of our NMT systems

TABLE I
CORPUS DESCRIPTION.

Nature of corpus	Name of corpus	Number of instances
Training	OPUS [4]	10,00000
	HindEnCorp [5]	2,73,885
	Hindi_MTIL2017-Training [6]	1,60,758
	Bible [7]	30,000
	Different Indian Government websites	28,184
	Total	14,92,827
Test	Test-Set-1 [8]	562
	Test-Set-2 [9]	2,507
Validation (Hindi)	Hindi_MTIL2017-Training [10]	4000

TABLE II
BLEU SCORES ON TEST-SET-1 (MTIL-2017) USING VARIOUS TOOLS.

Tools/Manual Translators	BLEU-1	BLEU-2	BLEU-3	AVG BLEU
Our NMT-1	54.48	**32.25**	20.95	**35.89**
Our NMT-2	53.35	**30.72**	19.2	**34.42**
Google Translator	69.14	**50.49**	38.46	**52.69**
Bing Translator	59.56	**38.46**	26.45	**41.49**
Top Scorer (MTIL-2017)	23.25	-	-	-

TABLE III
BLEU SCORES ON TEST-SET-2 (IIT BOMBAY ENGLISH-HINDI PARALLEL CORPUS) USING VARIOUS TOOLS.

Tools/Manual Translators	BLEU-1	BLEU-2	BLEU-3	AVG BLEU
Our NMT-1	38.17	**15.03**	6.54	**19.91**
Our NMT-2	45.35	**19.89**	8.99	**24.74**
Google Translator	57.27	**31.39**	18.33	**35.66**
Bing Translator	51	**26.11**	14.58	**30.56**
Other NMT System [11]	12.23	-	-	-

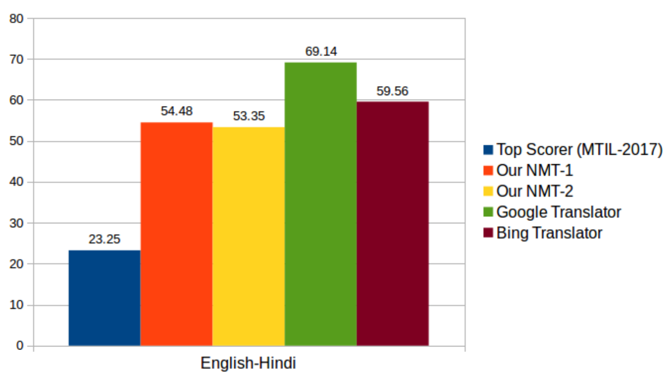

Fig. 2. Comparison of BLEU scores among various systems on Test-Set-1.

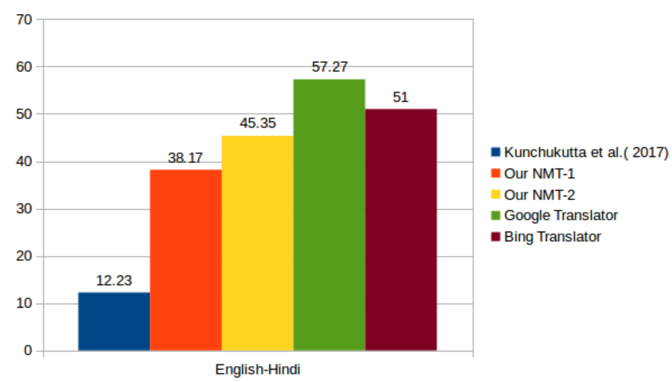

Fig. 3. Comparison of BLEU scores among various systems on Test-Set-2.

as shown in Example 4 and 5 of Table VII, VIII, where one NMT system perform well and other perform poorly. In Example 4, NMT-1 system makes a perfect prediction like Google and Bing systems for the given source sentence. Nevertheless, the interrogative mark (?) and translation of source word “road” are absent in the reference sentence. Our NMT-1 system accurately recognizes the interrogative nature of the sentence and keeps an interrogative mark at the appropriate place. But our NMT-2 system prediction is not appropriate for the translation of source word “road” and also it missed the interrogative mark. However, our NMT-2 system yields a precise prediction in the latter example, but our NMT-1 system unable to recognize the unknown word in the translation of the source sentence. In worst case scenario, Example 6, 7, 8 of Table IX, X, XI are presented. In Example 6, our NMT-1 system fails to predict unknown word and also, translates the source word “Nepal” two times. Also, our NMT-2 system prediction is not accurate since it excludes the translation of source word “Nepal”. In Example 7, our NMT-1 system fails to predict appropriate tense and NMT-2 system predicts the wrong word during translation. Moreover, in the case of a long sentence as shown in Example 8, our NMT-1 and NMT-2 system prediction is not accurate.

TABLE IV
EXAMPLE 1: BEST PERFORMANCE ON TEST-SET-1 (MTIL 2017).

Source	**It reads all right .**
Our NMT-1	यह सब ठीक है ।
Our NMT-2	यह सब ठीक है ।
Google	यह सब ठीक है ।
Bing	यह सब ठीक पढ़ता है ।
Reference	**यह सभी ठीक है ।**

TABLE V
EXAMPLE 2: BEST PERFORMANCE ON TEST-SET-1 (MTIL 2017).

Source	**It is not at the centre of the paper.**
Our NMT-1	यह कागज के केंद्र में नहीं है ।
Our NMT-2	यह केन्द्र में स्थित नहीं है।
Google	यह कागज के केंद्र में नहीं है।
Bing	यह कागज के केंद्र में नहीं है।
Reference	**यह पत्र के केंद्र में नहीं है।**

TABLE VI
EXAMPLE 3: BEST PERFORMANCE ON TEST-SET-1 (MTIL 2017).

Source	**But when my doctor made the decision that my disease had advanced to the point of needing more intensive care**
Our NMT-1	लेकिन जब मेरे डाक्टर ने निर्णय दिया कि मेरी बीमारी अधिक सघन देखभाल की आवश्यकता के लिए आगे बढ़ी है ।
Our NMT-2	लेकिन जब मेरे डाक्टर ने यह निर्णय दिया कि मेरी बीमारी जिस बिंदु तक उन्नत हो चुकी है , उस समय मैं अधिक गहन तन्द्रा की आवश्यकता महसूस कर रहा था ।
Google	लेकिन जब मेरे डॉक्टर ने निर्णय लिया कि मेरी बीमारी अधिक गहन देखभाल की आवश्यकता के बिंदु पर आगे बढ़ी है।
Bing	लेकिन जब मेरे डॉक्टर का निर्णय है कि मेरी बीमारी और अधिक गहन देखभाल की जरूरत के मुद्दे पर उंनत किया था बनाया है ।
Reference	**लेकिन जब मेरे डॉक्टर ने निर्णय किया था कि अपनी बीमारी को अधिक गहन देखभाल की दृष्टि से अधिक गहन ध्यान दिया गया था।**

TABLE VII
EXAMPLE 4: AVERAGE PERFORMANCE ON TEST-SET-1 (MTIL 2017).

Source	**Do you see the man who is running across the road ?**
Our NMT-1	क्या तुम उस आदमी को देखते हो जो सड़क पार कर रहा हो ?
Our NMT-2	क्या आप उस आदमी को देखते हैं जो पीठ के पार चल रहा है
Google	क्या आप उस आदमी को देखते हैं जो सड़क के पार चल रहा है?
Bing	क्या आप सड़क पार चल रहे आदमी को देखते हैं?
Reference	**के पार चलने वाले आदमी को आप देख सकते हैं**

[4]http://opus.nlpl.eu/.
[5]http://www.cfilt.iitb.ac.in/iitb_parallel/.
[6]https://nlp.amrita.edu/mtil_cen/.
[7]https://www.bible.com/.
[8]https://nlp.amrita.edu/mtil_cen/
[9]http://www.cfilt.iitb.ac.in/iitb_parallel/.
[10]https://nlp.amrita.edu/mtil_cen/.

TABLE VIII
EXAMPLE 5: AVERAGE PERFORMANCE ON TEST-SET-1 (MTIL 2017).

Source	**He executed quite a few major architectural and engineering works for domed churches and cathedrals.**
Our NMT-1	उन्होंने <unk> चर्च और <unk> के लिए बहुत कुछ <unk> और इंजीनियरी कार्य किया ।
Our NMT-2	उन्होंने काफी महत्वपूर्ण स्थापत्य और इंजीनियरिंग भवनों का निर्माण किया ।
Google	गुंबददार चर्चों और गिरिजाघरों के लिए उन्होंने कुछ प्रमुख वास्तुशिल्प और इंजीनियरिंग कार्यों का निष्पादन किया।
Bing	उसने गुंथे हुए चर्चों और कैथेडेलों के लिए बहुत कुछ प्रमुख वास्तुशिल्प और इंजीनियरिंग कार्य निष्पादित किए ।
Reference	**उन्होंने काफी कुछ प्रमुख वास्तुशिल्पीय और अभियांत्रिकी गिरजाघरों और कैथेड्रल के काम का परिचय दिया ।**

TABLE IX
EXAMPLE 6: WORST PERFORMANCE ON TEST-SET-1 (MTIL 2017).

Source	**The region shares its borders with china, nepal, bhutan and pakistan.**
Our NMT-1	चीन नेपाल , नेपाल , <unk> और पाकिस्तान के साथ इसकी सीमाओं का हिस्सा है
Our NMT-2	यह क्षेत्र भूटान और पाकिस्तान के साथ अपनी सीमाओं कों छूते हैं ।
Google	यह क्षेत्र चीन, नेपाल, भूटान और पाकिस्तान के साथ अपनी सीमाएँ साझा करता है।
Bing	यह क्षेत्र चीन, नेपाल, भूटान और पाकिस्तान के साथ अपनी सीमाएं साझा करता है ।
Reference	**इस क्षेत्र को चीन , नेपाल , भूटान और पाकिस्तान के साथ शेयर करता है।**

TABLE X
EXAMPLE 7: WORST PERFORMANCE ON TEST-SET-1 (MTIL 2017).

Source	**He was taken aback.**
Our NMT-1	वह चौंक गया ।
Our NMT-2	उसे फाँसी दी गयी ।
Google	उसे दबोच लिया गया।
Bing	वह दंग रह गए ।
Reference	**वह हैरान हो गया था।**

TABLE XI
EXAMPLE 8: WORST PERFORMANCE ON TEST-SET-1 (MTIL 2017).

Source	**I have spent the past 8 years of my life overweight not being able to participate in fun things happening around me .**
Our NMT-1	मैंने अपने जीवन के पिछले 8 साल की उम्र में अपने जीवन में जो कुछ भी भाग लेने में समर्थ न हो , उसके बारे में अधिक खर्च किया है ।
Our NMT-2	मैंने अपने जीवन के पिछले 8 वर्षों का भार अपने वजन के साथ बिताया है ।
Google	मैंने अपने जीवन के पिछले 8 वर्षों को अधिक वजन के साथ बिताया है जो मेरे आस-पास होने वाली मजेदार चीजों में भाग लेने में सक्षम नहीं हैं।
Bing	मैं अपने जीवन अधिक वजन के पिछले 8 साल बिताए है मज़ा मेरे आसपास हो रहा चीजों में भाग लेने में सक्षम नहीं किया जा रहा है ।
Reference	**मैं अपने जीवन के करीब पंद्रह साल का खर्च नहीं कर रहा हूँ , जो मेरे आस - पास होने वाली चीजों में शामिल हो रही हैं ।**

VII. DISCUSSION

In the current work, we have observed that our NMT systems, NMT-1 which is LSTM based encoder-decoder RNN with attention mechanism and NMT-2 which is based on the transformer model having self-attention mechanism yield higher BLEU score than existing systems [11], [12]. The reason behind is that our NMT-1, NMT2 systems have been trained on more instances than the existing NMT system [12]. Also, our NMT-2 system has adopted the transformer model, which reason about outperforms [11], [12]. The NMT system of the existing work [11] is based on conditional GRU with attention mechanism. The main goal of using LSTM or GRU is that to track long-term dependencies effectively. The LSTM performs the activity of tracking long-term dependencies via input, forget and output gates while the GRU operates through a reset gate and an update gate. LSTM controls the exposure of cell state via separate input and forgets gates while GRU exposes the entire cell state to other units in the network via a single reset gate. The attention mechanism recognizes parts of the input sequence that are relevant to each word in the output and incorporate relevant information to select the appropriate output as mentioned in Section IV-B. Our NMT system, NMT-1, LSTM based attention mechanism improve MT output of GRU based NMT system [11] in the context of English to Hindi translation. Moreover, we have analyzed the performance of MT output in the context of Indian language, Hindi on different BLEU scores i.e. BLEU-1, BLEU-2, BLEU-3. It is observed that in the case of Indian language like Hindi, it is not sufficient to analyze only on the ground of BLEU-1 score like existing works [11], [12]. It is realized that obtained average BLEU score in each configuration of our NMT systems are closed to the respective bi-gram BLEU score as shown in Table II and Table III.

VIII. Conclusion and Future Work

Although, NMT system provides better accuracy than conventional approaches like rule-based MT as well as SMT but it still lags behind compared to manual human translation. In this paper, our both NMT systems namely, sequence-to-sequence RNN amd transformer based model are used for English to Hindi translation and compared with the existing MT output of [11], [12] in terms of BLEU score. It shows better performance than the existing systems. However, close analysis of predicted translations remarks that our NMT systems need to be improved in case of recognition of unknown word, blank lines in output and diverse translation of the source sentence. Besides, with the realization of the effect of the bi-gram model in the Hindi language translation and relationship among Indian similar languages in [13], opens a new direction of research for direct translation between pairs of similar languages. By taking the advantage of the similarity between languages, it might be possible to overcome the limitation of available parallel data in case of low resource languages to produce accurate output.

Acknowledgement

Current research work is sponsored by TEQIP-III and authors would like to thank organizers of the CICT19 and also, thank Centre for Natural Language Processing (CNLP) and Department of Computer Science and Engineering at National Institute of Technology, Silchar for providing the requisite support and infrastructure to execute this work.

References

[1] Shachi Dave, Jignashu Parikh, and Pushpak Bhattacharyya, “Interlingua-based englishhindi machine translation and language divergence,” Machine Translation, vol. 16(4), pp.251–304 (2001).

[2] Kyunghyun Cho, Bart van Merrienboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio, “Learning phrase representations using rnn encoderdecoder for statistical machine translation,” In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), Association for Computational Linguistics, Doha, Qatar, pp. 1724–1734 (2014).

[3] lya Sutskever, Oriol Vinyals, and Quoc V. Le, “Sequence to sequence learning with neural networks,” In proceedings of the 27th international conference on neural information processing systems - Volume 2. MIT Press, Cambridge, MA, USA, NIPS14, pp. 3104–3112 (2014).

[4] Holger Schwenk, “Continuous space translation models for phrase-based statistical machine translation,” In Proceedings of COLING 2012: Posters. The Coling 2012 Organizing Committee, Mumbai, India, pp. 1071–1080 (2012).

[5] Jacob Devlin, Rabih Zbib, Zhongqiang Huang, Thomas Lamar, Richard Schwartz, and John Makhoul, “Fast and robust neural network joint models for statistical machine translation,” In proceedings of the 52nd annual meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, Baltimore, Maryland, pp. 1370–1380 (2014).

[6] Sainik Kumar Mahata, Dipankar Das, and Sivaji Bandyopadhyay, “Mtil2017: Machine translation using recurrent neural network on statistical machine translation,” Journal of Intelligent Systems pp. 1–7 (2018).

[7] Sandeep Saini and Vineet Sahula, “ Neural machine translation for english to hindi ,” In proceedings of the conference: 2018 fourth international conference on information retrieval and knowledge management (CAMP), Sabah, Malaysia (2018).

[8] Thang Luong, Hieu Pham, and Christopher D. Manning, “Effective approaches to attention-based neural machine translation,” In proceedings of the 2015 conference on empirical methods in natural language processing, Association for Computational Linguistics, Lisbon, Portugal, DOI: 10.18653/v1/D15-1166, pp. 1412–1421 (2015).

[9] Guillaume Klein, Yoon Kim, Yuntian Deng, Jean Senellart, and Alexander Rush, “Open-nmt:Open-source toolkit for neural machine translation,” In proceedings of ACL 2017, system demonstrations, Association for Computational Linguistics, Vancouver, Canada, pp. 67–72 (2017).

[10] Marcin Junczys-Dowmunt, Roman Grundkiewicz, Tomasz Dwojak, Hieu Hoang, Kenneth Heafield, Tom Neckermann, Frank Seide, Ulrich Germann Alham Fikri Aji, Nikolay Bogoychev, Andre F. T. Martins, and Alexandra Birch, “Marian: Fast neural machine translation in C++,” In proceedings of ACL 2018, system demonstrations, Association for Computational Linguistics, Melbourne, Australia, pp. 116–121 (2018).

[11] Anoop Kunchukuttan, Pratik Mehta, and Pushpak Bhattacharyya, “The IIT bombay english-hindi parallel corpus,” In proceedings of the eleventh international conference on language resources and evaluation (LREC 2018), May 7-12, 2018, Miyazaki, Japan, isbn: 979-10-95546-00-9, (2018).

[12] Amarnath Pathak and Partha Pakray,“Neural machine translation for Indian languages,” Journal of Intelligent Systems (JISYS), DOI: https://doi.org/10.1515/jisys-2018-0065 (2018).

[13] Sayan Ghosh, K V Vijay Girish and Tv Sreenivas, “Relationship between indian languages using long distance bigram language models,” In proceedings of ICON-2011: 9th international conference on natural language processing, Macmillan, pp. 104–113 (2011).

[14] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu,“Bleu: A method for automatic evaluation of machine translation,” In proceedings of the 40th annual meeting on Association for Computational Linguistics, Stroudsburg, PA, USA, ACL 02, pp. 311–318 (2002).

[15] Rico Sennrich, Orhan Firat, Kyunghyun Cho, Alexandra Birch, Barry Haddow, Julian Hitschler, Marcin Junczys-Dowmunt, Samuel Läubli, Antonio Valerio Miceli Barone, Jozef Mokry and Maria Nadejde,“Nematus: a Toolkit for Neural Machine Translation”, In proceedings of the EACL 2017 software demonstrations, Valencia, Spain, April 3-7 2017, pp. 65–68, Association for Computational Linguistics (2017).

[16] Dzmitry Bahdanau, Kyunghyun Cho, Yoshua Bengio,“Neural Machine Translation by Jointly Learning to Align and Translate,” In 3rd international conference on learning representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, conference track proceedings (2015).