International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 05 Issue: 04 | Apr-2018 www.irjet.net p-ISSN: 2395-0072

Survey and Analysis on Language Translator Using Neural Machine Translation

Ms. Neeta Verma[1], Abhay Jain[2], Animesh Basak[3], Kshitij Bharti Saksena[4]

[1]Associate professor, Dept. of Computer Science Engineering, Inderprastha Engineering College, Uttar Pradesh, India
[2,3,4] Student, Dept. of Computer Science Engineering, Inderprastha Engineering College, Uttar Pradesh, India

---

ABSTRACT: *Neural machine translation or NMT is a new proposed defined approach to the machine translation. Unlike the traditional SMT i.e. statistical machine translation, NMT focuses on constructing a single neural network that can be jointly aligned to maximize the performance, translation and efficiency. The models that are proposed for NMT belongs to a group of encoders and decoders and encode a source text or sentence into a vector of fixed length from which a decoder generates an appropriate translation.*

This paper discussed different approaches for language translation. The HMM based encoder decoder models are used in the survey. The comparison in NMT and SMT is also analysed in this survey paper

Keywords: SMT, NMT, Encoder, Decoder, Language translator

I. INTRODUCTION:

Neural machine Translation (NMT) is an empirical approach to the known process of Machine Translation that uses a large artificially constructed neural network to predict or to know the occurrence of a large sequence of words, basically modelling or generating the entire sentences in a single integrated model. It has gained adoption in many large-scale functionalities. NMT systems take advantage of continuous representations that greatly ease the sparsity problem, and make use of much larger contexts, thus lessen the locality problem. Many issues and shortcomings of the traditional machine translation systems is eradicated by the new approach i.e. NMT. Deep Neural Machine Translation i.e. Deep NMT is an extension of NMT. Both of them uses a large neural network with the only difference that exist is that deep neural machine translation processes multiple neural network layers instead of just one as it incorporates higher proficiency compared to NMT. Deep learning applications was first appeared in speech recognition in 1990s. The first scientific research paper on neural networks in machine translation appeared in 2014, followed by many advances in the following few years. (Large-vocabulary NMT, application to Image captioning, Subword-NMT, Multilingual NMT, Multi-Source NMT, Character-dec NMT, Zero-Resource NMT, Google, Fully Character-NMT, Zero-Shot NMT in 2017). In 2015, the first appearance of a NMT system that was in a public machine translation competition (OpenMT'15). All of the parts of the NMT model are trained jointly i.e. (end-to-end) to maximize the translation performance. NMT models use deep learning and representation learning. The word sequence modelling was at first typically done using a RNN Recurrent Neural Network. A bidirectional recurrent neural network which was used known as an *encoder* used by the neural network to encode a source sentence and a second RNN known as a *decoder*, that is used to predict words or sentences in the target language was used.

II. LITERATURE REVIEW:

This section provides the brief description of various research papers studied for this study. The given below table 1 represents the summarization of various methods applied for Neural Machine Translation.:

TABLE I: METHODS APPLIED FOR LANGUAGE TRANSLATION

S.no	Topic	Technology	Conclusion	Reference
1	Speechalator: Two-way speech-to-speech translation on a consumer PDA	HMM-based recognition, interlingua translation (both rule and statistically	Alex Waibel et. al [1] shows a working two-way speech-to-speech translation system that works in real-time on the user's computer. It can translate text from English language to Arabic language and vice-versa .	http://www.aclweb.org/anthology/N03-4015

	By Alex Waibel, et.al [1]	based), and unit selection synthesis		
2	Speech to Speech Language Translator By Umeaz Kheradia and Abha Kondwilkar	The input speech first goes to the speech IC (HM2007) of the speech processing unit. This IC works in two modes: · Training: Stores the database. ·Recognition: Compares with the database.	Umeaz Kheradia et.al [2] describes a prototype which uses a speech processing hardware & online translator to provide the user with real time translation. Speech processing hardware works on the principle of 'compare / analyse and then forward', i.e., a stored database is in the system is used for comparing with the input or the user's speech and then the result is sent forward for further processing.	www.ijsrp.org /research-paper-1212/ijsrp-p1242.pdf
3	Neural Machine Translation by Jointly Learning To Align And Translate By Dzmitry Bahdanau, KyungHyun Cho and Yoshua Bengio	(Soft)alignment generated by the RNN search Encoder–Decoders Model	Dzmitry Bahdanau et.al [3] in this paper, proposed a novel architecture that addresses this issue. They extended the basic encoder and decoder model by introducing a model soft search mechanism for a set of input words, when generating each target word. Hence allowing the model from having to encode a full source sentence into a vector of fixed length, and also allows the model to focus only on the information relevant to the generation of the next target word. This has a major positive impact on the ability of the neural machine translation system to yield better results for longer sentences.	https://arxiv.org/pdf/1409.0473
4	English to Sanskrit Transliteration: an effective approach to design Natural Language Translation Tool By Leena Jain and Prateek Agrawal	Transliterating: Algorithm which will automatically convert the text typed in English to Hindi language	Leena Jain et.al [4] developed a transliteration tool that translates English to Sanskrit. Transliteration is the process of converting the letters of typed text in one language to the letters of another language. The methodology used is to design an algorithm that uses Unicode for transliteration. The Unicode for English and Hindi are mapped to each other. The input is taken in English and the letters are matched to Hindi through the mapped Unicode. The Output is the text in Hindi. Result and Conclusion All test cases passed. 100% accuracy. The tool can be used for ML and Natural Language Translations. The interface is user-friendly.	www.ijarcs.info/index.php/Ijarcs/article/download/2860/2843
5	Six Challenges for Neural Machine Translation By Philipp Koehn and Rebecca Knowles	NMT and SMT	Philipp Koehn et.al [5] in this paper provides a contrast between NMT (Neural Machine Translation) and SMT (Statistical Machine Translation) , They used common toolkits for NMT (Nematus) and traditional phrase-based statistical machine translation (Moses) with common data sets, drawn from WMT and OPUS. They carry out experiments on English–Spanish and German–English as for these language pairs, large training data sets are available. Found out that a known challenge in translation is that in different domains, 6 words have different translations and meaning is expressed in different styles. They trained both NMT and SMT systems for all domains	https://arxiv.org/pdf/1706.03872.pdf

			When translating took place they found that the output of the NMT system is often quite fluent but completely unrelated to the input while the SMT output betrays its difficulties with coping with the out-of-domain input by leaving some words untranslated. They showed that, despite the recent successes, NMT still has to overcome various challenges, most notably performance out of domain and under low resource conditions. What a lot of the problems have in common is that the neural translation models do not show robust behaviour when confronted with conditions that differ significantly from training conditions.	
6	Neural Machine Translation of Indian Languages By Karthik Revanuru et.al [6]	NMT BY DIFFERENT APPROACHES INCLUDING LSTM, Bi-LSTM, SGD, ATTENTION MECHANISM AND OTHER APPROACHES	Karthik Revanuru et.al [6] worked on neural machine translation on Indian languages This is the first work to have applied NMT on Indian language pairs. They trained their models using 8 different configurations and evaluated them using five different standard and evaluation metrics. Their models had simpler architecture, require fewer sources and take less time. Their best model outperformed Google Translate by a margin of 17 BLEU score	https://www.researchgate.net/publication/319351932_Neural_Machine_Translation_of_Indian_Languages
7	Google's Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation By Melvin Johnson, et.al [7].	NMT and Multilingual model architecture	Melvin Johnson et.al [7] use a single Neural Machine Translation (NMT) model to translate between multiple languages. There is no change to the default model architecture from the standard NMT system but instead they introduces an artificial token at the beginning of the input sentence to denote the required target language. The model includes an encoder, decoder and attention module, that remains unchanged and is shared across all languages. They used a shared word piece vocabulary. Their approach enables Multilingual NMT using a single model without any increase or addition in parameters, which is significantly simpler than previous proposals for Multilingual NMT. In addition to make a better translation quality of language pairs that the model was trained with, their models can also learn to play implicit bridging between language pairs never seen before or during training, showing that transfer learning and zero-shot translation is possible for neural translation.	https://arxiv.org/pdf/1611.04558.pdf
8	Neural Machine Translation of Rare Words with Subword Units By Rico Sennrich, et.al [8]	Byte Pair Encoding is a simple data compression technique that iteratively replaces the most frequent pair of bytes in a sequence with a single, unused byte.	Rico Sennrich et.al [8] introduced a easier and more effective approach, making the NMT model capable of open-vocabulary translation by encoding rare and unknown words as sequences of sub word unit as NMT models typically operate with a fixed vocabulary, but translation is an open-vocabulary problem. Their Previous work addresses the translation of words that are not in vocabulary by backing off to a dictionary.	https://arxiv.org/pdf/1508.07909.pdf

9	UNSUPERVISED NEURAL MACHINE TRANSLATION By Mikel Artetxe, et.al [9]	UNSUPERVISED CROSS-LINGUAL EMBEDDINGS and STATISTICAL DECIPHERMENT FOR MACHINE TRANSLATION	Mikel Artetxe et.al [9] completely removed the need of parallel data and provided a novel method to train an NMT system in a unsupervised manner, relying only on monolingual corpora. Their recent work on unsupervised embedding mappings and consists of a slightly modified attentional encoder-decoder model that can be trained on monolingual corpora alone using a combination of denoising and backtranslation. The model can also profit from small parallel corpora, and attains 21.81 and 15.24 points when combined with 100,000 parallel sentences, respectively.	https://arxiv.org/pdf/1710.11041.pdf

III.METHODOLOGY:

For this research work Neural Machine Translation (NMT) [13] is taken as methodology. NMT introduced as a new approach with the capability of addressing many shortcomings of traditional machine translation systems. The enefit or advantage of NMT lies in its ability to learn precisely, i.e. mapping of input text to associated output text. Its architecture consists of two recurrent neural networks (RNNs), one is use to absorb the input text sequence i.e. encoder and one is use to generate translated output text i.e. decoder.

As a human, we read the full source sentence or text, then understand its meaning, and then provide a translation. Neural Machine Translation (NMT) mimics that!

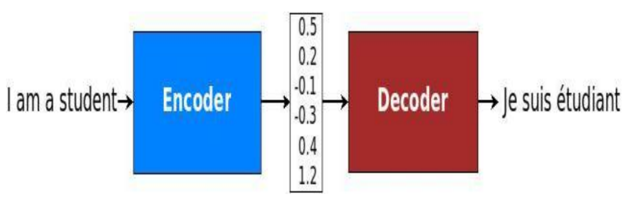

Figure 1. Encoder-decoder architecture .

An encoder converts a input sentence into a "meaning" vector which is passed through a *decoder* to provide a translation.

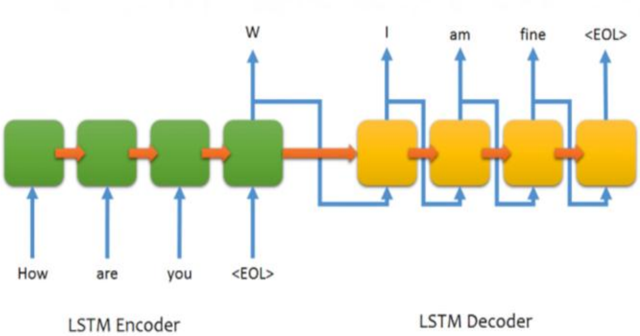

Figure 2. LSTM Encoder Decoder Model

Specifically, an NMT system first reads the input sentence using an encoder to build a "thought" vector , a sequence or series of numbers that represents the sentence meaning; a decoder, then, processes and provides the sentence vector to produce a translation, as illustrated in Figure 1. This is often referred to as the encoder-decoder architecture.

NMT models vary in terms of their exact architectures. An obvious choice for the sequential data is the RNN i.e. Recurrent Neural Network, which is used by most NMT models. Usually an RNN is used for both the encoder and decoder. These models, however, differ in terms of:

(a) **directionality** –i.e. can be unidirectional or bidirectional;

(b) **depth** – i.e. can be single- or multi-layer;

(c) **type** – often either a vanilla RNN, a Long Short-term Memory (LSTM), or a gated recurrent unit (GRU).

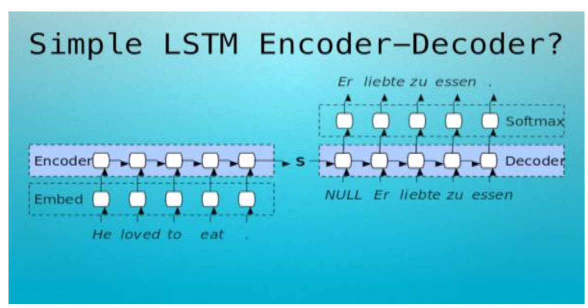

Figure 3. (a) Directionality

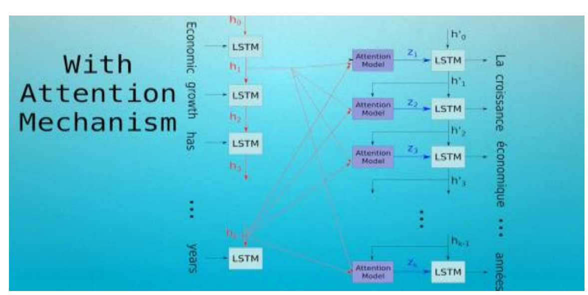

Figure 3. (b)Depth

A deep multi-layer Recurrent Neural Network which is unidirectional and uses LSTM as a recurrent unit. At a high level, the NMT model consists of two recurrent neural networks: the encoder RNN simply consumes the input source words without making any prediction; the decoder, on the other hand, processes the target sentence while predicting the next words.

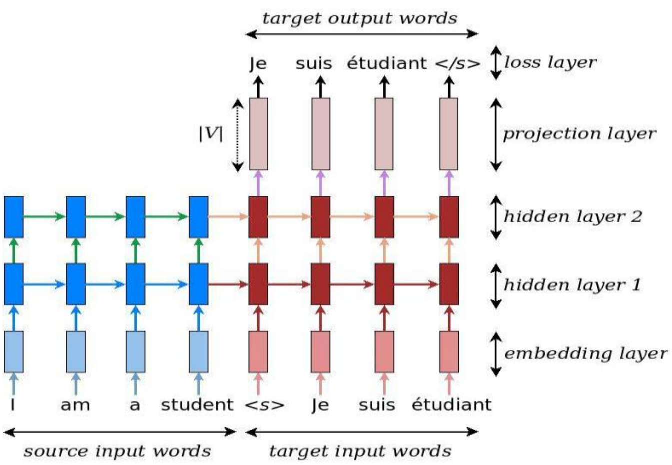

Figure 4. Neural machine translation

example translating a source sentence "I am a student" into a target sentence "Je suis étudiant". Here, "<s>" denotes the start of the decoding process while "</s>" informs the decoder to stop.

NMT often lead by an attention mechanism, which helps it cope effectively with long input sequences. Advantage of NMT is that it eludes many delicate design choices in traditional phrase-based machine translation.

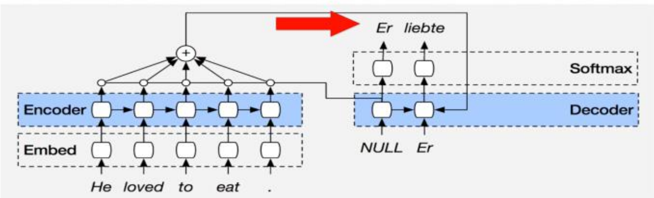

Figure 5. (a)Single Layer Encoder Decoder with attention mechanism

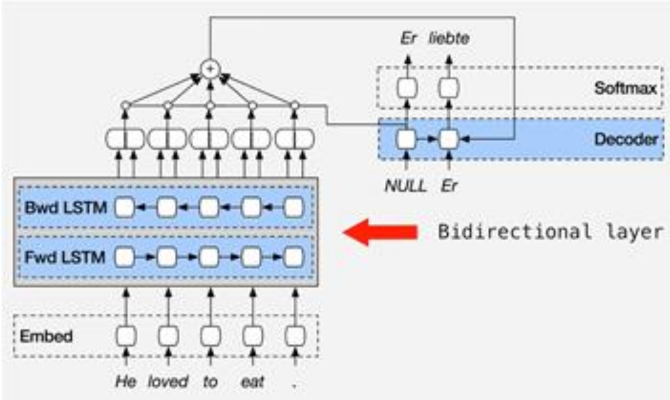

Figure 5. (b) Bidirectional Encoder Decoder with attention mechanism. It's two RNNs. One that goes forward over the sentence and the other goes backwards. So for each word it concatenates the vector outputs which produces a vector with context from both sides.

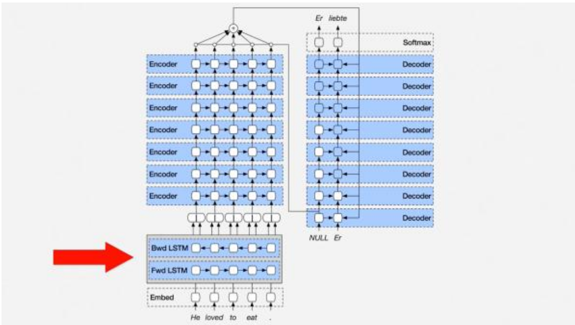

Figure 5. (c) Bidirectional Encoder Decoder with attention mechanism . The encoder has one bi-directional RNN layer and seven unidirectional RNN layers. The decoder has eight unidirectional RNN layers.

The more layers the longer the training times so that's why we use a single bi-directional layer ,if all the layers were bi-directional the whole layer would have to finish before layer dependencies could start computing But by using unidirectional layers, computation is going to be more parallel.

IV.CONCLUSION / FUTURE WORK:

In practice, however, NMT systems used to be better in accuracy than phrase-based translation systems, especially when training on very large-scale datasets as used for the very best publicly available translation systems.

Weaknesses of NMT are responsible for this gap: slower training and inference speed, ineffectiveness in dealing with rare words, and sometimes failure to translate all words in the source sentence.

Firstly, it takes considerable amount of time and computational resources to train an NMT system on a large-scale translation dataset, thus slowing the rate of experimental turnaround time and innovation. For inference, they are generally much slower than phrase-based systems due to the large number of parameters used.

Secondly, NMT lacks robustness in translating rare words. Though this can be addressed in principle by training a "copy model" to mimic a traditional alignment model, or by using the attention mechanism to copy rare words, these approaches are both unreliable at scale, since the quality of the alignments varies across languages, and the latent alignments produced by the attention mechanism are unstable when the network is deep. In addition, simple copying may not always be the best strategy to cope with rare words, for example, when a transliteration is more appropriate. Finally, NMT systems sometimes produce output sentences that do not translate all parts of the input sentence – in other words, they fail to completely cover the input, which can result in surprising translations.

In implementation, the recurrent networks are Long Short-Term Memory (LSTM) RNNs. LSTM RNNs have 8 layers, with residual connections between layers to encourage gradient flow. For parallelism, we connect the attention from the bottom layer of the decoder network to the top layer of the encoder network. To improve inference time, we employ low-precision arithmetic for inference, which is further accelerated by special hardware. To effectively deal with rare words, we use sub-word units for inputs and outputs in our system. Using sub-words gives a good balance between the flexibility of single characters and the efficiency of full words for decoding, and also eludes the need for special treatment of unknown words. Experiments suggest the quality of the resulting translation system gets closer to that of average human translators.

So we can conclude that after reviewing different papers the work done by Melvin Johnson et. al [7] was remarkably the best as the use of zero shot translation results the best output and efficient result.

Zero shot learning[15] refers to computers recognizing or "learning" new concepts without previous knowledge of them.

"…a multilingual NMT model trained with Portuguese>English and English>Spanish examples can generate reasonable translation for Portuguese>Spanish although it has not seen any data for that language pair. We show that the quality of zero shot language pairs can easily be improved with little additional data of the language pair in question," the paper states.

According to Google, the research indicates the "first demonstration of true multilingual zero-shot translation."

The ability to base multiple language pairs off of a single NMT model means Google will be able to drastically cut down on the number of machine translation models it needs to create. Google Translate works in more than 100 languages and would theoretically have to create thousands of translation models if it weren't for the single model system.

According to Google, the method improves the translation quality of "low-resource languages," too. Low resource languages are those that don't contain a good amount of reference material for translation.

They show that zero-shot translation without explicit bridging is possible, which is in a form of true transfer learning has been shown to work for machine translation

V. REFRENCES:

[1] A. Waibel, A. Badran, A. W Black, R. Frederking, D. Gates, A. Lavie, K. Lenzo, L. Tomokiyo, J. Reichert, T. Schultz, D. Wallace, M. Woszczyna and J. Zhang, "Speechalator: two-way speech-to-speech translation on a consumer PDA", EUROSPEECH 2003 – GENEVA.

[2] U. Kheradia and A. Kondwilkar, "Speech To Speech Language Translator", International Journal of Scientific and Research Publications, Volume 2, Issue 12, December 2012 1 ISSN 2250-3153.

[3] D. Bahdanau, K. Cho and Y. Bengio, "NEURAL MACHINE TRANSLATION BY JOINTLY LEARNING TO ALIGN AND TRANSLATE", Published as a conference paper at ICLR 2015.

[4] L. Jain and P. Agrawal, "English to Sanskrit Transliteration: an effective approach to design Natural Language Translation Tool", International Journal of Advanced Research in Computer Science, Volume 8, No. 1, Jan-Feb 2017

[5] P. Koehn and R. Knowles, "Six Challenges for Neural Machine Translation", Proceedings of the First Workshop on Neural Machine Translation, pages 28–39(2017).

[6] K. Revanuru, K. Turlapaty, S. Rao, "Neural Machine Translation of Indian Languages" Published as a Conference paper at COMPUTE 2017: 10th Annual ACM India Conference, At Bhopal, India, November 2017

[7] M. Johnson, M. Schuster, Q. V. Le, M. Krikun, Y. Wu, Z. Chen and N. Thorat, "Google's Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation",(2017).

[8] R. Sennrich, B. Haddow and A. Birch, "Neural Machine Translation of Rare Words with Subword Units" , Submitted on 31 Aug 2015 (v1), last revised 10 Jun 2016 (this version, v5)) , The research presented in this publication was conducted in cooperation with Samsung Electronics Polska sp. z o.o. -Samsung R&D Institute Poland.

[9] K. Cho, M.Artetxe, G. Labaka & E. Agirre, "UNSUPERVISED NEURAL MACHINE TRANSLATION", Published as a conference paper at ICLR 2018.

[10] G. Lample, L. Denoyer and M. Ranzato, "UNSUPERVISED MACHINE TRANSLATION USING MONOLINGUAL CORPORA ONLY", Under review as a conference paper at ICLR 2018.

[11] B. Turovsky, "Found in translation: More accurate, fluent sentences in Google Translate", Published Nov 15, 2016.

[12] Information on NMT : https://github.com/tensorflow/nmt

[13] How to make a language translator by Siraj raval : https://www.youtube.com/watch?v=nRBnh4qbPHI

[14] Google's zero shot translation : http://daily.unitedlanguagegroup.com/stories/editorials/google-translate-adds-nmt-to-8-language-pairs-unveils-zero-shot-translation