IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

Neural Machine Translation Using Attention Mechanism

Vishwas Dhabhai, Vatsal Joshi, Akash Jadhav
Department Of Information Technology
SVKM's NMIMS MPSTME Shirpur, India

1st Author-Vishwas Dhabhai
Designation-Student, Dept. of Information technology
Name of organization of 1st Author-SVKM'S NMIMS MPSTME Shirpur, India

Abstract: The development of Neural Machine Translation (NMT) has revolutionized the field of machine translation. This review paper provides a comprehensive overview of the recent advances in NMT, including its history, architecture, evaluation methods, and applications. The paper starts with a background on machine translation and a brief history of NMT. It then discusses the various NMT architectures, such as Encoder-Decoder, Transformer, and Hybrid models, and compares their strengths and weaknesses. The evaluation methods for NMT, including Bleu score, meteor score, and human evaluation, are also covered in detail. The paper concludes with a discussion of the various applications of NMT, including its use in multilingual communication, cross-lingual information retrieval, and machine-aided language learning. The paper provides insights into the current state-of-the-art NMT systems and identifies future research directions in this field.

INTRODUCTION

Neural machine translation (NMT) is an approach to machine translation that uses an artificial neural network to predict the likelihood of a sequence of words, typically modelling entire sentences in a single integrated model. The proposed system uses technologies such as Neural Networks and Attention Mechanism.

The re-discovery of neural networks has provided a significant boost to a variety of subfields within the realm of natural language processing (NLP). But for a considerable amount of time, the incorporation of neural networks into machine translation (MT) systems was just superficial at best.

Prior attempts used feedforward neural language models to re-rank translation lattices for the target language (Schwenk et al., 2006). Further on this idea, the first neural models to take into account the source language did things like scoring phrase pairings directly with a feedforward net or adding a source context window to the neural language model (Zamora-Martinez et al., 2010). The same team of scientists employed all of these techniques. Kalchbrenner, Blunsom and Cho et al. (2015) were the first to propose using recurrent networks in translation modelling. In each of these systems, neural networks were implemented as components of a more conventional statistical machine translation system. As a result, they continued to use the log-linear model combination but modified the standard design by exchanging some of its components.

3. BACKGROUND

Using an artificial neural network to estimate the likelihood of a word sequence and often modelling full sentences in a single integrated model, neural machine translation (NMT) is a method of machine translation. The online application's user can translate spoken or written language by using the algorithms in conjunction with the data in the language database.

Neural Machine Translation (NMT) is a cutting-edge machine translation method that uses neural network techniques to estimate the possibility of a set of words occurring in a given order.

This approach is sometimes referred to as Neural Machine Translation, NMT, Deep Neural Machine Translation, Deep NMT, or DNMT. This might be a single word, a complete sentence, or, in light of recent advancements, an entire piece of written material.

NMT employs deep neural networks and artificial intelligence to train neural models to address the issue of language translation and localization in a fundamentally different way. With a significant shift from SMT to NMT in just three years, NMT has quickly overtaken other machine translation approaches as the dominant one.

When compared to statistical machine translation methods, neural machine translation often delivers translations of a significantly higher quality, with greater fluency and appropriateness. Only a small portion of the memory utilised by conventional Statistical Machine Translation (SMT) models is used by neural machine translation. Since the neural translation model is trained end-to-end to maximise translation performance, this NMT approach differs from traditional translation SMT systems.

Neural machine translation aims to develop and train a single, sizable neural network that can read a sentence and produce an accurate translation, as opposed to the conventional phrase-based translation system, which is composed of numerous small sub-components that are tweaked separately.

4. PAPERS:

4.1 A Survey of Multilingual NMT. [1]

This paper give an overview of multilingual neural machine translation (MNMT), a field that has gotten a lot of attention in the past few years. Because of the transfer of translation knowledge, MNMT has helped improve the quality of translation (transfer learning). MNMT has more potential and is more interesting than statistical machine translation because it uses end-to-end modelling and distributed representations, which open up new ways to study machine translation. Using multilingual parallel corpora to improve translation quality has been suggested in many different ways. But because there isn't a complete survey, it's hard to know which approaches are promising and should be looked into further.

In this article, the authors took a close look at the literature on MNMT that is already out there. First, they putted different approaches into groups based on their main use case. Then, they putted them into more groups based on resource scenarios, modelling principles, core issues, and challenges.

After that they compared the pros and cons of different techniques by analysing them with each other. They also discussed about where MNMT is going in the future. We realized that this article is for both people who are new to NMT and those who know a lot about it. And we do certainly hope that this article will help researchers and engineers like us who are interested in MNMT get started and give them new ideas.

There are many different applications of MNMT that have been developed depending on the resources and use-cases that have been made accessible. The following is a list of the most important situations in which MNMT has been investigated in the published research: (See figure-1)

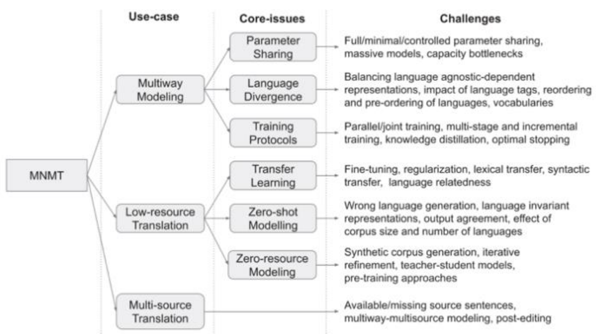


Figure 1 MNMT research is put into groups based on use cases, core issues, and challenges. Use-cases are where the focus is.

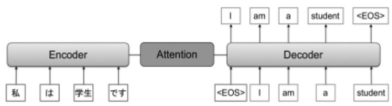


Figure 2 A standard NMT model based on the encode-attend-decode modeling approach.

Figure 3 provides a high-level summary of multi-way NMT with respect to the degree of sharing, as well as the benefits and drawbacks of the various sharing methods. Each variant of MNMT models has its own unique set of challenges during training, such as batching, language clustering, and knowledge distillation. Finding the sweet spot between language-specific and language-independent representations is also crucial.

Then the paper shed some light on multiway-NMT. considering transfer learning between languages can improve translation quality for many directions and permit translations between language pairs without data. Multiway NMT systems can also help statistical and linguistically analyse language relationships.

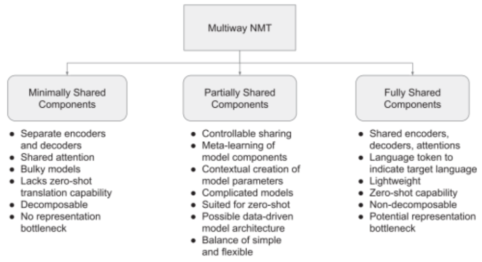


Figure 3 An overview of multi-way NMT.

Figure-4 MNMT for low-resource language pairs: an overview of potential solutions.

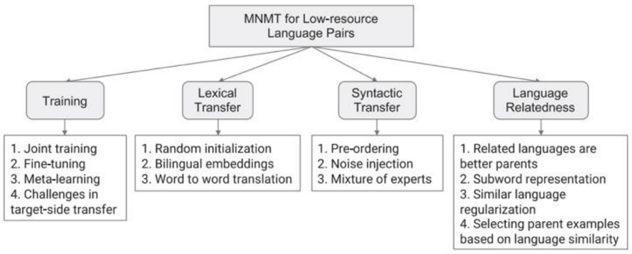


Figure 4

Table-1 An examination of potential new research paths and concerns in the field of NMT.

Central Goal	Possible Directions/Issues/Approaches
Language Representation Learning	1. Balancing the sharing of representations between languages. 2. Handling language divergence. 3. Addressing code-switching and dialects. 4. Identifying language families computationally.
Leveraging Pre-trained Models	1. Pre-trained BERT/GPT/Transformer XL encoders and decoders. 2. Incorporating web-level knowledge into translation process. 3. Designing pre-training objectives for multilingualism. 4. Dealing with large model sizes. 5. Universal parent (pre-trained) models.
One Model for All Languages	1. A single model for all languages, domains, dialects, and code-switching. 2. Possible improvement from multi-modal knowledge. 3. A model to explain multilingualism. 4. Handling representation bottleneck.

They have methodically collated the primary design methods, along with their variants, as well as the primary MNMT difficulties, along with the recommended remedies, as well as the strengths and limitations of each approach. In addition to this, they have provided a historical context for MNMT by comparing it to previous work on multilingual RBMT and SMT systems.

4.2 Neural Machine Translation System: Bridging the Gap between Human and Machine Translation. [2]

Despite the fact that MT-predicted translations are not identical to human translation, they are understandable and the translation process requires no human involvement. Proof of the translation method's efficacy is found in its ability to produce a target construct that is both semantically similar and grammatically correct. An intellectual translation method avoids literal word-for-word translation in favour of a deeper exploration of the languages' underlying concepts and meanings.

Traditional machine translation (MT) methods are rule-based, corpus-based, and hybrid. Transfer-based and interlingua-based processes use a set of translation rules to study the syntax, semantics, and morphology of the two languages to produce a target representation. Interlingua-based and transfer-based techniques are also included.

Rule-based techniques use language models to make sure that translation rules are easy to understand and to make sure that the translation is correct both grammatically and syntactically. But techniques based on interlingua aren't very good because they try to translate using a representation that doesn't depend on the language.

In this article, test sentences are supplied into the system in batches, and a trained model is used to make predictions about how those test sentences will be translated. In order to locate the best translation or the list of translations that are the best, the process of translation makes use of beam search, which is an improved and heuristically-based form of best first search.

The efficiency of the search mechanism may be observed in its capacity to support a trade-off between translation time and search accuracy. This is accomplished by setting the beam size to a reasonably modest value, which in turn ensures that the trade-off is successful. In addition to that, the translator will utilise the symbol when it is unsure about the word that it is translating.

Key-words : LSTM, Context Vector, Attention Layer.

All of the scores were visualized, compared and analysed through a bar chart. Refer figure-5, figure-6 and figure-7.

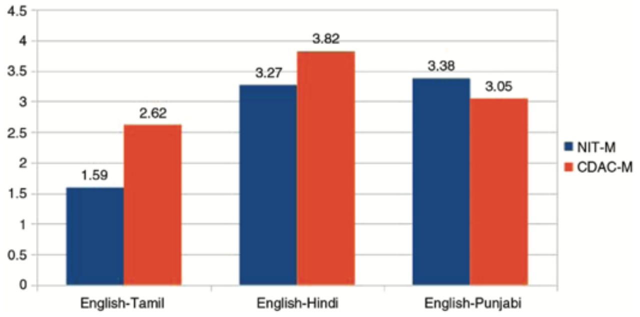


Figure 5 Comparison of scores of NIT-M & CDAC-M systems.

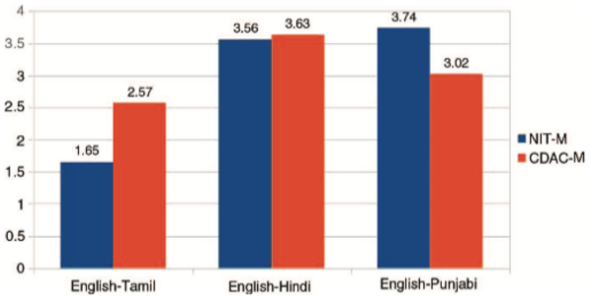


Figure 6 Comparison of fluency score of NIT-M & CDAC-M systems.

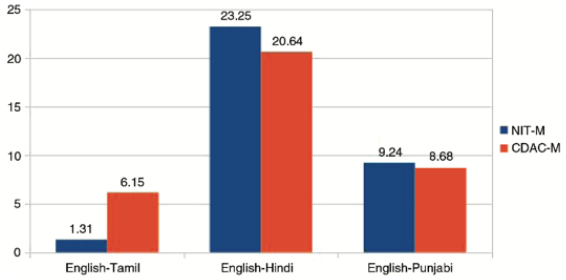


Figure 7 Comparison of BLEU score of NIT-M & CDAC-M systems.

Furthermore, they have investigated a variety of experimental designs in order to evaluate the performance of a system's translation capabilities in relation to the number of epochs, the size of the training data, and the average length of sentences in the test dataset. When comparing Gold Data and expected translations, they made use of a multi-BLEU, evaluator with a precision of 1g.

It' was surprising that the NMT system's BLEU score improves with test phrase length. NMT's context-analyzing capacity may explain it. Attention mechanism and context vectors boost translation performance with longer test sentences.

In conclusion, NMT, a new MT method, predicts translations using a well-trained end-to-end neural network. We investigated NMT for Indian languages because of its fluent translation, context-analyzing, and performance advantages over SMT. We trained and tested English-Tamil, English-Hindi, and English-Punjabi NMT systems. MTIL organisers offered predicted translations for human and BLEU evaluations on sufficiency, fluency, and overall rating. Different experimental setups have been created to examine the English-Hindi NMT system's translation performance as epochs, training data, and test sentence length change. Analysing projected translations shows that NMT systems create fluent translations and improve with training data and test phrase length. A translation performance vs epoch graphic can help verify system training convergence.

4.3 Machine translation of English speech: Comparison of multiple algorithms [23]

This paper gives a concise overview of the neural network approach that is used for speech recognition. Encoding was done using a technique called long short-term memory (LSTM), rather than the more conventional recurrent neural network (RNN), while decoding was done using RNN. LSTM was utilised for the encoder, and RNN was used for the decoder. The machine translation algorithm was then subjected to simulation experiments, after which it was compared to two other machine translation algorithms.

The findings showed that the back-propagation (BP) neural network correctly identified the test samples in a shorter amount of time and with a lower word mistake rate than artificial recognition. In addition, the LSTM-RNN strategy resulted in less incorrectly spelled words when contrasted with the BP-RNN and RNN-RNN approaches. In the real-world speech translation evaluation, the LSTM-RNN algorithm demonstrated the least variation in translation score and word error rate across all speech lengths. Furthermore, when the length of speech was held constant, the LSTM-RNN algorithm demonstrated the greatest translation score and the lowest word error rate across all speech lengths. Even though the duration of the speech was held constant, the speech still managed to get the best possible score for translation.

Machine translation uses computers to translate English documents accurately and efficiently, but computers, which lack vision and hearing, need input text before transforming characters [7]. The report found that human input is ineffective for querying and translating English words and phrases in practical applications. However, voice input is convenient and allows computer-aided simultaneous interpretation. Before voice-translating English, the user's voice must be recognised and the audio turned into text characters..

A BP neural network was utilised in this investigation for the purpose of voice character recognition. The BP neural network was chosen because it is the most fundamental type of neural network and has a high degree of generalizability in comparison to other types of neural networks.

And figure-9 explains the typical process of speech recognition.

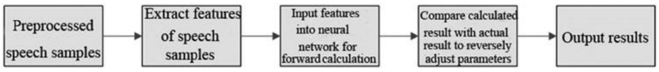


Figure-8

This study first extracts features from English voice samples using Mel-scale Frequency Cepstral Coefficients. The retrieved features train the BP neural network. The BP neural network trains with extracted feature samples. The hidden layer uses the activation function to perform multilayer forward calculation of the extracted features, and the results are compared to the training data.

The hidden layer hyperparameters are reversed according to the results gap. Then, the layer-by-layer forward computation is repeated, compared to the actual results, and the parameters are changed in reverse. Repeat the preceding procedures until the difference between calculated and real results reaches the threshold value. After feature extraction, the trained neural network model receives speech samples for testing and calculates recognition results.

4.4 Machine translation algorithm based on deep learning.

The conventional method of machine translation for recognising voice in text is to translate the speech text word by word utilising a word bank of Chinese-English mutual translation.

This method is used for translating recognised speech texts. Long phrases in English are difficult to translate effectively using this way of machine translation, despite the fact that it is simple and quick to use, and the traditional method of translation can guarantee the overall meaning of the text that is being translated.

Word-for-word translation is problematic for a number of reasons, the first of which is that because Chinese and English syntax are so dissimilar, it frequently results in grammatical confusion or even interpretations that are the exact opposite of what the words mean.

The second reason is that the translation is made less apparent because several English auxiliary words that don't have precise meanings are also translated.

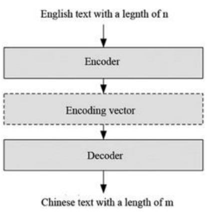


Figure 9 Deep learning based algorithm.

The above problems with machine translation have been fixed by the development of intelligent algorithms based on deep learning. Figure 10 shows how intelligent algorithm-based machine translation works at its core. The structure as a whole is made up of an encoder and a decoder. Deep learning algorithms, such as BP neural networks, convolutional neural networks, and so on, are used in the encoder and decoder. When the English text is being translated, the encoder first turns the English text into an encoding vector, and then the decoder turns the encoding vector into Chinese.

4.5 Machine translation algorithm with a LSTM encoder and a RNN decoder.

In this study, the encoder turns the text into a coded vector so that the text can be translated by a machine when it is decoded. However, the sequence of words in an utterance will change the meaning of the words, and the traditional BP network does not take this into account when encoding the text. When training and using the RNN algorithm, the current results are affected by the previous results. This is similar to how language works, where word sequences can change the meaning of words. The RNN algorithm is used for encoding and decoding machine translation because of this.

To begin, a training set was utilised to instruct the BP neural network that would later be used for speech recognition. The rectified linear unit activation function was buried within each of the five hidden layers that were present. In each of the buried layers, there were a total of 1,024 nodes. 500 iterations was the most that could be done in total. The results of the testing were compared with the results of artificial recognition, and the training set was used for testing purposes.

In the comparative experiment, all three machine translation algorithms shared a fundamental structure of "encoder + decoder," with RNN serving as the decoder in one of the algorithms. The BP neural network, the RNN, and the LSTM were each considered to be intelligent algorithms. The intelligent algorithm served as the distinguishing factor between the three different algorithms.

The primary goal of this experiment was to assess the generalisation performance of the machine translation algorithm in a real-world application situation after speech recognition and machine translation were combined. There were ten people who agreed to take part in the experiment. Then they each read aloud an English paragraph with a different word count: 100, 200, 300, 400, and 500 words. Volunteers' English voice was captured using a radio device and put into an algorithm trained to perform machine translation.

Table 2 shows that the word mistake rate and recognition time for English speech recognition using artificial recognition was 7.53 percent and 34 minutes, respectively, but these values were 1.54 percent and 8 minutes, respectively, when using a BP neural network. When comparing BP neural network with automatic recognition, the latter took longer but made fewer word errors. The above result was achieved because computers were more computationally efficient than humans, and because humans struggled to maintain their attention for an extended period of time when presented with a large number of speech sounds in the test set, leading to a higher word error rate and longer recognition time. However, the BP neural network avoided the drawback of inattentiveness by using computers to recognise speech sounds.

Table 2 Machine translation algorithm based on deep learning.

	Word error rate (%)	Recognition time (s)
Artificial recognition	7.53	34
BP neural network	1.54	8

The translation word error rate for the machine translation algorithm using the BP neural network for the encoder and the RNN for the decoder was 25.4% (see Table-3 for details), while the rate for the machine translation algorithm using the RNN for both the encoder and the decoder was 18.7%, and the rate for the machine translation algorithm using the LSTM neural network for the encoder and the RNN for the decoder was 3.2%. The error rate of translated words was highest for the BP-RNN approach, second highest for the RNN-RNN technique, and lowest for the LSTM-RNN approach. These two algorithms found application in automatic translation software.

	BP–RNN	RNN–RNN	LSTM–RNN
Word error rate (%)	25.4	18.7	3.2

Table 3 Calculation of word-error rate.

The two lab test results above were achieved after training using a predetermined set of samples. Three machine translation techniques used neural networks. When learning with a limited number of training samples, the neural network appears to improve, but when it tests on out-of-sample data, it falls into an overfitting condition and fails to generalise. In This study recruited ten volunteers to read aloud English scripts with varied word counts and translated their voice using the three trained machine translation algorithms to examine their generalisation performance.

This study provides a concise introduction to the neural network algorithm for voice recognition and uses LSTM rather than the more conventional RNN as the encoding algorithm for the encoder. On the other hand, typical RNN is used as the decoding algorithm for the decoder. The machine translation algorithm was subjected to simulation trials, and the results of those studies were compared with those of two other machine translation algorithms. The results are summarised in the following. When compared to the artificial recognition approach, the BP neural network had a lower word mistake rate and required less time to complete the recognition process.

LSTM-RNN has the lowest word error rate for English voice recognition, followed by RNN-RNN and BP-RNN. Three algorithms were RNN-based. Despite the speech duration being the same, the LSTM-RNN algorithm had the lowest word error rate and greatest translation score. Based on real-world data, three translation algorithms performed similarly in speech translation applications. The word mistake rate increased with speech length, lowering the translation scores of the three systems.

4.6 Information Retrieval System and Machine Translation [2]

Some of the most important subfields in information retrieval, including Cross-Lingual Information Retrieval (CLIR), Multi-Lingual Information Retrieval (MLIR), and Approaches and Techniques for Machine Translation, are discussed in this work. The ability to store and retrieve information locally is extremely important for growing nations in today's ever-expanding nation. The Center for Linguistic and Intercultural Research (CLIR) processes queries and requests for document retrieval in languages other than English. MLIR allows queries to be posed in one or more languages, and it retrieves documents in one or more languages. Both the CLIR and MLIR systems cannot work properly without the assistance of machine translation.

Within the field of computational linguistics, one aspect of language processing that is included is machine translation. The document or the query can be translated using the machine translation approach, which refers to a technology that translates text automatically.

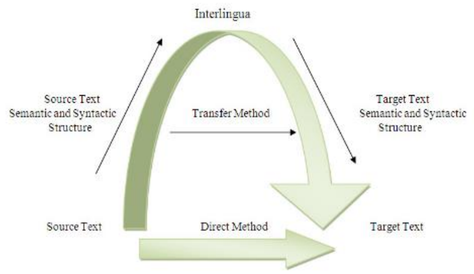


Figure 10

This can be the foundation for searching not only between two languages but also in multiple languages. Cross-lingual and Multi-lingual Information Retrieval (IR) provides new paradigms for searching documents through various varieties of languages all over the world, and this can be the basis for searching in multiple languages as well. For several years, the domain of artificial intelligence and information retrieval systems known as machine translation has been a hotbed of research activity. Because of their great level of complexity, natural languages provide a challenging problem for machine translation (MT). Because of the nature of how languages have developed over time, it is challenging to say whether or not a single strategy would be adequate to manage the translation process.

4.7 Machine translation using natural language processing. [11]

Machine translation, sometimes known as MT, is a subfield of computational linguistics that investigates the use of software to translate written text or spoken language from one language to another. Machine translation is nothing more than the process of words in one language being substituted for words in another language, however this does not necessarily guarantee an accurate translation. Statistical and neural methodology is a more advanced method that is also a burgeoning topic that is used to handle the problem of recognising multiple phrases. It is one of the ways that this issue is being addressed. The process of converting text from one language to another in this instance involves no involvement from human translators; rather, the conversion is carried out entirely by a machine.

Rule-based, statistical, and neural translation systems are the three primary categories of machine translation software. The rule-based method is a standard approach that combines language and grammar with the support of dictionaries. Rule-based methods can be broken down into three categories: The construction of a complete machine translation pipeline is the primary emphasis of this effort. We went over several different architectures that are already in use, and in the end, we came up with the idea of using a hybrid model to create a more effective machine translation system that goes from English to French.

Layer (type)	Output Shape	Param #
embedding_49 (Embedding)	(None, 15, 128)	25600
bidirectional_65 (Bidirectio	(None, 256)	197376
repeat_vector_45 (RepeatVect	(None, 21, 256)	0
bidirectional_66 (Bidirectio	(None, 21, 256)	295680
time_distributed_136 (TimeDi	(None, 21, 512)	131584
dropout_64 (Dropout)	(None, 21, 512)	0
time_distributed_137 (TimeDi	(None, 21, 345)	176985

Total params: 827,225
Trainable params: 827,225
Non-trainable params: 0

Therefore, in order to transform English text into French, we developed a deep neural network that operates as a component of a pipeline for end-to-end machine translation. The output of this network is the French translation of the English text.
In terms of the validation loss, the performance of the suggested network was significantly superior than that of the other designs that were mentioned earlier.

It was accurate 96.71% of the time. After determining a test harness for three-fold cross validation, compiling the model, and performing an evaluation on it, a random seed was fixed for reproducibility.

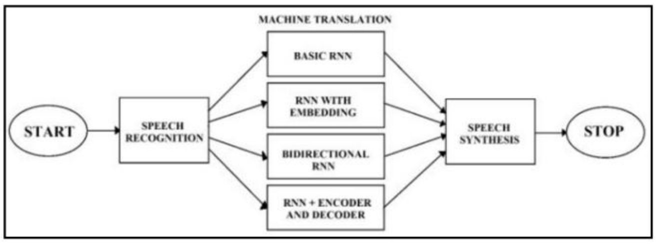
4.8 - Neural Machine Translation using Recurrent Neural Network [7]

In this day and age of increased globalisation, it is highly likely that we may come across individuals or communities that do not communicate with us using the same language that we use. In order to address the issues that this raises, we are currently working on improving our machine translation capabilities. Developers at a number of well-known companies, such as Google LLC, have been hard at work on the creation of algorithms to support machine translations.

These developers are utilising machine learning algorithms, such as Artificial Neural Networks (ANN), in order to make machine translation more accessible.

In this regard, several neural machine translations have been produced, but on the other hand, the development of recurrent neural networks (RNN) has not progressed very much in this particular subject.

In the course of our work, they have attempted to introduce RNN into the field of machine translations in order to draw attention to the advantages that RNN possesses in comparison to ANN. The findings demonstrate that RNN is capable of carrying out machine translations with the required degree of accuracy.



The built system integrates multiple recurrent neural network models within 10 epochs, achieving maximum accuracy. A better model is obtained by collecting multiple models at simultaneously, which improves the overall system's precision. The observational evidence for this was gathered. It has also been observed that, among the different models, the Recurrent Neural Network with embedding yields the maximum accuracy with increasing iterations. For example, two of the worst-performing individual models are the simple Recurrent Neural Network and the Recurrent Neural Network.

4.9 – Effective Approaches to Attention-based Neural Machine Translation [5]

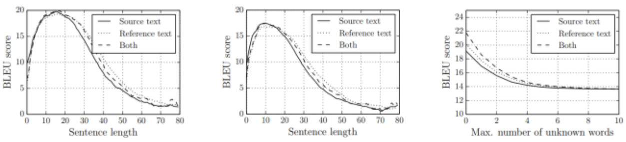
An attentional mechanism has recently been utilized to improve neural machine translation (NMT) by focusing selectively on parts of the source sentence while it is being translated. This allows NMT to better understand what is being translated. On the other hand, not much research has been done to investigate potentially helpful architectures for attention-based NMT. This study investigates two classes of attentional mechanisms that are both straightforward and efficient. The first is a global strategy that pays attention to each and every source word, while the second is a local strategy that focuses its attention on only a subset of source words at any given time.

We demonstrate the efficacy of both approaches on the WMT translation tasks between English and German in both directions by focusing on both directions simultaneously. We get a large gain of 5.0 BLEU points with local attention compared to non-attentional systems that already utilize well-known approaches such as dropout. This represents a major improvement. Our ensemble model, which makes use of a variety of attention architectures, achieved a new state-of-the-art result in the WMT'15 English to German translation task, scoring 25.9 BLEU points. This represents an improvement of 1.0 BLEU points over the previous best system, which was supported by NMT and an n-gram re-ranker. In this paper, we put out a proposition By analysing and contrasting a wide variety of alignment functions, we were able to shed light on which functions work best with which attentional models. According to the findings of our research, attention-based natural language processing models are superior to non-attentional ones in a variety of contexts, such as when translating names and managing lengthy phrases.

4.10 – On the Properties of Neural Machine Translation: Encoder–Decoder Approaches [4]

A relatively recent development in the field of statistical machine translation, neural machine translation is a technique that relies only on neural networks. In neural machine translation models, an encoder and a decoder are frequently present as separate components. The encoder takes a sentence with a variable length as input and generates a representation of a fixed length from it. The decoder then takes this representation and generates a translation that is correct.

In this paper, we focus on analysing the properties of neural machine translation utilising two models: RNN Encoder–Decoder and a newly suggested gated recursive convolutional neural network. Both of these models are gated recursive convolutional neural networks. We demonstrate that the neural machine translation performs quite well on short phrases that do not contain any unknown words; nevertheless, its performance deteriorates significantly as the length of the sentence and the number of unknown words increases. In addition, we have discovered that the gated recursive convolutional network that we presented is capable of autonomously learning the grammatical structure of a sentence.



(a) RNNenc (b) grConv (c) RNNenc

In this paper, we conducted research into the properties of a family of machine translation systems that was only recently developed and is solely based on neural networks. The evaluation of an encoder–decoder strategy, which was recently proposed in (Kalchbrenner and Blunsom, 2013; Cho et al., 2014; Sutskever et al., 2014), was the primary focus of our work when it came to the problem of sentence-to-sentence translation. We specifically chose two encoder–decoder models that differ in the choice of the encoder from the many possible encoder–decoder models. The first model is an RNN with gated hidden units, and the second model is a newly proposed gated recursive convolutional neural network. Both of these models can be found in the previous sentence.

Following the training of these two models on matched sets of English and French phrases, we evaluated their performance with BLEU scores, taking into consideration the lengths of the sentences as well as the presence of unfamiliar or uncommon words within the sentences. According to the findings of our investigation, the performance of the neural machine translation is greatly hindered by the length of the sentences being translated. On the other hand, we came to the conclusion that both models are capable of producing accurate translations in a highly satisfactory manner.

1.4 SYSTEM OVERVIEW

Implementing a Neural Machine Translation (NMT) system typically requires the following:

- Data: A large parallel corpus of source and target languages is required to train the NMT model.
- Pre-processing: The raw data must be pre-processed and cleaned to remove any irrelevant information, noise, and formatting inconsistencies.
- Tokenization: The text must be tokenized into words or sub-words units to provide inputs to the NMT model.
- Vocabulary: A vocabulary of the source and target languages must be created, mapping words and sub-words to unique numerical representations called word embeddings.
- Model architecture: The architecture of the NMT model must be selected based on the available resources, the size of the training data, and the desired level of accuracy. Common architectures include Recurrent Neural Networks (RNNs) and Transformer networks.

- Training: The NMT model must be trained using the parallel corpus to learn the mapping between the source and target languages. This is typically done using stochastic gradient descent or a variant thereof.
- Evaluation: The trained model must be evaluated on a held-out test set to measure its accuracy, speed, and ability to generate coherent translations.
- Deployment: The final trained model can be deployed in a translation API or integrated into a machine translation system.

Note that this is a high-level overview, and the specific requirements for implementing NMT can vary depending on the specific use case, available resources, and desired accuracy level.

CONCLUSION

In conclusion, this review paper has provided a comprehensive overview of the field of Neural Machine Translation. The various NMT architectures, such as Encoder-Decoder, Transformer, and Hybrid models, have been compared and evaluated, and their strengths and weaknesses have been discussed. The various evaluation methods for NMT, including Bleu score, meteor score, and human evaluation, have been discussed in detail. The various applications of NMT, including its use in multilingual communication, cross-lingual information retrieval, and machine-aided language learning, have also been covered. The current state-of-the-art NMT systems have been described and future research directions in this field have been identified.

In general, NMT has made great progress in recent years and has shown great promise in revolutionizing the field of machine translation. Despite the progress made, there is still room for improvement and ongoing research in NMT will be necessary to further develop this technology and address its limitations.

BIBLIOGRAPHY

[1] Dabre R, Chu C, Kunchukuttan A. A Survey of Multilingual Neural Machine Translation. (2020)

[2] 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 (2016).

[3] *Sheila Castilho, Joss Moorkens, Federico Gaspari,a lacer Calixto, John Tinsley, Andy Way Is Neural Machine Translation the New State of the Art?,Prague Bulletin of Mathematical Linguistics,* June 2017.

[4] Kyunghyun Cho, Bart van Merrienboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, Yoshua Bengio. On the Properties of Neural Machine Translation: Encoder–Decoder Approaches," 3 June 2014.

[5] Minh-Thang Luong, Hieu Pham, Christopher D. Manning Effective Approaches to Attention-based Neural Machine Translation(2015)

[6] A. P. &. P. Pakray, *Journal of Intelligent Systems,* 1 7 2019.

[7] Debajit Datta, Preetha Evangeline David, Dhruv Mittal, Anukriti Jain, Debajit Datta, "Neural Machine Translation using Recurrent Neural Network," *International Journal of Engineering and Advanced Technology,* vol. 9, no. 4, p. 1395, April 2020.

[8] Muskaan Singh, Ravinder Kumar & Inderveer Chana. "Corpus based Machine Translation System with Deep Neural Network for Sanskrit to Hindi Translation," *International Conference on Computational Intelligence and Data Science (ICCIDS 2019),* 2019.

[9] Mangala Madanka, M.B. Chandak and Nekita Chavhan Information Retrieval System and Machine Translation : A Review.(2015)

[10] B. Premjith*, M. Anand Kumar and K.P. Soman Neural Machine Translation System for English to Indian Language Translation Using MTIL Parallel Corpus. Journal of Intelligent Systems, vol. 28, p. 387, 2019.

[11] Middi Venkata Sai Rishita, Middi Appala Raju and Tanvir Ahmed Harris "Machine translation using natural language processing.," *MATEC Web of Conferences,* 2019.

[12] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones , Aidan N. Gomez, tukasz Kaiser and Illia Polosukhin Ashish Vaswani, "Attention is all you need," *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems,* p. 6000, December 2017.

[13] Phillip Koehn. Rebecca, R. Knowles, "Six Challenges for Neural Machine Translation," *Proceedings of the First Workshop on Neural Machine Translation*, p. 28, 2017.

[14] Kyunghyun Cho, Bart van Merrienboer, etal. "Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation," *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, p. 1724.

[15] Shuoheng Yang, Yuxin Wang and Xiaowen Chu A Survey of Deep Learning Techniques for Neural Machine Translation, (2020).

[16] Sethunya R Joseph, Hlomani Hlomani, Keletso Letsholo, Freeson Kaniwa and Kutlwano Sedimo.. "Natural Language Processing: A Review," *International Journal of Research in Engineering and Applied Sciences,* vol. 6, no. 3, March 2016.

[17] Diksha Khurana, Aditya Koli, Kiran Khatter and Sukhdev Singh. Natural Language Processing: State of The Art, Current Trends and Challenges.

[18] Y Wen Zhang, Yang Feng, Fandong Meng, Di You, Qun Liu Wen Zhang, "Bridging the Gap between Training and Inference for Neural Machine Translation," vol. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, p. 4334–4343, July 2019.

[19] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, Peter J. Liu. "Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer," *Journal of Machine Learning Research 21,* 2020.

[20] "On Application of Natural Language Processing in Machine Translation," *2018 3rd International Conference on Mechanical, Control and Computer Engineering,* 2018.

[21] A. Navon, "Parameter Sharing in Deep Learning," [Online].

Available: https://avivnavon.github.io/blog/parameter-sharing-in-deep-learning/.

[22] Amarnath Pathak, Partha Pakray Neural Machine Translation for Indian Languages, DE Gruyter, (2018)

[23] Qin, Yijun Wu and Yonghong Machine translation of English speech: Comparison of multiple algorithms, (2022) *Journal of Intelligent Systems,* vol. 31, p. 159, 2022.