

Recurrent Neural Network Techniques: Emphasis on Use in Neural Machine Translation

Dima Suleiman

King Abdullah II School of Information Technology, The University of Jordan, Amman, Jordan

E-mail: dima.suleiman@ju.edu.jo

Wael Etaiwi

King Talal School of Business Technology, Princess Sumaya University for Technology, Amman, Jordan

E-mail: w.etaiwi@psut.edu.jo

Arafat Awajan

King Hussein School of Computing Sciences, Princess Sumaya University for Technology, Amman, Jordan

College of Information Technology, Mutah University, AlKarak, Jordan

E-mail: awajan@psut.edu.jo, awajan@mutah.edu.jo

Keywords: machine translation, graph-based, phrase-based, BLEU

Received: September 6, 2021

Natural Language Processing (NLP) is the processing and the representation of human language in a way that accommodate its use in modern computer technology. Several techniques including deep learning, graph-based, rule-based and word embedding can be used in variety of NLP application such as text summarization, question and answering and sentiment analysis. In this paper, machine translation techniques based on using recurrent neural networks are analyzed and discussed. The techniques are divided into three categories including recurrent neural network, recurrent neural network with phrase-based models and recurrent neural techniques with graph-based models. Several experiments are performed in several datasets to make translation between different languages. In addition, in most of techniques, BLEU is used in evaluating the performance of different translation models.

Povzetek: Opisan je pristop s povratnimi globokimi nevronskimi mrežami za jezikovno prevajanje.

1 Introduction

Natural Language Processing (NLP) is a subset of artificial intelligence that can automatically represents and processes human language using computational techniques [1-4]. There are several NLP tasks and applications such as machine translation, information extraction, question answering, and text summarization [5-7].

Machine translation in one of natural language processing computer applications that receives a sentence in certain natural language called source input and translates it into a target sentence of another natural language where both the source and the target sentences must have the same meaning [8]. Machine translation is crucial and essential in natural language processing for many reasons. The first reason related to the benefits of people communication over the world who speaks different languages. The second reason is the lack of machine translation that perfectly translates and satisfies the user requirements. Another important reason is the cost, speed and throughput of using machine translation tools which will be less than the cost of human translation. Finally, machine translation is used in several fields of natural language processing, thus it must be efficient [8].

There are three categories of machine translation including semantic web machine translation, statistical machine translation and neural machine translation [9]. In this paper, the focus is on neural machine translation where neural networks and deep learning techniques are used in translation. Neural network is one of machine learning techniques that enable learning using several layers. The basic structure of neural networks consists of three layers which are input, hidden and output layers. Each layer consists of one or more processing units called neurons or hidden states. The lines connect the neurons consist of weights that are initialized randomly and updated during the training of the network. In the case of the machine translation, the inputs for the neural network are the words of the text.

Deep learning is a complex neural network that consists of many hidden layers and several hidden states in each layer. Deep learning can be used for extracting features with different level of abstraction, either high level with fewer details or low level with more details [10]. There are several types of deep learning such as recurrent neural network (RNN), convolutional neural network (CNN) and auto-encoder (AE). In this research, the main focus is on recurrent neural network.

Several review papers compared the deep learning techniques used in neural machine learning [11-12]. In [11], resources and tools used in neural machine translation were summarized. In addition, comparisons were made in terms of decoding, modelling, interpretation, augmentation of data and evaluation. On the other hand, Stahlberg [12] traced several neural language models in addition to trace the using of words and sentences embedding in representation. Moreover, the neural machine learning architectures including convolutional or recurrent neural networks were reviewed in addition to reviewing the segmentation, decoding and training techniques used. This paper provides a comprehensive analysis of recent neural machine learning techniques based on recurrent neural networks. The main focus of this paper is the combination of RNN and other modeling. The techniques are divided into three categories including: recurrent neural network, recurrent neural network with phrase-based models, and recurrent neural techniques with graph-based models. Comparisons are made in terms of techniques, modeling, and using of attention and copy mechanisms, datasets and evaluation. The main difference between this paper and similar papers is the focus of using recurrent neural network in neural machine translation. The rest of this paper is organized as follows: Section 2 explains material and methods. Section 3 investigates the results and the discussions. Finally, section 4 presents the conclusion.

2 Neural machine learning techniques

In this section, we divided the neural machine learning techniques into three categories: recurrent neural network, recurrent neural network with phrase-based models and recurrent neural techniques with graph-based models.

Most of the models based on sequence-to-sequence encoder-decoder model. Sequence-to-sequence model can be seen in Figure 1. In sequence-to-sequence model, the input is a sequence of words and the output is sequence of words. RNN consists of sequence of hidden states where the output of each hidden states is passed as input to the next hidden state. In machine translation, the inputs at the encoder are the words of the text of the source language. On the other hand, the outputs at the decoder are the words of the target language.

2.1 Recurrent Neural Networks Techniques:

One of the models of recurrent neural networks (RNN) is called Encoder-Decoder was proposed in [14]. The proposed model consists of two RNN where one RNN is called Encoder since it encodes symbols sequence into a fixed length vector and the other RNN is called Decoder since it decodes the fixed length vector into another symbols sequence as shown in Figure 2 [14]. In order to maximize the probability of extracting the target sequence of symbols from the source sequence of symbols, both RNNs must be trained jointly. On the other hand, in order to facilitate the training process in addition to enhancing

the capacity of the memory, the authors proposed to use the hidden units.

In their research, Cho et al. [14] focused on translation from the English phrase to the French phrase by training the proposed model to learn the corresponding translation. After that, the scoring of each pair of phrases which exists in the phrase table was calculated in order to utilize the proposed model within the standard system of phrase-based statistical machine translation. On the other hand, in order to analyse the quality of the proposed model, comparisons between its phrase score and phrase score of already existing translation models were made. Moreover, the experiments showed that the phrase continuous space representation can be learned using RNN Encoder – Decoder which plays a significant role in keeping the semantic and syntactic phrase structure.

In addition, the proposed model can learn the conditional probability over the variable length sequence on another one, where the two sequences may differ in length. After training, the Encoder – Decoder RNN model can be used for two purposes: the first one is to find the target sequence given the source, while the second purpose is to find the score of given target and source sequences. Another advantage of the proposed model is that it discriminates between the sequences that have the same words but in a different order. Also, it produced word embedding matrix learned from the model which display the relationship between the words. Finally, several models were implemented and evaluated using BLEU evaluation metric. The best result achieved was 34.54.

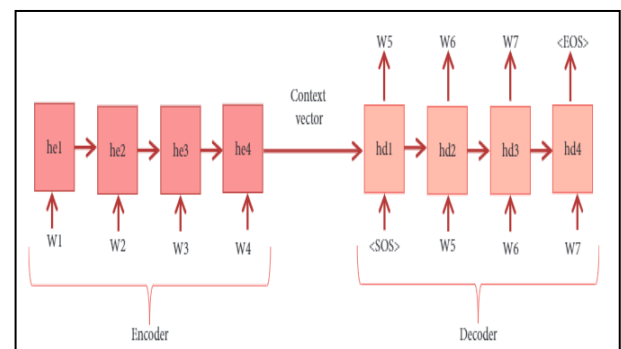


Figure 1: Sequence-to-Sequence Model [13].

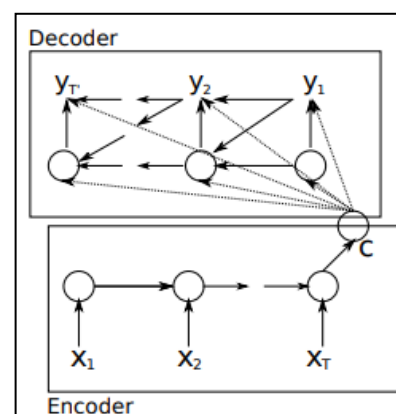


Figure 2: RNN encoder decoder model [14].

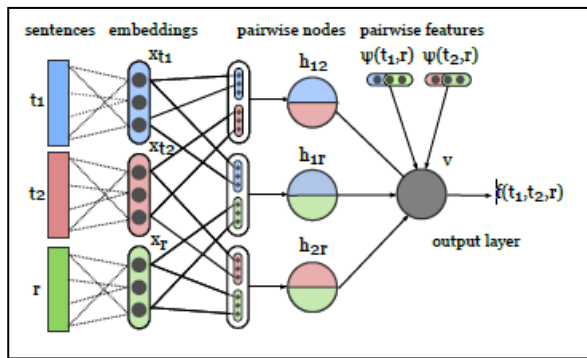


Figure 3: The architecture of the model [15].

Framework for using a neural network in evaluating the machine translation process was proposed in [15], where for a given reference translation, the best translation will be chosen from hypotheses pair using the proposed framework as shown in Figure 3 [24]. Multi-layer neural network was used to model the interaction and nonlinear relationship from two sides: the first one between the two hypotheses and the second one between the two hypotheses and the reference. Moreover, the input for the multi-layer neural network will be a vector representation which consists of a compact distribution of the two hypotheses in addition to the reference semantic, lexical and syntax information. Therefore, syntax and semantic information are crucial to get the relationship between the reference and the two hypotheses. Thus, in order to represent the relation, Glove and word2vec embedding representation for the input of neural network was used to represent syntax and semantic vectors. In addition, the experiments were made using WMT Metrics shared task datasets which are: WMT11, WMT12, WMT13, and WMT1. Furthermore, the extension of this framework was modelled using recurrent and convolutional neural networks. Accordingly, the results of using the proposed model provided efficient learning because of its flexibility and generality. They used several languages during experiments such as the translation from Hindi to English, German to English, Russian to English and others. The best value of BLEU was 44.1 which was obtained during the translating from Hindi to English.

The performance of machine translation from Japanese to English using recurrent neural network was examined in [16]. In spite of the fact that there are large freely available corpora such as Kyoto wiki corpus and TED corpus which consists of 500,000 sentence-pair and 150,000 sentence-pair respectively, however, according to a limited number of resources they created parallel hand-crafted corpora. Moreover, the evaluation of the models was conducted using BLEU which is a metric for evaluating machine translation. BLEU metric was used to measure the precision of the translation process by comparing the phrase translated by machine with a phrase translated by a human. As a conclusion, training the model on small parallel corpus give reasonable results with BLEU value is equal to 73. In addition, it is expected to perform well on a large corpus.

Recurrent Highway Networks (RHN) encoder-decoder with attentions is used by Parmer and Devi [17] in natural machine translation tasks. The authors demonstrate the effectiveness of RHN approach as well as LSTM encoder and decoder on the IWSLT English-Vietnamese dataset. The experimental results indicate that RHN performs the same with LSTM based models and, in some cases, even better. The BLEU value of their model was 24.9.

Datta et al. [18] developed a three stages model to facilitate speech translation using RNN. These three main modules are Speech Recognition, Machine Translation and Speech Synthesis. The authors used Google APIs to convert text to speech and speech to text. English to French dataset is used in the experiments. The English corpus consists of 1,823,250 English words, while the French corpus contains 1,961,295 French words. The authors concluded that using multiple models at a time for machine translation resulted in higher accuracy for the proposed framework as a whole. Accuracy was used to measure the performance of the model where its value approached from 97.37.

Liu et al. [19] proposed an approach that based on the agreement between a pair of targeted directional RNNs to translate from Japanese to English and from English to Japanese. Two efficient approximate search methods have been developed for agreement. In terms of either non-sequence level or sequence level metrics, the search methods are empirically shown to be almost optimal. Three standard sequence-to-sequence transduction tasks were used in the experiments to validate the proposed approach: machine translation and machine transliteration, grapheme-to-phoneme transformation. The experimental results show that the proposed approach achieves substantial improvements and consistent, compared to many state-of-the-art systems. The best result of BLEU obtained is 35.

2.2 Recurrent Neural Network Techniques with Phrase-Based Model:

Huan et al. [20] proposed neural machine translation model (NPMT) based on phrases where the model produced the output sequence based on using an already existing model called Sleep-Wake Networks (SWAN) that depend on segmentation. Nevertheless, a new layer before the SWAN layer was added to reorder the local sequence of input slightly in order to minimize the requirement of monotonic alignment in SWAN. However, the proposed model differed from the previous neural machine translation in that it can decode in real time the sequential order of output phrases instead of using decoding mechanisms which is attention based. The structure of NPMT consists of soft reordering of the phrases in German sentences after representing it using word embedding. In the next step, the reordered phrases were passed to Bi-directional RNN. Moreover, after that, the results passed to SWAN for monotonic alignment. Finally, the phrases were translated one by one to English to form the target sentence in English. On the other hand, given the sequence of the output, SWAN can model all

valid segmentations of the output by defining a distribution for the probability of the output and using dynamic programming. In addition, SWAN modelled the alignment between the input sequence and output segments where empty output is possible and no assumptions for input and output sequences length. As a result, the experiments were conducted in IWSLT 2014 and IWSLT 2015 tasks and showed that the output phrases are meaningful. Also, the performance was significantly improved. The overall architecture is displayed in Figure 4 (a) and an example of translation from Germany to English can be seen in Figure 4 (b). The model was evaluated using BLEU with value 25.36.

Attention mechanism achieved significant improvement in performance in machine translation using a sequence-to-sequence neural network. The reason for this improvement is related to capturing the contextual information from the source side continuously during prediction. However, this is not the same from the target side since extracting contextual information for non-sequential words dependence is not an easy process. Thus, Werlen et al. [21] proposed to use the self-attentive residual recurrent network for decoding. Therefore, the self-attentive residual was used within the attention base neural network and focused on propagating useful contextual information from the translation of the previous words to the output of the decoder. The translation included three pairs of languages which are: English to German, English to Chinese, and Spanish to English. In addition, the datasets were used were a complete set from WMT 2016 for English to German translation, a subset of the UN parallel corpus for English to Chinese translation and subset of WMT 2013 for Spanish to English translation. Several models were implanted and the best results were achieved using self-attentive residual connections model with BLEU values 29.7 for the translation from English to Germany.

A new approach of using RNNs over traditional statistical MT (SMT) for machine translation is proposed by Mahata et al. [22]. The performance of the proposed RNN is compared with the performance of the phrase table of SMT. Traditional machine translation model has been constructed using Moses toolkit in addition to enriching the language model using external data sets provided by MTIL2017 for translating from English to Hindi. Furthermore, the phrase tables are ranked using an RNN encoder-decoder module. The experimental results showed that for long sentences SMT works well and for short sentences neural machine translation works well. Their model BLEU value was 3.57.

2.3 Recurrent Neural Network Techniques with Graph-Based Model:

Encoding the semantic meaning of the sentence as a rooted directed graph is called Abstract Meaning Representation (AMR) where the nodes represent the concepts and the relations between the concepts were represented using edges. However, recovering the text from AMR graph and preserving the meaning of the original text was considered a problem. In order to overcome this problem, in their

research [23], authors proposed to use novel LSTM structure to directly encode the structure of AMR in graph to sequence model as shown in Figure 5 [23]. At the decoder, attention mechanism was used in addition to using the copy mechanism. Moreover, the dataset was used in experiments was the standard AMR corpus LDC2015E86 with 1368 instances for development, 16833 instances for training and 1371 instances for testing. As a result, the proposed model outperformed others in the literature. The experimental results showed that the proposed model outperformed the other model with BLEU value equal to 33.

In addition to machine translation of the source language, Hashimoto et al. [24] proposed attention based on neural machine translation model that learn the representation of the source sentences as a part of the encoder using task-specific latent graph parser. However, there is a similarity between the structure of the dependency of the sentence and the structure of the graph,

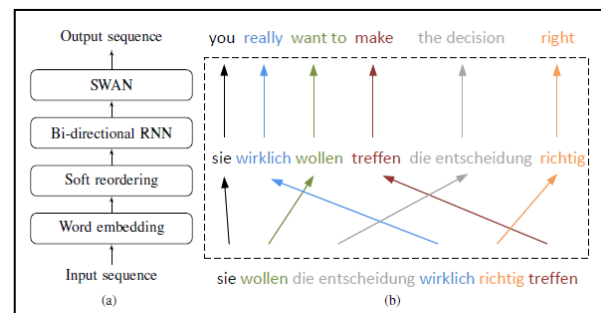


Figure 4: (a) The overall architecture (b) An example of translation from Germany to English [20].

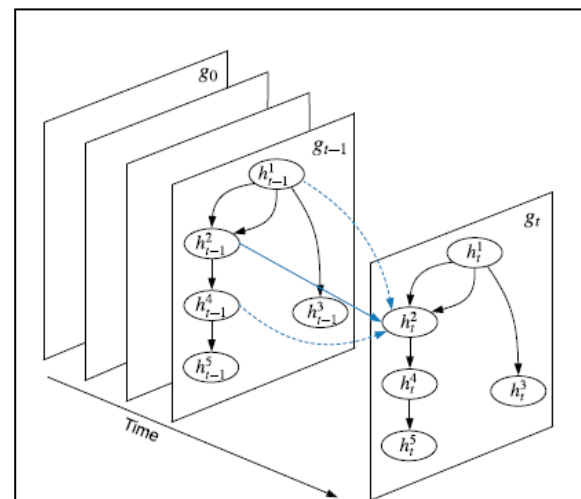


Figure 5: Graph state LSTM [23].

in addition to the possibility of having a cycle in the graph. Also, each graph edge has a real value, thus the connection is soft. There are two parts of the proposed model; the first one was the latent graph parser which is task independent and pre-trained independently with Treebanks, while the second part is the attention-based part. Moreover, the latent parse built upon Recurrent Neural Networks (RNNs) which is bi-directional and utilized Long Short-Term Memory (LSTM). The experiments were conducted

Ref	Year	Machine Translation Technique	Attention mechanism	Copy mechanism
[14]	2014	RNN		
[15]	2017	RNN		
[16]	2015	RNN		
[17]	2019	RNN	✓	
[18]	2020	RNN		

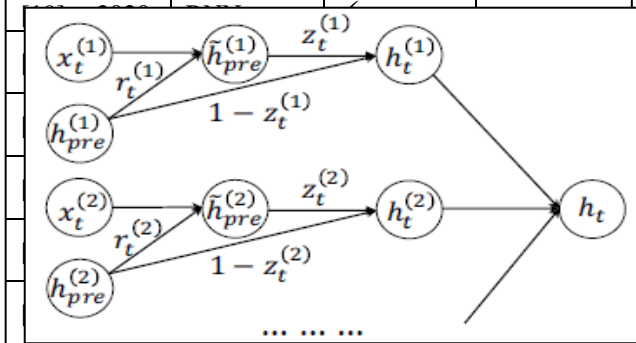


Figure 20: Deep Graph-Based Word-Lattice [25].

Table 1: Machine learning techniques, attention mechanism and copy mechanism.

Ref	Source Language	Target Language	Results
[14]	English	French	BLEU = 34.54
[15]	Hindi	English	BLEU = 44.1
	German	English	
	Russian	English	
	Others		
[16]	Japanese	English	BLEU = 73
[17]	English	Vietnamese	BLEU = 24.9
[18]	English	French	Accuracy = 97.37
[19]	Japanese	English	BLEU = 35
	English	Japanese	
[20]	Germany	English	BLEU = 25.36
[21]	English	German	BLEU = 29.7
	English	Chinese	
	Spanish	English	
[22]	English	Hindi	BLEU = 3.57
[23]			BLEU = 33
[24]	English	Japanese	BLEU = 39.42
[25]	English	Japanese	BLEU = 36.50

Table 2: Result and the source and target language of the machine Translation Techniques.

in Asian Scientific Paper Excerpt Corpus (ASPEC) to train the model to translate from English to Japanese. As a conclusion, the performance as along as BLEU and RIBES scores of the proposed model was improved compared with previous models. Even more, pre-training the model with small amount of annotation Treebank will be adding further improvements. The best result was achieved by Latent Graph Parsing for neural machine translation model with BLEU value equal to 39.42.

In order to use neural machine translation (NMT) to learn the input sentence semantic representation, then the word level modeling must be used. Thus, the sentences must be tokenized to get the words where the tokenization may cause two issues when using conventional NMT. The first issue was finding the best granularities of the tokenization process, while the second issue related to the possibility of propagating errors to the encoder by 1-best tokenization. On the other hand, to handle those problems, Su et al. [25] proposed to use NMT with word-lattice based Recurrent Neural Network (RNN) encoders, where the word lattice is a directed graph. The proposed encoder generalized RNN to word lattice topology by taking the word lattice as input where the word lattice encoded multiple tokens compactly as shown in Figure 6 [25]. However, the experiments were conducted on 1.25M sentence pairs with 27.9M Chinese words and 34.5M English words extracted from LDC2002E18, LDC2003E07, LDC2003E14, LDC2004T08, LDC2004T07 and LDC2005T06 to translate from Chinese to English. The best result for BLEU was 36.50.

3 Results and discussion

As shown in Table 1, much of the reviewed work concentrated on using encoder-decoder deep learning in order to produce the translated text. Since the encoder-decoder deep learning approach is commonly used from different size input-output applications. Also, we can see that some of the techniques are based on using graph while others are based on using phrase models.

4 Conclusion

Machine translation is one of the most important NLP applications. Several deep learning techniques can be used in machine translation but the main focus of this research is on recurrent neural networks. Recurrent neural network encoder-decoder model was used in most of techniques since the machine translation is based on having source language which we want to translate from and target language we want to translate to. We divided the techniques into three categories which are: recurrent neural network, recurrent neural network with phrase-based models and recurrent neural network with graph-based model. It be clearly seen that, most of the models used attention mechanism. The experiments were conducted on different datasets and several languages were used. In most of experiments, the machine

translation models were evaluated using BLEU evaluation measure.

5 References

- [1] D. Suleiman and A. Awajan, "Comparative study of word embeddings models and their usage in Arabic language applications," International Arab Conference on Information Technology (ACIT), Werdanye, Lebanon, pp. 1-7.2018. <https://doi.org/10.1109/ACIT.2018.8672674>
- [2] D. Suleiman, A. Awajan, and N. Al-Madi, "Deep Learning Based Technique for Plagiarism Detection in Arabic Texts," in 2017 International Conference on

Ref	Dataset
[14]	news commentary (5.5M), UN (421M), Bilingual corpora include Europarl (61M words) and two crawled corpora of 90M and 780M words
[15]	WMT11, WMT12, WMT13 and WMT14
[16]	Tanaka corpus consists of 150,000 sentence-pair collections Hand-Crafted Corpus
[17]	IWSLT English Vietnamese dataset
[18]	Windows Speech Recognition in the Microsoft Operating System, IBM Via Voice software
[19]	From Fukunishi, Yamamoto, Finch, and Sumita, Wikipedia inter-language link titles7 (2013)
[20]	TED talks dataset contains roughly 7K test, 7K development sentences, and sentences 153K training sentences
[21]	WMT 2016, UN parallel corpus and WMT 2013
[22]	MTIL2017
[23]	AMR corpus LDC2015E86 16833 instances for training, 1368 for development and 1371 for the test.
[24]	Asian Scientific Paper Excerpt Corpus (ASPEC)
[25]	Hansards portion of LDC2004T08, LDC2004T07, and LDC2005T06, 1.25M sentence pairs extracted from LDC2002E07, LDC2003E18, q LDC2003E14, With 27.9M Chinese words and 34.5M English words.

Table 3: The Datasets that are used in experiments.

- New Trends in Computing Sciences (ICTCS), Amman, pp. 216-222, 2017. <https://doi.org/10.1109/ICTCS.2017.42>
- [3] D. Suleiman, A. A. Awajan, and W. al Etaiwi, "Arabic Text Keywords Extraction using Word2Vec," in 2019 2nd International Conference on new Trends in Computing Sciences (ICTCS),

Amman, Jordan, pp. 1-7, 2019.

<https://doi.org/10.1109/ICTCS.2019.8923034>

- [4] C. Tapsai, P. Meesad, and C. Haruechaiyasak, "Natural Language Interface to Database for Data Retrieval and Processing," j.asep, May 2020. <https://doi.org/10.14416/j.asep.2020.05.003>
- [5] D. Suleiman and A. Awajan, "Bag-of-concept based keyword extraction from Arabic documents," in 2017 8th International Conference on Information Technology (ICIT), Amman, Jordan, pp. 863-869, 2017. <https://doi.org/10.1109/ICITECH.2017.8079959>
- [6] D. Suleiman, G. Al-Naymat, M. Itriq, "Deep SMS Spam Detection using H2O Platform," International Journal of Advanced Trends in Computer Science and Engineering 9(5):9179-9188, 2020. <https://doi.org/10.30534/ijatcse/2020/326952020>
- [7] D. Suleiman, M. Al-Zewairi, W. Etaiwi, G. Al-Naymat, "Empirical Evaluation of the Classification of Deep Learning under Big Data Processing Platforms," International Journal of Advanced Trends in Computer Science and Engineering 9(5):9189-9196, 2020. <https://doi.org/10.30534/ijatcse/2020/327952020>
- [8] A. Alqudsi, N. Omar, and K. Shaker, "Arabic machine translation: a survey," Artif. Intell. Rev., vol. 42, no. 4, pp. 549–572, Dec. 2014. <https://doi.org/10.1007/s10462-012-9351-1>
- [9] M. R. Costa-jussà and J. A. R. Fonollosa, "Latest trends in hybrid machine translation and its applications," Comput. Speech Lang., vol. 32, no. 1, pp. 3–10, Jul. 2015. <https://doi.org/10.1016/j.csl.2014.11.001>
- [10] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
- [11] Tan, Zhixing, et al. "Neural machine translation: A review of methods, resources, and tools." AI Open 1 (2020): 5-21. <https://doi.org/10.1016/j.aiopen.2020.11.001>
- [12] Stahlberg, Felix. "Neural machine translation: A review." Journal of Artificial Intelligence Research 69 (2020): 343-418. <https://doi.org/10.1613/jair.1.12007>
- [13] D. Suleiman and A. Awajan, "Deep Learning Based Abstractive Text Summarization: Approaches, Datasets, Evaluation Measures, and Challenges," in 2020 Mathematical Problems in Engineering, vol. 2020, pp. 1-29. <https://doi.org/10.1155/2020/9365340>
- [14] K. Cho et al., "Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation," ArXiv14061078 Cs Stat, Jun. 2014, [Online]. Available: <http://arxiv.org/abs/1406.1078>.
- [15] F. Guzmán, S. R. Joty, L. Márquez, and P. Nakov, "Machine Translation Evaluation with Neural Networks," ArXiv171002095 Cs, Oct. 2017, [Online]. Available: <http://arxiv.org/abs/1710.02095>.
- [16] E. Greenstein and D. Penner, "Japanese-to-English Machine Translation Using Recurrent Neural

- Networks,” Stanford Deep Learning for NLP Course, 2015.
- [17] M. Parmar and V. S. Devi, “Neural Machine Translation with Recurrent Highway Networks,” in International Conference on Mining Intelligence and Knowledge Exploration, 2019, pp. 299–308. https://doi.org/10.1007/978-3-030-05918-7_27
- [18] D. Datta, P. E. David, D. Mittal, and A. Jain, “Neural Machine Translation using Recurrent Neural Network,” *Int. J. Eng. Adv. Technol.*, vol. 9, no. 4, pp. 1395–1400, 2020. <https://doi.org/10.35940/ijeat.D7637.049420>
- [19] L. Liu, A. Finch, M. Utiyama, and E. Sumita, “Agreement on Target-Bidirectional Recurrent Neural Networks for Sequence-to-Sequence Learning,” *J. Artif. Intell. Res.*, vol. 67, pp. 581–606, 2020. <https://doi.org/10.1613/jair.1.12008>
- [20] P.-S. Huang, C. Wang, S. Huang, D. Zhou, and L. Deng, “Towards Neural Phrase-based Machine Translation,” Feb. 2018, Accessed: Jun. 01, 2018. <https://doi.org/10.1613/jair.1.12008>
- [21] L. M. Werlen, N. Pappas, D. Ram, and A. Popescu-Belis, “Self-Attentive Residual Decoder for Neural Machine Translation,” *ArXiv170904849 Cs*, Sep. 2017, [Online]. Available: <http://arxiv.org/abs/1709.04849>.
- [22] S. K. Mahata, D. Das, and S. Bandyopadhyay, “Mtil2017: Machine translation using recurrent neural network on statistical machine translation,” *J. Intell. Syst.*, vol. 28, no. 3, pp. 447–453, 2019. <https://doi.org/10.1515/jisys-2018-0016>
- [23] L. Song, Y. Zhang, Z. Wang, and D. Gildea, “A Graph-to-Sequence Model for AMR-to-Text Generation,” *ArXiv180502473 Cs*, May 2018, [Online]. Available: <http://arxiv.org/abs/1805.02473>.
- [24] K. Hashimoto and Y. Tsuruoka, “Neural Machine Translation with Source-Side Latent Graph Parsing,” *ArXiv170202265 Cs*, Feb. 2017, [Online]. Available: <http://arxiv.org/abs/1702.02265>.
- [25] J. Su, Z. Tan, D. Xiong, R. Ji, X. Shi, and Y. Liu, “Lattice-Based Recurrent Neural Network Encoders for Neural Machine Translation,” *ArXiv160907730 Cs*, Sep. 2016, [Online]. Available: <http://arxiv.org/abs/1609.07730>.

