Educalitra
English Education, Linguistics, and Literature Journal

Machine Translation Based on Neural Network: Challenge or Chance?

Xu Minyun

Hunan Normal University, China. E-mail: xuminyun221@qq.com

ARTICLE INFO

ABSTRACT

With the acceleration of economic globalization, there is a growing demand for translation services. In recent years, the quality of machine translation has been significantly improved. Compared with human translation, machine translation has the advantages of low cost and high speed. However, there is still little discussion on this scholarship, as this neural machine translation brings both convenience and pressure to translators. To address this issue, this paper will analyze the advantages and disadvantages of neural machine translation and discuss whether neural machine translation is a chance or a challenge for human translators.

1. Introduction

Machine Translation, a branch of Natural Language Processing (NLP), refers to the process of using a machine to automatically translate a natural language (source language) sentence into another language (target language) sentence (Li et al., 2018). The natural language here refers to the human language used daily (such as English, Chinese, Japanese, etc.), which is different from languages created by humans for specific purposes (such as computer programming languages).

According to statistics, about 5600 human languages exist (Jin, 2019). In China, as a family composed of 56 ethnic groups, some ethnic minorities also have their languages and scripts (Li, 2018). In other countries due to the history of colonization, these countries usually have multiple official languages, so some official documents usually need to be written in more than two languages. In the context of the "One Belt, One Road" initiative, communication among different languages has become an important part of building a community with a shared future for mankind. Therefore, machine translation technology can help promote national unity, communication between different languages, and cross-cultural communication.

Although the latest machine translation method, neural machine translation, has advantages such as speed and low cost, machine translation is still far from being as effective as human translation. Li (2021) selects Chronicle of a Blood Merchant, translated by Andrew F. Jones, a classic work of Yu Hua, and the versions of Baidu Translation, Youdao Translation, and

Google Translation as a corpus to conduct a comparative study on translation quality. Starting from the development of machine translation, Jin (2019) analyzed the advantages and disadvantages of machine translation and manual translation, discussing whether machine translation can replace human translation. To further explore the impact of machine translation on translators, this paper will take neural machine translation - the latest machine translation as an example to discuss whether machine translation is a chance or a challenge for translators.

2. Literature Review

2.1 Comparison of different machine translation methods

The development of Machine Translation methods goes through four stages: rule-based, instance-based, statistical, and neural machine translation. Thanks to the application of deep learning methods, neural machine translation has become mainstream. Compared with statistical machine translation, neural machine translation has the following advantages:

1) End-to-end learning does not rely on too many prior assumptions. In the era of statistical machine translation, model design makes more or less assumptions about the translation process. Phrase-based models, for example, assume that source and target languages are sliced into sequences of phrases with some alignment between them. This hypothesis has both advantages and disadvantages. On the one hand, it draws lessons from the relevant concepts of linguistics and helps to integrate the model into human prior knowledge. On the other hand, the more assumptions, the more constrained the model. If the assumptions are correct, the model can describe the problem well. But if the assumptions are wrong, the model can be biased. Deep learning does not rely on prior knowledge, nor does it require manual design of features. The model learns directly from the mapping of input and output (end-to-end learning), which also avoids possible deviations caused by assumptions to a certain extent.

2) The continuous space model of neural networks has better representation ability. A basic problem in machine translation is how to represent a sentence. Statistical machine translation regards the sentence generation process as the derivation of phrases or rules, essentially a symbol system in discrete space. Deep learning transforms traditional discrete-based representations into representations of continuous space. For example, a distributed representation of the space of real numbers replaces the discrete lexical representation, and the entire sentence can be described as a vector of real numbers. Therefore, the translation problem can be described in continuous space, which greatly alleviates the dimension disaster of the traditional discrete space model. More importantly, a continuous space model can be optimized by gradient descent and other methods, which have good mathematical properties and is easy to implement.

2.2 Principles of Neural Machine Translation

2.2.1 Word Representation

We know that machine translation is a branch of natural language processing. We need to decide how to represent individual words in a sentence. The first thing we do is come up with a vocabulary called a dictionary, which means making a list of the words we will use in our representations. We can then use one-hot representation to represent each word in a sentence, representing each word as a long vector. The dimension of this vector is the size of

the vocabulary, most of the elements are 0, and only one dimension has a value of 1, and this dimension represents the current word.

For example, "tiger" is represented as [0, 0, 0, 0, 1, 0, 0, 0, 0 ...] and "panda" is represented as [0, 1, 0, 0, 0, 0, 0, 0, 0 ...]. Therefore, we can label "tiger" as 4 and label "panda" as "1". However, a critical problem exists in one-hot representation: any two words are isolated. It's impossible to tell whether the words are related to these two vectors. A key idea is a new way of representing words called word embeddings. In Deep Learning, we commonly use not just one-hot representation but distributed representation, often called word representation or word embedding. Such a vector would look like this: [0.123, –0.258, –0.762, 0.556, –0.131 ...]. And the dimension of this vector is far smaller than the vector dimension represented by one-hot representation.

It can often let our algorithms automatically understand analogies like that man is to a woman. A king is to a queen and many other examples. And we will find a way to learn word embeddings later. We should know that these high dimensional feature vectors can better represent different words than one-hot vectors.

If words are represented in one-hot representation, it may cause a dimension disaster when solving certain tasks, such as building language models (Bengio 2003, 1137–1155). But using lower-dimensional feature vectors doesn't have this problem. Their complexity is almost unacceptable if high dimensional feature vectors are applied to deep learning. Therefore, low dimensional feature vectors are also popular here. In my opinion, the biggest contribution of word embedding is to make related or similar words closer in distance.

2.2.2 Recurrent Neural Network Language Model

Language modeling is one of the basic and important tasks in natural language processing. There's also one that Recurrent Neural Networks (RNNs) do very well. The language model built as RNNs is called Recurrent Neural Network Language Model (RNNLM).

What is a language model? For example, if we build a speech recognition system and hear the sentence, "The apple and pear(pair) salad were delicious.", what exactly did we hear? "The apple and pear salad was delicious." or "The apple and pair salad was delicious."? Humans might think that what we hear is more like the second. That's what a good speech recognition system helps output, even if the two sentences sound so similar. The way to get speech recognition to choose the second sentence is to use a language model that can calculate the probability of each sentence.

For example, a speech recognition model might calculate the probability of the first sentence being: P(The Apple and pear salad) = 2.6 × 10-13, whereas P(The Apple and pear salad) = 4.3 × 10-10. Compare the two values because the second sentence is 1,000 times more likely than the first, so the speech recognition system can choose the correct answer between the two sentences.

What the language model does is that it tells you what the probability of a particular sentence is. As we just mentioned, the language model is a fundamental component of speech recognition and machine translation systems, where translation systems want to output only likely sentences. The basic job of a language model is to input a sentence. Then the language model will estimate the probability of that particular word in a sequence of sentences.

How to build a language model with an RNN? First, we need a training set comprising a large corpus of English text or text from whatever language we want to build a language model. Here is the architecture of RNNML, which Mikolov references (Mikolov 2012). As the picture, RNNML predicts the probability of the 5th position word based on the previous 4 words, so this model is more likely outputs the sentence: "The students opened their books".
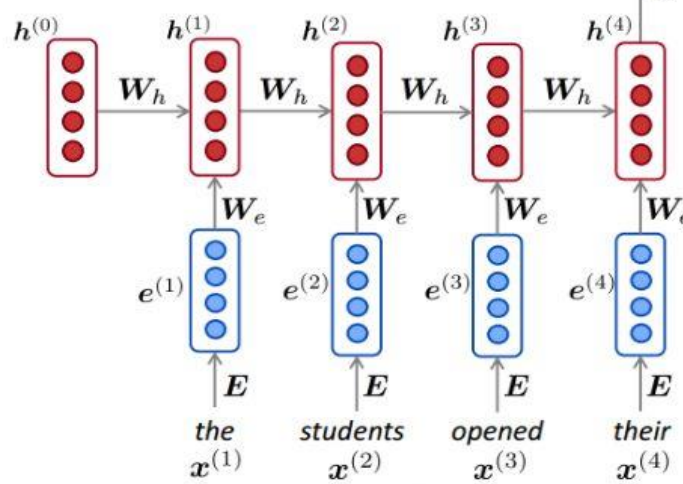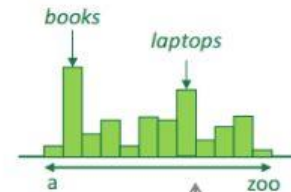
Figure 1. A RR Language Model

2.2.3 Encoders and Decoders

Actually, in machine translation, the length of the source language sentence and the target language sentence are generally different. Therefore, the common language models cannot meet the needs of machine translation. In 2014, the researchers (Sutskever et al. 2014) (Cho et al. 2014) designed new sequence-to-sequence models, often called encoder-decoder models.

Firstly, we should create a network, which we will call the encoder network, to be built as RNNs. Then, we should feed in the input a source language word at a time. And after ingesting the input sequence, the RNNs will output a vector representing the input sequence. And after that, we can build a decoder network, which takes the encoding output by the encoder network as input. The decoder network can be trained to output a translated word at a time until it eventually outputs the end of the sequence or the sentence token.

As shown in the figure below, the upper and lower dotted boxes respectively represent the encoder and decoder of neural machine translation, S and T sequences respectively represent

the source language sentences and the target language sentences, </S> represents the end of a sentence, and small circles represent feature vectors and neural network hidden layer.
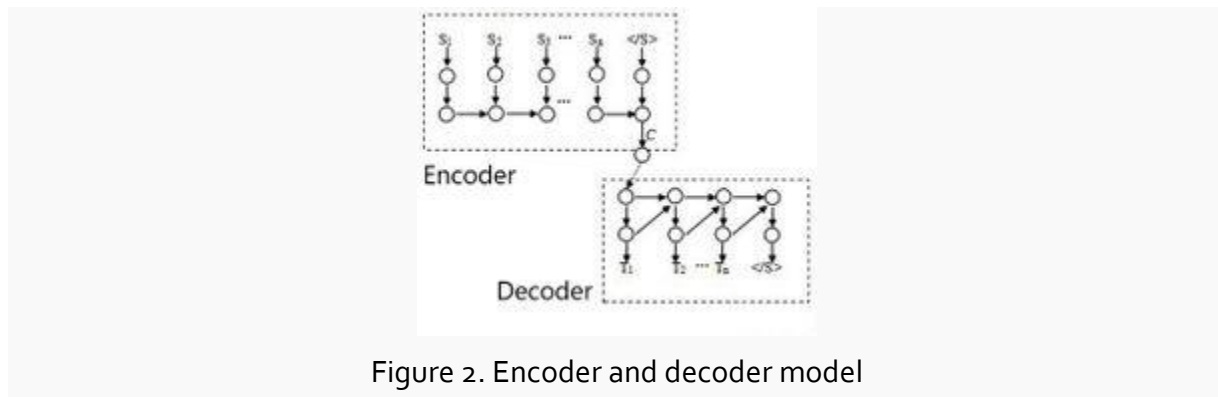

Figure 2. Encoder and decoder model

However, the model also has some problems, and the main problem is that only constant length vector C can be used to represent the entire source language sentences. In other words, regardless of the length of source language sentences, it can only be encoded as a vector C of fixed length. If it is a long sentence, the information contained by the constant length vector C will decrease or even disappear. Therefore, the problem of gradient disappearance or explosion exists in model training.

2.2.4 Attention Model

To solve the above problems, the researchers (Junczys-Dowmunt et al. 2016) introduced an attention model that dynamically captures context. The attention model is an improvement of the traditional neural network model. The basic idea is that each target language word has nothing to do with most of the source language words but only some words. By improving the representation of source language using a bidirectional recurrent neural network, vector representation containing global information can be generated for each source word.

As the picture shows, in the attention model, the encoder will use the forward recurrent neural network to encode the source language sentence from left to right in turn, generating a set of hidden states, and then use the backward recurrent neural network to generate another set of hidden states. Finally, the two sets of hidden states at the corresponding moments are spliced together to generate a new set of hidden states (Li Mu et al. 2018, 165), represented as the source feature vectors. Its advantage is that each source language feature vector representation contains the context information on the left and right sides, and the concatenation of the two means that it contains the entire source sentence information, so both can be used as the vector representation of the entire sentence.

There are some attention weights between encoder hidden states and decoder hidden states, and the weights are trained by comparing the hidden states of the encoder and decoder. The attention model will use these attention wights to weigh all the encoded hidden states by bit to obtain the source language sentence context vector C' at that moment. For example, when the attention model is generating the target word "T2", and "T2" is only most relevant to the source word "S1", but is irrelevant or under-relevant to other words, the weights between "T2" and "S1" will be large, and the other weights will be small. Repeatedly, we will get the target sentence finally.

3. Drawbacks of Neural Machine Translation

As a new technology, neural machine translation still has many problems. Recently, relevant, comprehensive studies mainly include improving attention mechanism, integrating priors and constraints, model training and fusion, new model, architecture construction, and evaluation of neural machine translation. The research on improving the quality of neural machine translation mainly includes the following:

1)Due to insufficient computational space, we cannot save all the words in word embeddings, so the problem of out-of-vocabulary words translation exists when we train models.

2)Because of the scarcity of language sources and insufficient training corpus, the minority languages translation becomes extremely difficult.

3)Because the model is sensitive to sentence length, it produces short results, resulting in translation difficulties in long and difficult sentences.

3.1 Translation of Out-of-Vocabulary Words

In the decoding process, neural machine translation needs to normalize the probability distribution with the help of the whole translation word embeddings, which has a large time and space consumption and high computational complexity. To control the temporal and spatial overhead, only high-frequency words are generally used in model training, and the number is limited to 30,000 to 80,000. All other uncovered words, often called out-of-set words, are uniformly identified as <UNK> characters. <UNK> characters mean that the semantic structure of parallel sentences in the training set is damaged, and the quality of model parameters will be affected. Also, it means that source language sentences are difficult to be expressed correctly, and the risk of ambiguity increases. Last but not least, unknown words will appear in the target language, and the quality and readability of the target language will be damaged. Moreover, language changes rapidly in this modern information society: old words add new ideas, new words continue to emerge, and named entities appear frequently. Therefore, the translation of out-of-vocabulary words is one of the basic topics in the research of neural machine translation, which is of great significance in improving the quality of neural machine translation.

Out-of-vocabulary word translation is a difficult issue in the current neural machine translation research. There are two ways to alleviate it. One is word granularity processing, mainly achieved by replacing out-of-vocabulary words. Although this method can reduce sentence ambiguity and improve the quality of translation and model parameters, it is not accurate enough to replace low-frequency and polysemous words, so it is difficult to deal with the problem in certain cases effectively. The other is the sub-word and character granularity processing method, which is the most popular method, and mainly solves the problems of parataxis language segmentation and hypotactic language deformation by reducing translation granularity and data sparsity. The sub-word and character granularity processing method does not need to use the out-of-vocabulary words processing module alone but decomposes out-of-vocabulary words into sub-words and characters, which are simple, effective, and widely used. However, fine-grained lexical segmentation may change semantic information and increase the number of sentence words, the risk of ambiguity, and the difficulty of training.

3.2 Translation of Minority Languages

According to statistics, there are about 5,600 languages in the world. The current neural network model for major languages will not be able to cope with the increasing translation needs in the era of big data. Therefore, it is of great practical value to improve the translation performance of neural network models under the condition of resource scarcity. Like statistical machine translation, neural machine translation is also a data-driven translation model, and its performance is highly dependent on the scale, quality, and breadth of the parallel corpus. The scale of artificial neural network parameters is huge, and only when the training corpus reaches a certain level neural machine translation will significantly surpass traditional statistical machine translation (Zoph et al. 2016, 1568). However, the reality is that, except for some vertical fields in major languages with relatively rich resources, most minority languages or vertical fields still lack large-scale, high-quality, and broad-covered parallel corpora. Therefore, how to use existing resources to alleviate the translation problems of minority languages is a focus of current research. The latest methods to deal with this problem mainly include zero-resource, data augmentation, and diverse learning methods.

The zero-resource method is an effective way to alleviate the neural machine translation problem in minority languages. The specific method is as follows: If there are three languages of, A, B, and C, and you want to realize the translation between A and C, there is no parallel corpus between them. But the parallel corpus of A and B is sufficient, and the parallel corpus of B and C is sufficient too. Then you can choose B as the pivot to realize the translation between A and C.

The data augmentation method can also effectively alleviate the insufficient generalization ability of the model due to the scarcity of training data. According to the currently available paper, data augmentation techniques used for neural machine translation mainly include back translation and word exchange.

Diverse learning methods, such as meta-learning, transfer learning, multi-task learning, unsupervised learning, and so on, are also effective ways to solve the shortage of minority language resources. However, the latest results are mostly seen in the latter two.

Although the methods mentioned above can alleviate the translation problems of minority languages to some extent, the breadth and depth of the experiment are still limited. Whether each method applies to all language pairs is worthy of in-depth discussion. In addition, for major languages, the information processing technology of minority languages is often more backward. In some languages, such as Mongolian, even the basic problems of part-of-speech tagging and named entity recognition are still poorly solved (Bao et al. 2018, 61). Therefore, while actively developing new training methods, we must also improve information processing technology in minor languages.

3.3 Translation of Long and Difficult Sentences

Due to the insufficient number of long sentences in the training corpus and the long-term memory problem of the cyclic neural network (Li Yachao et al. 2018, 2745). Therefore, the neural network model cannot translate long and difficult sentences. It only has a comparative advantage over traditional statistical machine translation in translating sentences within about 60 words (Koehn, Knowles 2017, 28). If this limit is exceeded, the quality will drop

sharply. Although the encoder-decoder model based on the attention mechanism can dynamically capture context information, solve the problem of information transmission over long distances, and improve the performance of neural machine translation, due to the complex structure of natural language, even the attention model cannot properly focus on all the information in the source language sentences. Therefore, in the translation of long and difficult sentences, there will be mistranslations such as over-translation and under-translation. Although the fluency of the target language has improved, semantic fidelity is worrying.

There are two main solutions to the problem of long and difficult sentence translation: one is to improve the capture ability of the model in the long distance; the other is to adopt the long sentence divide and conquer strategy. These two methods can alleviate the sensitivity of sentence length to a certain extent, but their effects still need to be improved. In view of the complexity and diversity of languages, not all languages can be divided and conquered, so the first method can be considered to improve the quality of long and difficult sentence translation in the future.

4. The prospect of Neural Machine Translation

In the long term, as a new technology, neural machine translation is in the ascendant and has a promising future. In recent years, especially since 2014, neural machine translation has rapidly progressed and developed. It is not only outstanding in traditional text translation but also excellent in image and speech translation. It is a machine translation model with great potential. The following trends are promising, and we believe that neural machine translation will have a brighter future as time goes on.

4.1 Unsupervised Translation

The recurrent neural network is suitable for processing sequential data, especially variable length sequential data, and is the mainstream implementation of traditional neural machine translation. The recurrent neural network is a typical supervised learning model. During the training process, it is highly dependent on a bilingual or multilingual parallel corpus, and the scale and quality of the corpus will directly restrict the translation performance of the model. However, the reality is that most languages do not have ready-made parallel corpora, and they are not naturally tagged, so it is expensive to process the corpus, such as alignment and labeling. At present, a number of scholars have tried to use different methods to achieve unsupervised translation, such as the unsupervised cross-language embedding method (Artetxe et al. 2018) and the latent semantic space-sharing method (Lample et al. 2018). Unsupervised translation has shown great development potential and will surely become one of the key exploration objects in the future.

4.2 Multilingual Translation

Barrier-free communication has always been a dream of mankind, and word vector technology allows it to realize the dream of Babel Tower. By mapping the words to the latent semantic space and using low-dimensional continuous real number vectors to describe their features (Li Feng et al. 2017, 610), we can avoid dimension disaster and improve semantic representation accuracy. More importantly, different languages can not only share the same semantic space but also share the same attention mechanism (Firat et al. 2016, 866), which lays a good foundation for multilingual neural machine translation. Although in practice,

whether word embeddings can be used to represent all language vocabulary remains to be verified, in theory, it does play the role of universal language. Using word embedding technology to improve the existing neural network model and achieve barrier-free communication is a topic to be explored.

4.3 Cross-cultural Communication

What is cross-cultural communication? Cross-cultural communication is the communication and information dissemination between social members with different cultural backgrounds. It also involves the migration process, diffusion, and change of various cultural elements in the global society and its impact on different groups, cultures, countries, and even the human community. Language is the carrier of culture.

Nowadays, cultural differences cannot be ignored. In my opinion, the quality of translation determines the quality of cross-cultural communication. If different cultures are compared to two lands that have never communicated with each other, then translation is a bridge of cross-cultural communication. The width and flatness of the bridge determine how many people can cross it on both sides, and the capacity and flow of the bridge are determined by the meaning that the translation can convey. Only more accurate and rigorous translations based on different cultures can make participants more extensive and enthusiastic, facilitating the smooth dissemination of culture.

Therefore, the ultimate goal of neural machine translation is to help translate high-quality cultural works, such as classic literature and good domestic animation, to let Chinese culture go out and absorb excellent foreign cultures.

5. Impact on Human Translation Market

With the development of neural machine translation, what impact has it been or will have on the human translation market? Whether neural machine translation is a challenge for human translators?

Firstly, the human translation market will not shrink, just as the textile machine has expanded the textile market. Some applications that require only "near" translation, such as the localization of some e-commerce sites, might be abandoned because of the high cost of human translation. But it might be survived with the help of machine translation. And in this process, the human translation market has created new demand called post-editing. When there was only human translation, this demand did not exist. But under machine translation, there were some human translation businesses. The machine translation market and the human translation market have expanded.

Secondly, high-end human translation will not die. Any excellent handmade work is still expensive today. After all, some translation tasks, such as the translation of poetry and literature is, a creative labor. We should know that machine translation is not sensitive to culture. Different cultures have unique language systems, so the translations they produce may not conform to the values and specific norms of the culture. Therefore, humans can use their unique advantages to translate certain translation tasks.

Finally, technological improvements often lead to improvements in efficiency, allowing people to complete their work more efficiently. The steam engine improves the efficiency of moving bricks but still requires a driver. Technology development will bring various positions

around machine translation, such as post-editing, quality controller, etc. Machine Assisted Translation (CAT) has become compulsory in many translation schools. If you can master these technologies proficiently, you will have the upper hand in the market.

6. Conclusion

Let's go back to the question at the beginning of this article: whether neural machine translation is a chance or a challenge for human translators. We can say that neural machine translation is a chance and a challenge for human translators.

From this paper, we know that neural machine translation has been developed a lot because of some critical methods such as deep learning, and machine translation quality has been greatly improved. Nowadays, machine translation also has a place in the translation market. However, as far as I am concerned, we should not pay too much attention to the impact of neural machine translation on human translation. We should discuss how to combine two types of translation services effectively.

As a new model, neural machine translation still has a long way to go. Improving the existing neural network model and making itself more intelligent is a major challenge. Therefore, we should look at machine translation from a dialectical and developmental perspective. Humans do not need to be afraid of technology but should learn to use it to enhance their work's efficiency and value.

References

Artetxe et al. (2018). *Unsupervised Neural Machine Translation*.

Bao et al. 包乌格德勒 等. (2018). 基于RNN和CNN的蒙汉神经机器翻译研究 [Mongolian-Chinese Neural Machine Translation Based on RNN and CNN]. 中文信息学报 *Journal of Chinese Information Processing (8)* 61.

Cho et al. (2014). *Learning phrase representation using RNN encoder-decoder for statistical machine translation*.

Firat et al. (2016). *Multi-way, Multilingual Neural Machine Translation with a Shared Attention Mechanism.*

Jin Wenlu 靳文璐. (2019). 机器翻译可以取代人工翻译吗? [Can machine translation replace human translation?]. 智库时代 *Think Tank Times* (40) 282-284.

Junczys-Dowmunt et al. (2016*). Is Neural Machine Translation Ready for Deployment*? A Case Study on 30 Translation Directions.

Koehn, Knowles. (2017). *Six Challenges for Neural Machine Translation*.

Lample et al. (2018). *Unsupervised Machine Translation Using Monolingual Corpora Only.*

Li Mu et al. 李沐 等. (2018). 机器翻译 [*Machine Translation*]. Beijing: Higher Education Press 高等教育出版社.

Li Yachao et al. 李亚超 等. (2018). 神经机器翻译综述 [A Survey of Neural Machine Translation]. 计算学报 *Chinese Journal of Computers (12)* 2745.

Li Yao 李耀. (2021). 基于语料库的机器翻译文学作品质量研究--以《许三观卖血记》为例 [A Corpus-based Study on the Quality of Machine Translation Literary Works-- Taking *Chronicle of a Blood Merchant* as an example]. 海外英语 *Overseas English* (18) 39-40+42.

Mikolov Tomáš. (2012). Statistical Language Models based on Neural Networks. *Ph.D. thesis, Brno University of Technology.*

Sutskever et al. (2014). *Sequence to sequence learning with neural networks*.

Yoshua Bengio et al. (2003). A neural probabilistic language model. *Journal of Machine Learning Research (JMLR) (3)* 1137–1155.

Zoph et al. (2016). *Transfer Learning for Low-resource Neural Machine Translation*.