# Chapter 38
# Machine Translation Using Deep Learning: A Comparison

**S. Swathi and L. S. Jayashree**

**Abstract** In recent days, machine translation is rapidly evolving. Today one can find several machine translation systems that provide reasonable translations, although they are not perfect. The main objective of machine translation is to provide interaction among the people speaking two different languages. Machine translation, being an important task of natural language processing, leads to the development of different approaches, namely, rule-based machine translation, statistical machine translation, and neural machine translation for the translation process. The recently proposed method is the neural machine translation which improves the quality of translation between natural languages through neural networks. Neural machine translation led to remarkable improvements in the translation process by retaining the contextual information. End-to-end neural machine translation uses RNN Encoder-Decoder mechanism to train the neural translation model with bilingual corpora which is bilingual parallel sentence pairs, an important resource of machine translation. NMT has a reasonable BLEU score which is the evaluation metrics for machine translation. In this paper, we present a survey on the different kinds of machine translation approaches with their strengths and limitations and the various evaluation metrics to measure the accuracy of the translation.

**Keywords** Machine translation · Natural language processing · Rule-based translation · Statistical machine translation · Neural machine translation

## Abbreviations

RNN     Recurrent neural network
LSTM    Long short-term memory
NMT     Neural machine translation

S. Swathi (✉) · L. S. Jayashree (✉)
Department of Computer Science and Engineering, PSG College of Technology, Coimbatore, Tamil Nadu, India

BLEU    Bilingual evaluation understudy
NIST    National Institute of Standards and Technology
TER     Translation error rate

## 38.1   Introduction

Machine translation is one of the most active research areas in natural language processing. Machine translation refers to the software that can perform the task of instant translation of the input from the source language to the target language in a short period of time. While performing the translation task, either human or automated, the meaning of an input sentence in the source language must be fully restored in the target language. Machine translation does not mean simply a word-for-word substitution. The translator must interpret and analyze the meaning of all the words in a sentence and know how each word influences the others. To achieve this kind of better translation, various approaches of machine translations evolved. Machine translation has been adapted which decrease the effect of language as a barrier of communication. Also the reason for choosing automatic machine translation rather than human translation is that they are better, faster, and cheaper than human translation.

Today a number of machine translation system are available that are capable of translating to a sufficient extent but not a perfect system which are used in several fields. The great challenge of machine translation is to handle various types of idioms and phrases which may vary according to the language, and it may take several trainings to improve the translation. Hence it will be very difficult for humans to design the perfect translation system. In this paper, we presented the study of various machine translation approaches that have been proposed for the task of translation.

## 38.2   Various Approaches for Machine Translation

### 38.2.1   Rule-Based Machine Translation

Rule-based machine translation mostly involves the dictionary creation for each word in the parallel corpus. It translates based on the syntactic and semantic rules constructed based on the linguistic knowledge of the languages used. It requires a large number of linguistic rules generated by analyzing the grammar of both the source and the target languages. This involves generating the intermediate representation from which the sentence in the target language is obtained [1]. It has the following sub-categories.

#### 38.2.1.1 Direct Machine Translation

Direct machine translation is just a word-by-word substitution based on the bilingual-dictionary generated. This system takes the source language as an input and produces the target language as an output based on the linguistic rules in the dictionary. It does not consider the relationship of the words in the sentence and hence results in ambiguity problem.

#### 38.2.1.2 Interlingual Machine Translation

Interlingual machine translation is not merely a dictionary mapping. This system transforms the source sentence into an interlingual language, i.e., a neural language independent of both the source and the target languages [2]. Thus the target sentences are produced from the abstract interlingual representation. But the problem is in defining such an interlingual language, and the translation is independent of target language.

#### 38.2.1.3 Transfer-Based Machine Translation

Transfer-based system, on the analysis of source language, generates the source intermediate (bilingual dictionary) which is then transferred into the target intermediate (target dictionary), i.e., both these dictionaries are used to produce the target language [3]. This retains the meaning of the sentence to an extent, which is better than the other two rule-based machine translation systems.

### 38.2.2 Corpus-Based Machine Translation

Corpus-based machine translation is based on the statistical analysis of the source and the target language which requires a huge volume of data for training the system. This includes two approaches for machine translation.

#### 38.2.2.1 Statistical Machine Translation

Statistical machine translation system performs translation better than the traditional rule-based translation system. It works based on statistics, i.e., they think in probabilities. It has different models for translation [1]. At the first stage, it generates many possible target sentences for a given input source sentence, i.e., it generates the possible translation for each word (or) phrase of a sentence and assigns the probability for each translation. The language model computes the probability of the target

language P(T). In the next stage, it selects the best candidate based on the scoring function. The translation model computes the probability of the translation for the target sentence given the source sentence P(T|S) which has to be maximized results in the better translation to an extent.

$$P(T|S) = \text{argmax } P(T) * P(S|T) \qquad (38.1)$$

The statistical machine translation system generates many possible translations of a word and ranks them rather than simple word-by-word substitution. This reduces the ambiguity of a sentence but hugely depends on the bilingual corpora.

#### 38.2.2.2 Example-Based Machine Translation

An example-based machine translation system has two modules: retrieval module and an adaptation module. The input source for the translation process is the bilingual corpus. The retrieval module retrieves a sentence from the input bilingual corpus which is similar to the given source sentence along with its translation. The adaptation module then adapts the retrieved translation to get the final corrected translation. It depends on the similarity and hence is suitable for the similar structure languages.

### 38.2.3 Neural Machine Translation

The recently used machine translation approach is the neural machine translation which depends on purely deep neural networks. This system uses the RNN Encoder-Decoder LSTM model to translate the source language to the target language which will retain the contextual information of the sentence [4].

The RNN Encoder-Decoder model consists of two recurrent neural networks (RNN). Among them, one does the encoding process and the other the decoding which are required for the translation. The encoder maps a variable-length input in a source language to a fixed-length vector, and the decoder maps the vector representation back to a variable-length output in a target language.

LSTM Neural Networks, which stand for Long Short-Term Memory, are a particular type of recurrent neural networks that have some internal contextual state cells that act as long-term or short-term memory cells. Machine translation needs to depend on the historical context of inputs to retain the meaning of the given sentence, rather than only on the very last input. RNN Encoder-Decoder together with LSTM captures both semantic and syntactic structures of the phrases from the source sentence to restore the contextual information [5].

Let S and T represent the source and target words of those sentences, respectively. The encoder RNN converts the source sentences S1, S2...SM into vectors of fixed dimensions. The decoder outputs one word at a time using conditional probability

$$P(T|S) = P(T|S1, S2, S3, \ldots, SM) \qquad (38.2)$$

that is, the occurrence of a word depends on the sequence of previous input words to restore the meaning of the given sentence [6].

## 38.3    Comparison Among the Different Machine Translation Techniques

From the earlier days, machine translation has been adapted in various fields with various machine translation techniques. For instance, machine translation for watcher (MT-W) focused on the readers to gain information sourced in other languages who accepted the bad translation rather than nothing. Similarly, a type of machine translation called machine translation for translator had been proposed which focused on the human translators by providing online dictionaries. Those types of machine translation adapted different machine translation techniques to perform the task of translation. On summarizing the evolution of the different types of machine translation, the recently proposed technique called neural machine translation results in better quality of translation. The comparison of the various machine translation techniques with their advantages and disadvantages is shown in Table 38.1.

**Table 38.1**  Machine translation techniques comparison

| Machine translation | Advantages | Disadvantages |
|---|---|---|
| Direct machine translation | Easy to implement | Considers only lexical words |
| | Effective for languages with similar grammar rules | Does not consider the relationship among the words |
| Interlingual machine translation | It suits only when the number of target languages is more | Defining the neutral language representation (universal abstract) |
| Transfer-based machine translation | It depends partially on the language pair and hence less complex | Can have loss in meaning for some source text |
| Statistical machine translation | No need for manual development of linguistic rules | Depend on huge amount of parallel corpus |
| | Generic applicability to any pair of languages | Results are not accurate |
| Neural machine translation | Focuses on contextual information than other MT | Make use of efficient semantic structure of data and hence the complex system |

## 38.4   Evaluation Metrics

Evaluation of performance of the translation system by humans is not always accurate. Also it takes time and is very expensive. Hence the evaluation is done via the following metrics.

### 38.4.1   Bleu Score

The Bilingual Evaluation Understudy Score, or BLEU, is a metric for evaluating a generated sentence to a reference sentence. BLEU is an algorithm for evaluating the quality of text which has been machine-translated from one natural language to another, i.e., the accuracy of the translation. A perfect match of the output sequence with the referred sentence results in a score of 1.0, whereas a perfect mismatch results in a score of 0.0. Few translations will attain a score of 1 unless they are identical to a reference translation, but this is not possible even with the human translation. This metric is based on n-gram score.

Ex**:**

System Output:  "the cat", "cat is", "is here"
Reference Sentence: "the cat", "cat is", "is on", "on the", "the wall"

"the cat": 1
"cat is": 1
"is here": 0
Bigrams  score = (1+1+0)/3 = 2/3 = 0.67

### 38.4.2   NIST Score

NIST is also the metric for evaluating the quality of machine translation based on BLEU score. Rather than calculating the simple n-gram score, this metric also calculates how informative the particular n-gram is. The high preference is given to the n-grams which are more informative. This score is designed to improve the BLEU score.

Ex: The cat sat on the wall.
The bigram "cat sat" are given more weightage than for the bigram "on the".

### 38.4.3   Translation Error Rate

The Translation Error Rate (TER) metric calculates the number of post-editing actions required to match the translated sentence with the reference sentence. This metric can be defined as follows:

$$\text{TER} = \frac{E}{w_R} \qquad (38.3)$$

where

$E$ denotes the minimum number of edits required for an exact match with the reference text.

$w_R$ denotes the average length of the reference text.

## 38.5   Conclusion

In this paper, various machine translation techniques which are adapted to provide the automated translation process have been reviewed. Also, this paper has presented about the various evaluation metrics that can be used to estimate the quality of different automated machine translation systems. Among those machine translation systems, neural machine translation provides more accuracy with the huge bilingual corpora as LSTM cells are used which can store the meaning of the text for a period of time. This helps in restoring the contextual information in the translated text and achieves greater quality in the translation process.

## References

1. Saini S, Sahula V (2015) A survey of machine translation techniques and systems for Indian languages. In: IEEE international conference on Computational Intelligence & Communication Technology, pp 675–681
2. Ashraf N, Ahmad M (2015) Machine translation techniques and their comparative study. Int J Comput Appl 125(7):25–31
3. Harjinder Kaur, Vijay Laxmi (2013) A survey of machine translation approaches. Int J Sci Eng Technol Res 2(3)
4. Zakaria El Maazouzi, BadrEddine El Mohajir, Mohammed Al Achhab (2017) A technical reading in statistical and neural machine translation system. In: IEEE international conference on Information Technology, pp 157–165
5. Xiao-Xue Wang, Cong-Huizhu, Sheng Li, Tie-Jun zhao (2016) Neural machine translation research based on the semantic vector of the tri-lingual parallel corpus. In: International conference on Machine Learning and Cybernetics
6. Nakamura N, Isahara H (2017) Effect of linguistic information in neural machine translation. International conference on Advanced Informatics, Concepts, Theory and Applications