# **Research of types and current state of machine translation**

### Yanran Wang

School of Information Science and Technology, Beijing University of Chemical Technology, Beijing, China

### 2021040107@buct.edu.cn

Abstract. The background of machine translation dates back to the 1950s when scientists began exploring the use of computers for translation. Motivated by various challenges and needs, such as the high cost of manual translation and cross-cultural communication, machine translation has become a pivotal field. This overview delves into the research background, content, methods, key figures, conclusions, and future prospects of machine translation. It summarizes automatic evaluation metrics, corpus construction, and transfer learning, all of which contribute to enhancing translation performance. Currently, there are three mainstream categories of methods, which include rule-based translation, statistical translation, and neural network-based translation. The rule-based translation involves the use of extensive bilingual corpora for identification and translation. The conclusion emphasizes the potential of neural machine translation, yet acknowledges challenges in diverse languages, low-resource languages, and specialized terminology.

Keywords: Machine translation, types, analysis, deep learning.

### 1. Introduction

The development of machine translation holds immense significance for humanity, as it saves time and costs associated with manual translation [1-2] and enables better educational and academic exchanges between different countries. With improved means of information dissemination, nations can also foster economic development through enhanced communication. Machine translation also serves as a driving force [3] behind the advancement of natural language processing and other artificial intelligence technologies. On an individual level, machine translation aids in breaking down language barriers. Currently, various machine translation methods are rapidly progressing and evolving, promising a bright future for the field.

Machine translation, as a pivotal research direction in the field of natural language processing, has made significant strides in recent years. Traditional rule-based and statistical methods have gradually been supplanted by neural network models, with the emergence of the Transformer model, in particular, leading a new wave of technological revolution. Pretrained language models like Bidirectional Encoder Representations from Transformers (BERT) and the Generative Pre-trained Transformer (GPT) [4] series have also paved new paths for machine translation, substantially enhancing translation quality and efficiency.

Machine translation finds a wide array of applications in the backdrop of globalization. From business communication to cultural exchanges, cross-border e-commerce to travel guides, machine translation plays a crucial role in breaking language barriers and facilitating cross-cultural communication. Additionally, in domains such as emergency response and medical diagnostics, machine translation provides convenience in transmitting information between different languages.

The research in machine translation not only caters to the needs of global communication [5], but also propels the development of natural language processing and the field of artificial intelligence. By addressing the communication gap between various languages, machine translation contributes to fostering cross-cultural understanding, driving international collaboration, and thus promoting societal progress. Moreover, the research in machine translation spans across semantic understanding, syntactic analysis, multimodal translation, and various other domains, fueling innovation in related technologies. This article will analyze machine translation from the aspects of its functionalities, advantages and disadvantages, and future prospects.

### 2. Evolution of Machine Translation

The trajectory of machine translation's development has been remarkably swift, progressing from its inception in the 1940s to a more refined state within less than a century. The rudiments of machine translation emerged even before the full-fledged advent of computers. However, serious research in this area began in the late 1940s, led by linguistic pioneer Warren Weaver and others. The evolution of machine translation can be roughly categorized into three stages spanning from the 1940s to the present:

Stage One: Emergence and Initial Development (1949-1960)

In 1949, Warren Weaver, a pioneer in information theory, published a memorandum discussing the feasibility of machine translation, which expedited the rapid progress of the field. Taking inspiration from Weaver's ideas, the first Machine Translation Conference was organized by Bar-Hillel in 1952. In 1954, with the assistance of IBM Georgetown University [6] conducted translation experiments, marking the commencement of automated translation experiments. During this phase, the translation quality was limited, resources were constrained, and importantly, the practicality of machine translation was lacking.

### Stage Two: Stagnation in Development (1960-1990)

In the late 1950s and early 1960s, researchers encountered inconsistent translation output quality from machine translation systems, raising doubts about the feasibility of further development. Bar-Hillel attributed the subpar quality to semantic ambiguities, highlighting the limitations of syntactic systems and semantic analysis [7] at that time. These concerns led to the establishment of the Automatic Language Processing Advisory Committee (ALPAC), tasked with evaluating machine translation efforts. In 1966, ALPAC issued the report "Language and Machines: Computers in Translation and Linguistics," which critically reviewed machine translation techniques, noting the difficulties of achieving high-quality translations through traditional rule-based methods. These findings resulted in reduced funding for machine translation research and ushered in a relatively quiet period for early machine translation efforts.

Stage Three: New Era in Machine Translation Development (1990-Present)

Starting from the 1990s, the emergence of substantial bilingual and multilingual corpora facilitated the development of machine translation approaches that departed from traditional methods. With the evolution of deep learning, machine translation gained greater accuracy in modeling and semantic analysis. The advent of translation patterns like multimodal translation and low-resource translation expanded the applicability of machine translation to various domains and directions. The emergence of methods like neural network machine translation enhanced accuracy and speed through multi-level nonlinear processing, outperforming traditional lower-level linear processing in problem analysis and performance.

Furthermore, the introduction of the Transformer model reshaped the landscape of machine translation. Proposed by Vaswani and others in 2017, the Transformer model, a neural network architecture based on attention mechanisms, addressed limitations posed by traditional recurrent and

convolutional networks in processing lengthy sequences. The attention mechanism in the Transformer model allowed it to consider various aspects within sentences, facilitating better capture of semantic relationships. Overall, the introduction of the Transformer model revolutionized the field of natural language processing, becoming a cornerstone in various NLP tasks, notably achieving remarkable success in tasks like machine translation, text generation, and text classification.

The emergence of GPT showcased powerful capabilities in diversified text generation [8]. GPT, a series of pretrained language models introduced by OpenAI, aimed to capture statistical properties and semantic relationships in language through extensive pretraining on text data. Particularly, GPT-3, unveiled in 2020, garnered significant attention as the largest-scale pretrained language model thus far.

### 3. Machine Translation Methods

Machine translation methods can be broadly categorized into rule-based, statistical, and neural networkbased approaches. This section will specifically discuss these machine translation methods.

# 3.1. Rule-Based Machine Translation

Rule-based machine translation represents a more traditional approach to translation. Its fundamental concept involves transforming source language text into an intermediate representation and then converting it to the target language text to achieve translation. Machines encode the meaning of the source language through decoding [9], combining it with the linguistic traits and grammatical rules of the target language to synthesize the translation.

# • Basic Steps of Rule-Based Machine Translation

1. Language Analysis:\*\* Analyze the source language text to identify lexical items, grammatical structures, and semantic information.

2. Rule Matching:\*\* Match the analyzed source language text with target language rules based on pre-defined translation rules [10].

3. Structural Transformation: Apply the matched rules to the source language text, performing structural transformation and transcription to convert it into the target language's structure.

4. Language Generation: Generate the translation of the target language based on the transformed structure.

Source language sentences are initially parsed into an intermediate language (interlingua) representation, a universal language structure independent of specific source or target languages. The intermediate language representation is then converted into the target language sentence. The core idea here is to bridge the source and target languages using the intermediate language, requiring the implementation of conversion rules for both source-to-interlingua and interlingua-to-target (Figure 1).

Source language sentences are first parsed into an intermediate representation, followed by the utilization of a series of rules to convert the intermediate representation into a target language sentence. The Transfer method focuses on establishing structural and syntactic correspondences between source and target languages, using rules to describe how to transform the source language structure into the target language structure for translation.



Figure 1. Basic Steps of Rule-Based Machine Translation

# • Advantages and Disadvantages of Rule-Based Machine Translation

Rule-based machine translation relies heavily on linguistic knowledge for the construction of translation dictionaries containing information on the morphological, syntactic, and semantic aspects of the source language. This approach necessitates the establishment of a knowledge base containing linguistic knowledge and transformation rules [11]. As a result, rule-based machine translation heavily depends on parallel corpora, faces difficulties in low-resource and specialized domain translation, and demands extensive manual work and domain expertise for rule writing and maintenance. Moreover, for complex language structures and instances of polysemy, rule-based machine translation might not yield ideal results.

The strengths of rule-based machine translation lie in its interpretability and controllability, allowing flexibility in adjusting and modifying rules to suit translation tasks in different domains, resulting in more accurate and specialized translations. Compared to neural network and statistical machine translation, it involves smaller models and requires less training data.

### 3.2. Statistical Machine Translation

Statistical machine translation, from its emergence in the 1990s to the present, was pioneered by IBM's Brown et al. in 1993, who proposed complex models of increasing complexity known as IBM Models 1 through 5, forming the basis of statistical machine translation [12]. It's a machine translation approach based on statistical models, involving learning translation rules and patterns from extensive parallel corpora to automate the translation process. Parallel corpora consist of pairs of bilingual sentences: one in the source language and its corresponding translation in the target language. Statistical machine translation can be further divided into word-based machine translation, phrase-based machine translation, and syntax-based statistical machine translation.

# • Basic Process of Statistical Machine Translation

1. Data Preprocessing: The first step involves preprocessing the training data, which consists of large-scale parallel corpora. The purpose of preprocessing is to prepare the data for building statistical models.

2. Model Training: This stage comprises two phases: translation model and language model. The former evaluates translation probabilities, while the latter assesses the fluency of target language sentences.

3. Decoding: Given a source language sentence, the statistical machine translation (SMT) system selects the optimal translation from a set of possible target language sentence candidates (Figure 2).

4. Post-processing: After obtaining the optimal target language sentence, post-processing tasks such as spell-checking and grammar correction are performed to enhance overall translation quality [13].

5. Evaluation: Once translation is complete, the translation results need to be evaluated.

Different statistical machine translation systems may employ varying translation steps to maximize the enhancement of sentence quality, such as translation rule extraction and rule filtering.



Figure 2. Architecture of SMT system

Statistical machine translation involves translation inference based on probability models. The system calculates probabilities for different translation results and selects the highest probability translation as the output. Contextual information and the sentence's context are taken into account to improve translation accuracy and fluency.

Statistical machine translation can automatically learn translation rules from data, exhibiting strong adaptability without the need for manual rule writing. Additionally, it can extensively utilize external resources to enhance translation quality, making it a mainstream machine translation method.

However, the performance of statistical machine translation is limited by data quality and scale. It requires extensive training data to achieve favorable translation outcomes. Furthermore, it might perform inadequately in scenarios involving complexities such as polysemy.

With the rise of deep learning, neural network machine translation has gradually replaced statistical machine translation as the mainstream approach. Nonetheless, statistical machine translation still holds research value. Representative approach has Maximum Entropy Model (MEM). In MEM,

Features from source and target languages are extracted and modeled using the Maximum Entropy Model. This model captures relationships between these features, enabling computation of translation probabilities [14]. The Maximum Entropy Model selects the most suitable probability distribution by maximizing the principle of entropy. In machine translation, it is employed to model translation probabilities, reordering probabilities, and dependencies between features.

#### 3.3. Neural Network Machine Translation

Similar to how the human brain processes complex information, neural network machine translation is a natural language processing technique that employs neural networks to translate text from one language to another. With the development of deep learning and neural networks, researchers began exploring the application of neural networks to machine translation around 2014. The Seq2Seq architecture, first introduced in this context, achieved certain successes in translation tasks. In 2015, Dzmitry Bahdanau et al. proposed the attention mechanism, which improved the connection between source and target languages, elevating translation quality. Neural network translation models have made significant achievements in the field of translation, including more accurate and fluent translation results and enhanced contextual understanding. Neural network translation models excel in translating between different language pairs, facilitating easier communication in cross-lingual and cross-cultural interactions. Additionally, neural network machine translation provides robust support for cross-lingual information retrieval, multilingual content creation, and globalized businesses.

### 4. Principles and Modules of Neural Network Machine Translation

Encoder-Decoder Architecture: Neural network machine translation employs an encoder-decoder structure. The encoder converts the input sentence into a fixed-length vector, while the decoder generates the output sentence using the context vector. The encoder typically uses recurrent neural networks to process the input sentence word by word, encoding each word's information as hidden states and generating the final context vector. The decoder starts with the context vector produced by the encoder and progressively generates words or characters of the output sentence.

Attention Mechanism: The attention mechanism is a crucial component that allows the decoder to focus on different parts of the input sentence during output generation, enhancing performance, particularly in translating longer sentences [15].

Loss Function and Training: Defining an appropriate loss function is crucial for training neural network machine translation models. Cross-entropy loss is commonly used to measure the difference between predicted outputs and actual outputs. Training data consists of pairs of parallel sentences containing the source language sentence and its corresponding target language sentence.

### • Advantages of Neural Network Machine Translation over Traditional Methods

Enhanced translation quality by better capturing contextual information between sentences.

End-to-end learning approach eliminates the need for manual rule writing, providing translation results directly from the input source language sentence.

Leverages large-scale bilingual data for training, offering good automation and scalability.

### 4.1. Efficiency Comparison and Combination of Translation Methods

Efficiency comparison among machine translation methods requires preparation in terms of data readiness, training and inference speed, and translation quality.

In data preparation, rule-based machine translation involves manual creation of complex translation rules, requiring time and specialized knowledge. Statistical machine translation demands alignment of parallel corpora, phrase extraction, and the like, and is more efficient than rule-based but still involves some manual engineering. Neural network machine translation simplifies data preparation, approaching end-to-end training and reducing the initialization workload of translation systems.

In terms of training and inference speed [16], rule-based machine translation is slow due to its complexity. Statistical machine translation speed depends on feature extraction and phrase table construction, making it slightly slower than neural network machine translation. Neural network machine translation benefits from hardware acceleration, achieving faster training speeds.

Regarding translation quality: Rule-based machine translation performs well on simple sentences. Statistical machine translation improves translation quality, while neural network machine translation generally excels across most tasks.

Combining different machine translation methods is a way to address their individual limitations. This combination aims to utilize complementary features of different methods to achieve better performance in translation [17]. Errors and weaknesses of one method can be rectified and compensated for by another, leading to enhanced translation quality. Such combination can occur at various levels, including word-level, phrase-level, and even sentence-level.

### 5. Conclusion

This article delves into the evolution of machine translation across different stages and presents the characteristics and principles of three distinct machine translation methods. From the rule-based

traditional approach to statistical machine translation, and to the recent emergence of neural network machine translation, the article comprehensively outlines their respective features and development trajectories. The advantages of neural networks, particularly their attention mechanism's prowess in handling long sentences and complex structures, are highlighted. Additionally, the article mentions the trend of integrating different methods, such as combining data preprocessing, model architecture, and post-processing, to achieve higher translation quality and efficiency. However, the review notes that challenges persist in the machine translation field, including translating low-resource languages and tackling cross-lingual difficulties. As technology continues to advance, emerging areas like transfer learning and multimodal translation are poised to bring new breakthroughs to machine translation. In essence, it is hoped that readers gain a comprehensive guide to the evolution of machine translation and a glimpse into the potential future directions.

# References

- [1] Liu Qun. Review and Prospect of Machine Translation Research. *Journal of Chinese Information Processing*, 2018, **29(1)**: 1-9.
- [2] Hutchins W. J. Machine Translation: A Concise History. *In Machine Translation History*. 3-48, 2007.
- [3] Bai Shize. Development Process of Machine Translation Technology. *Computer Engineering*, 2017, **31**(3): 1-3.
- [4] Brown, T. B., Mann, B., Ryder, N., Subbiah, Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 2020, **33**.
- [5] Hutchins W. J. Machine Translation: A Concise History. John Benjamins Publishing, 1986.
- [6] Koehn, P., & Knight, K. Empirical methods for compound splitting. *Conference of the European Chapter of the Association for Computational Linguistics* 2003. 187-193.
- [7] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Attention is all you need. *Advances in neural information processing systems*, 2017, **30**.
- [8] Radford, A., & Salimans, T. Improving language understanding by generative pretraining. *OpenAI*, 2018.
- [9] Rajman, M. Experience with a rule-based machine translation system for French-English. *Machine translation*, 1998 **3(2-3)**, 163-184.
- [10] Liu, Y., & Jiang, W. Rule-based statistical machine translation. *Tsinghua University*. 2013
- [11] Hutchins, W. J. Machine translation: A concise history. In History of machine translation, 2007 3-48.
- [12] Koehn, P., & Hoang, H. Factored translation models. 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. 868-876.
- [13] Och, F. J., & Ney, H. A systematic comparison of various statistical alignment models. Computational Linguistics, 2003 **29(1)**, 19-51.
- [14] Brown, P. F., Della Pietra, S. A., Della Pietra, V. J., & Mercer, R. L. The mathematics of statistical machine translation: Parameter estimation. *Computational Linguistics*, 1993, **19**(2), 263-311.
- [15] Johnson, M., Schuster, M., Le, Q. V., Krikun, M., Wu, Y., Chen, Z.. Google's multilingual neural machine translation system: Enabling zero-shot translation. Transactions of the Association for Computational Linguistics, 2017, 5, 339-351.
- [16] Gehring, J., Auli, M., Grangier, D., Yarats, D., & Dauphin, Y. N. Convolutional Sequence to Sequence Learning. *34th International Conference on Machine Learning*, 2017, 1243-1252.
- [17] Vaswani, A., Bengio, S., Boulanger-Lewandowski, N., & Bengio, Y. Axiomatic Memory Networks. *arXiv preprint arXiv*:1310.6299, 2013.