On Neural Machine Translation Based on Cloud Platform in the Context of Artificial Intelligence

Renna Gui(✉)

School of Foreign Languages, Nanchang Institute of Technology, Nanchang, Jiangxi, China
86050651@qq.com

Abstract. The research of machine translation has gone through the era of machine translation based on grammar rules, case data and statistical methods. Up to now, it has become a neural machine translation designed with coding and decoding as the basic framework and using neural network to model the translation process. It is illustrated how the human-computer co-translation can be applied based on translation technology cloud platforms so as to greatly enhance the translation productivity.

Keywords: Machine Translation · Cloud Platform · Artificial Intelligence · Human-Computer Translation

1 Introduction

After decades of development, machine translation has completed the evolution and promotion from "mechanical brain" to rule-based machine translation, case-based machine translation, statistics-based machine translation and neural network machine translation based on deep learning [2]. People have actively explored in model research and development, pre-translation processing of original language, post-translation editing which have been widely used. The rapid development of machine translation strongly empowers the trend of human-computer co-translation communication in this era of artificial intelligence.

2 Transition from CAT to NMT

Since the birth of computer in 1947, Machine aided translation (CAT) began to be used in the academic circles. The development of CAT translation method reflects the different stages of automatic translation by computer. The first stage is based on language rules, which holds that human translation can be simulated as long as the theories and methods of human translation are "injected" into the computer, but language rules cannot fully cover the complex and changeable language phenomena. The second stage is an empirical approach to learn from the actual language. Obtaining translation templates from translation examples or statistical model translation (SMT) fall into this category.

K. Subramanian et al. (Eds.): CTMCD 2022, ACSR 99, pp. 436–443, 2023.
https://doi.org/10.2991/978-94-6463-046-6_52

The third stage is neural network translation (NMT), which is an end-to-end holistic translation model [1].

With the deepening of machine translation research and the progress of machine learning and other related disciplines, people gradually find that there are many unavoidable problems in statistical machine translation. For example, the translation process depends on the assumption of hidden structure, the definition of translation features requires manual design, feature engineering is time-consuming and laborious, and often does not have universal significance. For these problems, people have tried a new path - neural machine translation. The so-called neural machine translation is to use neural networks to directly model translation problems. This process does not assume that translation has hidden structure and does not rely on artificially-defined features. The whole translation model can be completed in an end-to-end mode, and translation decoding becomes a process of forward calculation or inference of neural networks.

Although neural networks have been applied in many tasks, substantial progress in machine translation has not achieved until 2013. The main reasons are as follows: 1) there is no very effective framework to deal with the transformation from text sequence to text sequence; 2) The learning of deep neural network is not very effective. [4] The problem of deep network learning has made continuous progress in recent years, and the problem of frame selection in neural machine translation has been alleviated by the "coding and decoding" structure. The so-called "encoding and decoding" structure defines the transformation from sequence to sequence as a two-stage modeling problem. Firstly, the input word sequence $X = x_1\ x_2 \ldots x_n$ containing N words is encoded. The encoding result is a real vector h_n, which represents the information of the whole input sequence up to the N^{th} word; In the second stage, the encoded vector is used for decoding to generate the output sequence $Y = y_1\ y_2 \ldots y_m$, which can be described as follows:

$$\hat{Y} = \text{argmax}_Y \prod_{i=1}^{m} \Pr(y_i | \{y_0, \ldots, y_{i-1}\}, X)$$

$\Pr(y_i|\{y_0,\ldots,y_{i-1}\},X)$ describes the generation probability of the ith word of the target language. Since the encoder has expressed x as hn, the conditional part of $\Pr(y_i|\{y_0,\ldots,y_{i-1}\},X)$ is only related to hn and $\{y_0,\ldots,y_{i-1}\}$. Since $\Pr(y_i|\{y_0,\ldots,y_{i-1}\},X)$ can be calculated by neural network, a network is used to complete the transformation from input sequence to output sequence, and the network can be trained by relatively mature back-propagation method. More importantly, this model completely uses the representation of continuous space when calculating $\Pr(y_i|\{y_0,\ldots,y_{i-1}\},X)$. Compared with the discrete space representation of traditional statistical machine translation, the representation ability of the model is greatly enhanced. Figure 1 shows an example of an encoding and decoding structure.

Since the encoding and decoding framework completely transforms the machine translation problem into a network computing problem from input sequence to output sequence, it does not rely on the characteristics of manual design, so it can better capture the complex correspondence between different languages. It can be said that neural machine translation based on coding and decoding framework has become the standard configuration of relevant research institutions and enterprises. Based on this framework, researchers have also made a lot of improvements and upgrades. Compared with the

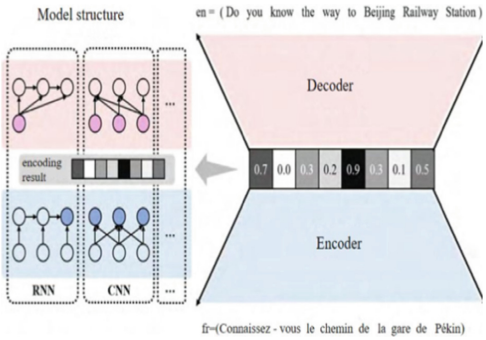

Fig. 1. Encoding and decoding structure of neural machine translation.

level three to five years ago, the quality of machine translation has been improved by leaps and bounds.

3 Application of Human-Computer Co-translation

3.1 Present Situation of Human-Computer Co-translation

With the application of deep learning in translation machine, the translation quality of translation machine based on artificial neural network has been significantly improved. The machine translation finds its way in political, economic, cultural and other communication activities, especially is widely used in the language service industry. According to China Language Service Industry Development Report 2022, with the innovation of artificial intelligence, machine translation is more and more widely used in the industry, and there are 252 enterprises with machine translation and artificial intelligence business. The service mode of "machine translation + Post editing" has been generally recognized by the market. According to the survey, more than 90% of enterprises said that adopting this mode can improve translation efficiency as well as translation quality and reduce translation costs. In the field of low-end text translation, machine translators have gradually replaced manual translators. In the translation and communication activities of other scenes, the participation of machine translators is becoming more and more common, such as using translation machines for cross language communication at symposiums, using machine translation systems for preliminary text translation in translation projects, and using their own translation systems to communicate with foreign friends on Wechat. Machine translator has become an important subject in translation practice and an unavoidable research topic in the study of translation and translation communications [6].

"Human-Computer Co-translation" is an intelligent translation mode based on big data, neural machine translation technology, artificial intelligence and mobile Internet, which combines machine intelligence and artificial intelligence, balances the high efficiency of machine translation and the high quality of human translation, and forms the output chain of "manuscript-machine translation-human translation". Neural network machine translation technology has greatly improved the speed and accuracy of the translation system.

3.2 Cloud Translation Platform Employed at Home and Abroad

In recent years, driven by cloud computing and big data technology, speech recognition, translation technology and translation platform technology have been continuously developed, and translation tools have shifted from local to network [5]. Some language service organizations and localization service companies began to apply cloud computing technology to the translation industry to build a service translation cloud platform that can meet the growing needs of users. Integrating the cloud platform with specific translation technology is the translation technology cloud platform. Its biggest advantage is that it will be distributed in different places. The translation service human resources are integrated to form a large-scale and standardized translation project management, so as to provide comprehensive services for the development of local business and trade and cultural exchanges [3].

Translation technology cloud platforms at home and abroad can be divided into four categories: translation trading platform, translation production platform, translation corpus data platform and artificial intelligence machine translation platform.

3.2.1 Translation Trading Platform

Translation trading platform is an online third-party trading security platform connecting language service talents and customers to ensure the safety and integrity of both parties in the transaction. The client finds the required translator through the translation trading platform and delivers the task to the translator who will complete the translation task with the agreement of both parties.

The most famous translation trading platform in the world is ProZ, which is the largest translation community in the world. It provides translation resources and employment opportunities for translators, translation companies and other personnel in the language industry, and reduces the cost of obtaining resources between language service talents and customers. ProZ also has a TM town platform, which can intelligently match customers with professional translators according to customer needs, translator qualifications, translation subjects and other screening conditions, select the best results, and help customers find translators faster.

3.2.2 Translation Production Platform

Under the traditional translation operation mode, translation companies often encounter problems such as difficult corpus reuse, tight cycle, complex software operation and inconsistent translation style, while translators will also encounter problems such as cumbersome term query, data loss and long typesetting time. Therefore, the online aided translation production platform came into being. The Internet+ language service translation production platform effectively combines online translation, corpus management, team management, project management, collaborative translation and other functions to help translation enterprises and translation teams improve translation efficiency and reduce translation costs.

Famous translation production platforms abroad include Memsource and Matecat. Take Memsource as an example. It is a Czech technology company that provides technical solutions such as cloud translation management system and CAT tools. It has won

international recognition for providing easy-to-use and fully functional translation tools and management system to customers, translation companies and language service personnel. Now it has become a leader in cloud translation technology. Memsource cloud is an online translation management system of MEM source company, with embedded CAT function, which supports project managers to view work progress in real time and multi-person cooperative translation. Famous translation production platforms in China include Tmxmall, Yi-cat, jeemaa.com and Twinslator. Taking Tmxmall, Yi-cat as an example, it is an online translation management platform, which establishes a fully automated, scientific and process-based translation management model for freelancers, translation companies and enterprise translation departments. YiCat embedded CAT tool supports 46 languages, 27 file formats and 27 QA inspection rules. It has the functions of real-time project supervision, team management, synchronization of translating and reviewing and intelligent QA inspection. YiCat integrates hundreds of millions of sentence pairs of corpus data from Tmxmall platform. Users can retrieve the total memory of their public cloud corpus sharing platform in real time during translation, and store them in the personal memory of private cloud corpus management platform, which can help translators save time and improve translation quality.

In the translation stage, human translators can use online and offline computer-aided translation software such as Trados, Wordfast and Twinslator to preprocess documents, that is, match the repeated contents of the original text and the translation memory and carry out pre-translation, so as to reduce the repeated workload and improve the translation efficiency. At the same time, Twinslator, jeemaa.com, zhimaky.utranshub.com and other platforms integrate translation memory and translation management functions, which can realize the complex division of translation, translation reviewing and quality inspection, and greatly improve the efficiency of translation management.

As is shown in Table 1, Human and machine interact and cooperate to complete project, knowledge base construction and machine learning. After obtaining information through image recognition, character recognition and audio and video recognition, the term corpus and online translation complete the basic work of the machine part, including term extraction, term unification, term management and corpus recovery after translation. Online translation completes tasks through automatic text analysis, intelligent settings

Table 1. Human-computer cooperation in translation.

User	Projects and Tasks	Information Tools
Team/individual	terminology management	Terminology database management tool
individual	Memory-based translation	Translation memory management tool
	Spelling check	Spelling Check tool
	Syntax check	Syntax check tool
	Data inquiry	Search engine
Team	Quality monitoring	Translation quality inspection tool
	Schedule management	Translate project management tools

and a variety of machine assistance. The intelligent error correction system checks the low-level errors in translation, such as spelling errors, digital errors and omissions, unit errors and omissions, terminology omissions, sentence and paragraph omissions and punctuation errors and omissions, which saves the time and cost of proofreading.

3.2.3 Corpus Data Platform

With the development of machine translation and the wide application of deep learning technology, more and more neural network structures are introduced into machine translation, such as machine translation based on convolution neural network structure, and machine translation based on cyclic neural network, which improve the quality of machine translation. Machine translation based on neural network usually needs to use a large number of high-quality corpus for training in order to get good translation results.

Corpus refers to a large-scale electronic library with considerable capacity, which is based on the guidance of certain linguistic principles and the method of random sampling to collect the text or speech fragments of continuous language without any processing. At present, corpus has been widely used in language teaching, language research and language engineering.

Machine translation methods are divided into rule-based translation methods and corpus-based translation methods. Corpus based translation methods can be divided into case-based translation and statistics-based translation. The difference between them is that in the former case, corpus will participate in translation as a kind of translation knowledge for the translation subject to query while in the latter one corpus is used to find the sentences that are most likely to become the target language without specific translation practice.

The statistics-based machine translation method starts with language phenomena and obtains the translation through a relatively rational model. The case-based machine translation method obtains the translation from the perspective of machine learning through the analysis and reasoning of cases. However, the two methods are not mutually exclusive. The prospect of corpus-based machine translation is to organically combine the advantages of various methods to further improve the performance of machine translation system.

3.2.4 Machine Translation Platform

Today, with the rapid development of big data + Internet, the application scenarios of machine translation continue to expand. Google translation is a famous foreign artificial intelligence machine translation platform. Google translation is a multilingual translation platform developed by Google, which supports 103 languages.

In other words, users only need to input the content to be translated and select the language pair to generate the target language translation. The interface is friendly and the operation is simple. Corporate giants such as Facebook, Amazon and eBay are also developing machine translation engines. Domestic machine translation platforms include Baidu translation, Youdao translation, Sogou translation and “iFLYTEK voice cloud” platform, while technology giants such as Tencent and Alibaba are also strengthening

the research and development of machine translation products and emerging in the field of machine translation.

4 Prospect of Human-Computer Co-translation

Machine translation has experienced rule-based system and statistics-based system. At present, the system based on deep learning has reached a more mature stage. In 2015, baidu released the neural network translation system, becoming the world's first Internet neural network translation system. Google and Youdao followed suit and successively released neural network translation systems. Compared with the previous translation system, the translation quality is greatly improved. Some experts believe that at present, the translation quality of Google translation system has reached the level of medium-sized manual translators. Machine translation is widely used in various fields, such as business translation, conference and so on. According to the Research Report on the Market Prospect and Investment of China's Translator Industry in 2020 issued by askci Corporation, it is expected that the market scale of translation machines will reach 30 to 40 million in three to five years, and the sales volume of translation machines in China will exceed 100 billion yuan in 2025.

Although machine translation has many advantages such as high speed, wide coverage and high economic benefits, its technology is technical, formal and computational. Although machine translation continues to explore in the interpretation of human brain, cognition and language, it faces a series of problems and challenges because it involves the most advanced and deepest core problems of human intelligence. The current machine translation technology has not possessed the strength of human brain cell computing and ability of cell neural connection and operation. Therefore, although the mainstream machine translation technology has been neural machine translation based on deep learning, it has not revealed the mystery of human brain. Even if the deep learning model can be used to deal with translation problems, machine translation cannot completely replace human translation in a short time, The mode of human-computer co-translation is still the direction of development in the future. The human-computer co-translation platform will gather more machine capacity and the wisdom of manual translators, which will be more closely combined in the operation of actual translation projects. The machine capacity will make more use of the platform to learn manual capacity and output translations close to manual capacity and quality.

5 Conclusions

Human-Computer Co-translation will gather more machine capacity and artificial intelligence, which will be more closely combined in the operation of actual translation projects. With the development of neural network technology, machine translation improves the accuracy of translation, and the role of translator mainly turns to post-translational editing. The translation platform integrating machine translation, computer-aided translation and translation management systems provides a more efficient translation environment for human translation. The translation platform can be integrated, iterated and upgraded continuously, and effectively meet the needs of human personalized translation.

Acknowledgements. This work was financially supported by Nanchang Institute of Technology Teaching Reform Project: An Empirical Study of Blended Interpretation Teaching Mode under the Background of "Internet Plus" Project No. 2020JG043.

References

1. Bundgaard, K. , Christensen, T. P. , 2016, Schjoldager A. "Translator-computer interaction in action - an observational process study of computer-aided translation", The Journal of Specialised Translation
2. Christensen, T. P. , Schjoldager, A .2016 "Computer-aided translation tools - the uptake and use by Danish translation service providers, The Journal of Specialised Translation
3. Excell, D. 2019, Some Challenges in Using Computer-Aided Translation Tools to Facilitate Second Language Fluency in Education. Annals of Emerging Technologies in Computing
4. Xiao, T., Li, Y.Q., Chen, Q., Zhu, J.B., 2018, Machine Translation in the Era of Deep Learning. Artificial Intelligence
5. Yan, X., Chen, R.Z., Zhang, J. 2019, The development Status andTrend of Translation Technology Cloud Platform. Chinese Science & Technology Translator Journal
6. Yin, F.Z., Yu, C.F., Deng, Y.L., 2020, Ten Chapters about Translation Communications, Hunan Normal University Press, Changsha