Analyzing the Current Status and Prospects of the Development of Big Data Translation

Zhao Haiyan*, Wang Jue

University of Science and Technology Liaoning, Anshan, China
2309336084@qq.com
*Corresponding author

Abstract: *This paper discusses the development status of big data translation technology, analyzes the current challenges, and looks forward to its future development trend. With the rapid progress of artificial intelligence technology, big data translation technology has become an important research direction in the field of language translation. This paper firstly reviews the development history of translation technology, then elaborates on the translation technology based on big data, and analyzes its advantages and disadvantages. Finally, the development prospect of big data translation technology is predicted in the light of the current technology trend and market demand.*

Keywords: *Big Data Translation; Language; Artificial Intelligence; Machine Translation; Corpus; Terminology*

1. Analysis of the current status of the development of big data translation

Translation is an important bridge of human civilization and its history can be traced back to ancient times. With the advance of globalization, the importance of translation work has become more and more prominent, and the languages and contents involved have become more and more abundant. Nowadays, with the rapid development of science and technology, the wide application of big data technology has brought about a revolutionary change in the field of translation[1].

1.1 Evolution of translation technology

The evolution of translation technology has gone through the process from human translation to computer-aided translation, and then to machine translation. The early translation work was mainly completed manually, which was inefficient and greatly influenced by the translator's personal level. With the development of computer technology, computer aided translation (CAT) tools have emerged at the historic moment. By providing functions such as term base and memory base, translation efficiency and quality have been greatly improved. However, CAT tools still need human participation and cannot be truly automated[2].

In recent years, with the rapid development of artificial intelligence technology, machine translation technology has made a major breakthrough. In particular, machine translation models based on neural networks, such as Google's Neural Machine Translation (GNMT) system, can automatically learn the mapping relationship between source language and target language through deep learning technology to achieve high-quality automatic translation.

1.2 Translation technology based on big data

The application of big data technology brings unprecedented opportunities for the translation field. By collecting and analyzing a large amount of bilingual control data, big data translation technology can continuously optimize the translation model and improve translation quality and efficiency. At present, translation technologies based on big data mainly include the following types[3-4].

(1) Statistical machine translation: statistical machine translation is one of the earliest translation technologies based on big data. It realizes automatic translation from source language to target language by analyzing the linguistic phenomena such as words, phrases and sentences in a large amount of bilingual control data and constructing statistical models. However, statistical machine translation has limitations in dealing with complex syntactic and semantic relationships.

(2) Rule-based machine translation: rule-based machine translation technology realizes translation by constructing a series of linguistic rules. These rules can be lexical substitution rules, phrase structure rules, grammar rules and so on. Although rule-based machine translation can achieve better translation results in some specific fields and scenarios, it requires a lot of human intervention and linguistic knowledge support[5-6].

(3) Neural machine translation: Neural machine translation is one of the most advanced and mainstream translation technologies. It utilizes deep learning technology to build neural network models, and learns the mapping relationship between the source language and the target language by training a large amount of bilingual control data. Neural machine translation is able to deal with complex syntax and semantic relationships and realize high-quality automatic translation. However, neural machine translation still has some challenges in dealing with long and difficult sentences, specialized terminology and contextual understanding.

1.3 Advantages and disadvantages of big data translation techniques

The advantage of big data translation technology is that it can deal with a large amount of text data and multilingual translation. By collecting and analyzing a large amount of bilingual control data, big data translation technology can continuously optimize the translation model and improve the translation quality and efficiency. In addition, big data translation technology also has the advantages of high automation and fast processing speed[7].

However, big data translation technology also has some disadvantages. First of all, since the training of translation models requires a large amount of bilingual data, there may be insufficient data in some languages and domains. Secondly, big data translation technology still has some challenges in dealing with long and difficult sentences, specialized terminology and contextual understanding. In addition, the results of big data translation techniques may be affected by the quality and quantity of training data, and therefore need to be continuously optimized and adjusted[8].

1.4 Challenges to Big Data Translation

At present, big data translation technology still faces some challenges. First of all, long and difficult sentences and complex grammatical structures are one of the important problems to be solved by big data translation technology. When dealing with these sentences, big data translation technology is prone to problems such as information loss and lexical conversion errors. Secondly, the big data translation technology does not have a deep enough understanding of the context and cannot combine the context well to translate the sentences accurately. In addition, in terms of specialized vocabulary and phrases, big data translation technology is also prone to the problems of word confusion and wrong translation.

In order to solve these problems, future big data translation technology needs to be continuously optimized and improved. For example, the translation quality can be improved by introducing more linguistic knowledge and rules; the ability to handle long and difficult sentences and complex grammatical structures can be improved by constructing more complex neural network models; and the ability to understand the semantics of sentences can be improved by introducing more contextual information[9-10].

2. The main problems facing the development of big data translation

Although big data translation technology has made significant progress, it still faces many challenges in practical application. These problems not only involve language level problems such as vocabulary, syntax and semantics, but also involve many aspects such as culture, context and performance.

2.1 Depth and breadth of linguistic knowledge

The core of big data translation lies in language conversion, which requires the system to be able to deeply understand and utilize linguistic knowledge. However, current big data translation systems often have difficulty in accurately conveying the original meanings of technical terms, idioms, idioms and proverbs when dealing with them. In the face of long and difficult sentences and complex grammatical structures, the system is often difficult to accurately classify sentence components and select appropriate word meanings and lexemes, resulting in translation errors. In addition, for the translation of the whole text, the system often lacks sufficient contextual understanding and contextual analysis ability, which

leads to biased translation results and fails to fully express the meaning and true significance of the original text.

2.2 Language differences and cultural divides

Differences between different languages and cultural barriers bring great challenges to big data translation. Different languages have different ways of expressing time, space, etiquette and social relations, which requires the big data translation system to have the ability of cross-cultural communication. However, in practical application, it is often difficult for the system to fully understand and adapt to the differences between different languages and cultural backgrounds, resulting in misunderstanding or ambiguity in translation results. This problem is particularly prominent in the translation of literary works, advertising and publicity, etc., because these texts often need to accurately convey the author's intentions and emotions, rather than just the literal meaning.

2.3 Data quality and performance bottlenecks

The performance of big data translation systems is directly affected by the quality and quantity of data. For some small languages or languages in specialized fields, the translation quality of the system is often limited due to the lack of large-scale data sets for training. In addition, problems such as errors, ambiguities and biases in the corpus can also affect the accuracy of translation results. These problems make it difficult for big data translation systems to reach the level of human translators in some cases, and even serious translation errors may occur.

2.4 Repeated translations and expressions of rigidity

The big data translation system will in some cases appear the phenomenon of repeated translation. When facing a long and difficult sentence to be translated, if there are word meanings or lexemes in the sentence components that are difficult for the system to recognize accurately, the system may skip these parts, resulting in the repeated translation of the second half of the sentence. This phenomenon not only affects the translation efficiency, but also reduces the translation quality. In addition, the big data translation system often lacks flexibility in expression, especially when translating ancient poems and other literary works, the system is often difficult to accurately convey the poet's emotions and mood, resulting in the translation results appear to be hard and mechanical.

For example, when translating "I love three things in this world. Sun, moon and you. Sun for morning, moon for night, and you forever." Human translation can more accurately convey the artistic conception and feelings of the original text: "There are thousands of people in the world, and I love three things. The sun, the moon and Qing. The sun is the morning, the moon is the evening, and Qing is the morning and evening." Similarly, when translating "If you do not leave me, we will die together." The big data translation system may give the result that "If you do not leave me, we will die together." Human translation can be more poetic: "What is love in the world? It can teach people to live and die together." These examples fully illustrate the limitations of big data translation system in terms of expression.

3. Project content

3.1 Neural network based text translation

The core of the translation system in this project is a well-designed neural network architecture. The neural network not only accepts the input sentences, but also accurately outputs the translation results. In fact, the neural translation system consists of two closely connected neural networks that form a seamless end-to-end interface. The first neural network acts as a linguistically proficient encoder, transforming the input word sequences (i.e., sentences) into meaningful numeric codes. The second neural network acts as a decoder, reassembling these numeric codes into meaningful word sequences in the target language. Through this process, the system learns how to build a bridge between the two human languages to achieve accurate mapping, Fig. 1 Principle of neural network-based text translation.
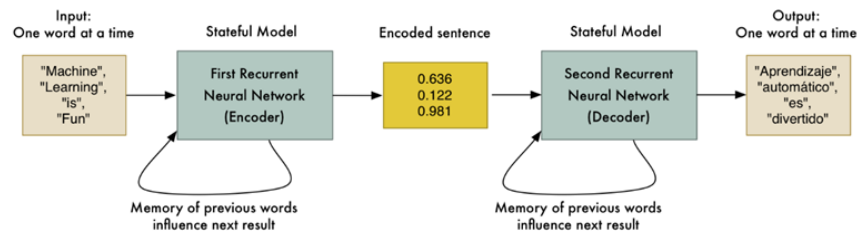
Figure 1: Principle of neural network-based text translation

3.2 The complete translation pipeline

In the translation process, we begin by normalizing the input text to eliminate formatting variations. We make sure that no matter what the user types, the system converts it to a uniform standard format, ensuring that the same sentences are always typed in exactly the same way. This step includes ensuring that words are always capitalized in the same way in the same context, fixing any formatting errors before and after punctuation marks, and cleaning up any unnecessary special characters or quotation marks, Figure 2.
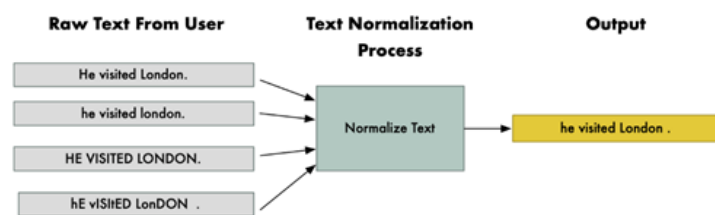

Figure 2: Texts are normalized

The translation process is clear, firstly, the system receives and segments the input text into sentences, then it carries out the normalization process, and finally it denormalizes the translated sentences, reassembles the translated sentences into a coherent text, and outputs the final result. The whole process is like a highly efficient assembly line, which ensures the efficiency and accuracy of the translation, as shown in Fig. 3 Translation process.
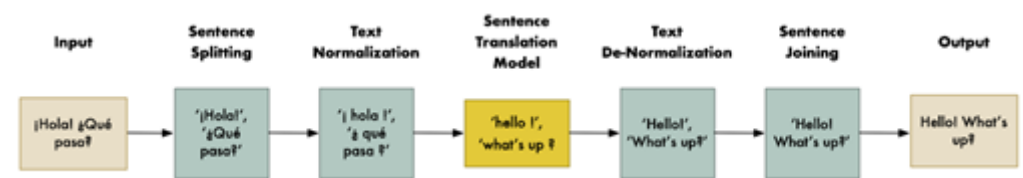

Figure 3: The translation process

3.3 System advantages

Neural network translation technology is an innovator in the field of machine translation in recent years. Compared with traditional translation methods, it utilizes deep learning models to train on massive bilingual corpus, thus realizing more accurate and smooth translation results. In the complex algorithm of hidden layer, the neural network can skillfully connect the words in the sentence, and output the optimal translation result through probabilistic analysis, which makes the semantic accuracy reach an unprecedented high level.

3.4 System functions and features

1) Document translation and text translation: the system supports various forms of text translation, whether it is a whole document or a single sentence, can be easily dealt with. In addition, users can click on the words in the translated text to check their meanings and derivatives, and the system will automatically provide other translations for users to choose. After the translation is completed, users can download the translated text directly from the system.

2) Picture translation: the system has powerful image recognition capability, which can recognize the text in the picture and translate it. At the same time, the system will automatically mark key words and

recommend related knowledge to help users better understand the translated content.

3) Web site translation: the user only needs to copy the Web site into the system, the system can automatically translate it into the target language, so as to facilitate the user to browse foreign language Web sites.

4) Domain-specific translation: the system supports translation needs in various domains, and users can select the corresponding domain for translation according to the text type. The system will automatically identify and mark the proper nouns and difficult words and phrases in specific fields to ensure the accuracy of translation results. In addition, users can also input commands to require the translation to conform to the linguistic characteristics of a certain field.

5) Terminology management: the system supports users to create and manage terminology databases in various fields, and call up specialized terms directly from the terminology databases when translating to ensure the accuracy and consistency of translation.

6) Memory function: Users can store the results of previous translations in the memory for easy subsequent search and reference. This helps to save translation time and improve translation efficiency.

7) Voice input: the system supports voice input function, which allows users to input sentences to be translated in the form of spoken audio, and the system will automatically generate the translated text. This function not only improves the user's operation convenience, but also makes the translation process more intelligent and efficient.

3.5 Prospects for application

1) Promoting the personalization of teaching and learning

This translation system has a wide range of applications in the field of education. Through this system, teachers can understand the problems of students in the process of translation, and provide personalized learning resources and tutoring for students based on these data. Students can also use the system to build up their professional knowledge system and test their translation practice ability, so as to improve their professionalism in translation.

2) Reconstructing the relationship between teaching and learning subjects

The system's information processing, semantic parsing and content generation capabilities make it an important support tool in the field of education. The system can provide strong support for teachers in all aspects of pre-, in- and post-course work, helping students to better understand and master translation knowledge. At the same time, the system can also provide students with translation services and tutorials anytime and anywhere, facilitating effective interaction and communication between teachers and students.

3) Intelligent counseling

Through the voice input function, the system not only improves students' speaking ability, but also realizes the function of intelligent tutoring. The system can provide targeted learning counseling and suggestions according to students' needs, helping students to better master translation knowledge and skills. This intelligent tutoring makes the learning process more efficient and interesting.

3.6 Translation Quality Profile

1) Translation of terms.

Translation of professional terms: This system can accurately translate the meaning of professional terms. For example, "The draft is accepted by the negotiating bank."

Translation of culture loaded words: for words and idioms with specific cultural meanings, this system can accurately grasp their core meanings and properly translate them. For example, "tender" in "Lymph nodes are often enlarged but are not tender." is accurately translated as "tender" in the medical field

2) Prospects for Big Data Translation

With the continuous development and application of artificial intelligence technology, the future of big data translation technology will be more broad. The following are some prospects for the future of big data translation technology.

(1) Multilingual translation capacity will continue to improve: with the advancement of globalization and the increase of international trade, the demand for multilingual translation will continue to increase. Big data translation technology will be able to support more languages and fields and realize more efficient and accurate translation.

(2) Translation quality will continue to improve: with the continuous development and optimization of technology, the translation quality of big data translation technology will continue to improve. The future big data translation technology will be able to better deal with long and difficult sentences, terminology and contextual understanding, and realize more accurate translation.

(3) The degree of intelligence will continue to improve: the future big data translation technology will be more intelligent. For example, the text can be realized by introducing natural language processing technology.

4. Trends and prospects for development

In the wave of artificial intelligence, translation technology is experiencing an unprecedented change, which is not only the deep integration of humanities and technology under digital humanism, but also the new technical requirements of the times and the language service industry for translation practitioners, becoming an indispensable part of the translation ecosystem. With the continuous development of machine translation technology, people can understand the information carried by different languages more conveniently, thus reducing the reliance on traditional human translation to a certain extent.

4.1 Personalized translation

With the help of big data, the future translation system will pay more attention to personalized needs. Through user feedback and customized settings, the system can accurately capture each user's translation preferences, carry out personalized translation output, ensure the accuracy and applicability of translation results, and provide users with more intimate and personalized translation services.

4.2 Cross-cutting applications

Big data translation will no longer be limited to traditional fields, but will be extended to more specialized fields, such as medical and financial fields. At the same time, in order to cope with the specialization and peculiarities of these fields, the system will constantly expand and improve the relevant corpus, including idioms, poems, proverbs, idioms and so on, so as to ensure the accuracy and specialization of the translation results.

4.3 Diversity of translation objects

With the progress of technology, the object of machine translation will also be expanded. From a single word to a whole sentence to a whole paragraph of text, the ability of machine translation will continue to improve. At the same time, the form of translation will be more diversified, from text translation to picture translation, to voice translation, to meet the translation needs of different scenarios.

4.4 Utilizing a parallel corpus of translations

In the process of translation, parallel corpus plays a crucial role. With the continuous improvement of the corpus, the system will add more parameter labels, such as context labels, culture labels, style labels, industry terminology labels, skill strategy labels, etc. These labels will help the system locate the translation content more accurately and improve the uniformity and accuracy of the translation results. These labels will help the system to locate the translation content more accurately and improve the uniformity and accuracy of the translation results. Meanwhile, the more detailed the labels are, the higher the quality of the translation results will be.

5. Conclusion

Driven by globalization and Internetization, machine translation industry has ushered in unprecedented development opportunities. With the support of neural network and big data and other advanced technologies, machine translation can show strong strength in both web and mobile terminals,

while expanding from text translation to speech and image translation, widely used in various fields, and providing a more convenient tool for human communication. Looking into the future, the technical development of big data translation will continue to move in a deeper, bigger and better direction, bringing more convenience and possibilities to our life.

Acknowledgements

References

[1] Liu Junpeng; Song Dingxin; Zhang Yiming; Huang Degen. Neural machine translation system based on fusion of multiple data generalization strategies [J]. Journal of Jiangxi Normal University (Natural Science Edition), 2020 (01)

[2] Liu Qingfeng; Liu Chenxuan; Wang Yanan; Zhang Weitai; Liu Junhua. Domain Personalized Machine Translation Method Integrating External Dictionary Knowledge in Conference Scenarios [J]. Chinese Information Journal, 2019 (10)

[3] Zhang Fawang; Shao Wenting. The Application of E-Commerce English Translation in the Multilingual System of Picking Robot -- Based on ESP Theory [J]. Research on Agricultural Mechanization, 2020 (02)

[4] Zhang Shenggang; Aishan Wumai Er; Turgen Ibrahim; Maihemuti · Maimaiti; Mirshati Litip. Implementation of Uyghur Chinese translation system based on neural network [J]. Modern Electronic Technology, 2018 (24)

[5] Wang Yajuan; Li Xiao; Yang Yating; Mi Chenggang. Research on Uyghur Chinese Machine Translation System Fusion Based on Interpretation Information [J]. Computer Engineering, 2019 (04)

[6] Yu Qian. Design of interactive English Chinese translation system based on feature extraction algorithm [J]. Modern Electronic Technology, 2018 (04)

[7] Xie Qianqian; Aishan Wumai Er; Turgen Ibrahim; Maihemuti · Maimaiti; Rexiti in Kabar River Abby. Design and implementation of a hybrid strategy Chinese Uyghur assisted translation system [J]. Modern Electronic Technology, 2017 (20)

[8] Weng Shenghui; Chen Weixuan; Chen Kuanglin. Design and Implementation of Sign Language Translator Based on Leap Motion [J]. Modern Electronic Technology, 2015 (24)

[9] Li Zhanhui; Meng Jianyi; Chen Zhijian; Yan Xiaolang. Dynamic translation algorithm based on the characteristics of branch instructions [J]. Journal of Shanghai Jiaotong University, 2015 (02)

[10] Wang Ronghua; Meng Jianyi; Chen Zhijian; Yan Xiaolang. High speed address translation method based on access region characteristics [J]. Journal of Zhejiang University (Engineering Edition), 2014 (02)