The Application of Big Data Analysis Technology in English Translation

Yuan Chang-pu

Department of Foreign Languages, Xi'an Jiaotong University City College, Xi'an, 710018, China

Abstract: Today, the times are developing, and science and technology are also advancing. English translation has also undergone tremendous changes. From the initial manual translation to the current intelligent translation, the speed is not as slow as before. Now, whether it is translating a word, a sentence or even an article, it can be translated in less than a minute. At the same time, fast translation is accompanied by a huge problem, which guarantees speed but cannot guarantee quality. Therefore, with the development of big data analysis technology, it is necessary to make up for the quality problems in the translation process, so that its application in translation can give full play to its actual value. The purpose of this article is to study the application of big data analysis technology in English translation. This article takes English translation as the main research object, and uses big data analysis technology to solve the problem of wrong translation in the process of English translation. In the intelligent English translation process, the original text must be analyzed first, and the expression sequence must be distinguished. Based on the language characteristics, appropriate translation strategies are adopted to effectively recombine the involved time, space and logic. This article uses a questionnaire survey method to understand translation errors in various English translation platforms. The experimental research methods used in this study, one of which is to analyze the translation errors in the sample and the types of errors; the other is to deal with the number of translation errors and the proportion of each type of error. The experimental results show that in text translation, the word order of the third paragraph has the highest accuracy rate, which is 85%, and the error rate is the context of the second paragraph, which is as high as 40%. In general, even Google Translate, which has the highest translation accuracy, has certain problems in terms of word order, grammar, and context. The application of big data analysis technology in English translation still needs to be improved. This article summarizes some countermeasures for better and more convenient application of big data analysis technology in English translation to improve the accuracy of English translation and bring convenience to the society.

1. Introduction

Today, the times are developing, and science and technology have been advancing [1]. English translation has also undergone tremendous changes. From the initial manual translation to the current intelligent translation, the speed is not as slow as before [2-3]. Now, whether it is translating a word, a sentence or even a piece of article, it can be translated in less than a minute [4]. At the same time, fast translation is accompanied by a huge problem, which guarantees speed but cannot guarantee quality [5]. Therefore, with the development of big data analysis technology, it is necessary to make up for the quality problems in the translation process, so that it can be used in translation to maximize its practical value [6-7].

In the research on the application of big data analysis technology in English translation, many scholars at home and abroad have conducted research on it, and achieved good results. Lin proposed a story grammar theory, thinking that a story can have its grammatical structure and sentence just like a sentence[8]. It consists of a series of grammatical rules with semantic rules, which is suitable for dealing with more formal story descriptions in style [9]. Miyazaki proposed the central theory, which divided the text structure into three types: language structure, intention structure and attention structure. At the level of language structure, the text is divided into language segments, and there is a nested relationship between language segments [10].

This article takes English translation as the main research object, and uses big data analysis

technology to solve the problem of wrong translation in the process of English translation. This article uses a questionnaire survey method to understand translation errors in various English translation platforms. The research methods used in this study are mainly qualitative research and quantitative research. Among them, qualitative research is used to analyze the translation errors in the sample and the types of errors; quantitative research is mainly used to deal with the number of translation errors and the proportion of each type of error. Through induction, some countermeasures for better and more convenient application of big data analysis technology in English translation are summarized. Improve the accuracy of English translation and bring convenience to the society.

2. Research on the Application of Big Data Analysis Technology in English Translation

2.1 Limitations of Smart English Translation

(1) Smart word translation

The initial machine translation is to convert the words of the source language sentence into the words of the target language, perform a certain arrangement, and form the final translation. This kind of translation does not perform sentence analysis, nor does it use the overall information of the sentence.

(2) Smart sentence translation

Sentence level translation uses grammatical analysis, semantic analysis and other technologies to use sentence structure information and context-related information in the sentence to deepen the depth of analysis, thereby greatly improving the accuracy of machine translation and the readability of the translation.

(3) Chapter intelligence translation

Language is a tool used by people to communicate, and only one chapter can complete the transmission of information independently. The most important thing in text translation is to construct a translation environment, save the global information of the text, and interact with this environment when translating sentences. People should have realized the importance of context a long time ago. It does not have a precise definition, it is mostly expressed in terms, and its role is to help the translation system determine the unique structure and meaning of the sentence.

2.2 Countermeasures Given by Big Data Analysis Technology

(1) Context-sensitive processing

Context correlation is one of the most difficult problems in natural language processing. It is designed based on uncertainty analysis. The execution process uses technologies such as controllable hierarchical compatible unification matching, forward prediction and backward feedback, and failure backtracking. When executing, it uses the rule header and the current form of the input to perform hierarchical matching. After the matching is successful, the relevant condition search function is adjusted to detect the context-related conditions. If the conditions are met, it will be reduced, otherwise it will be backtracked. Through big data analysis technology, context-sensitive operation definitions are embedded into specific rules and terms, so that the judgment of context-sensitive processing conditions is limited to the situations that may correspond to the current mode corresponding to it. In this way, the ambiguity of operation is reduced, and the global problem is localized, which not only improves the ability to distinguish ambiguity, but also simplifies complex context-related processing.

(2) Sentence analysis

In the dictionary of the intelligent English translation system, only some of the most commonly used words are provided with surface entry, and most of the words are accessed through the basic type. The form of the lexical analysis rule is consistent with the form of the SC grammar rule, and there is also a context-sensitive test function to initially deal with the relevance and improve the accuracy of the lexical analysis. The idea of sentence analysis algorithm is to find the phrase rules of all words appearing in the sentence from the dictionary, and use these phrase rules to form a

possible phrase list; then take out adjacent phrases from the phrase list to form a predicted path table, and analyze the formed parts The result is matched with the prediction rule. If the match is successful, operations such as semantic check and related detection below will be performed; if the match is unsuccessful or the test fails, it will be backtracked and the next rule or the next prediction path will be called. Repeat this way until a suitable description of the overall structure of the input sentence is formed.

(3) Text language processing technology

Text language processing refers to language processing in units of texts. It includes automatic abstract generation, story understanding, and intelligent translation technology based on texts. The technical difference between it and sentence-based language processing is that in the text processing system, the main consideration is the connection between the central idea of the text and the sentence, while the main task in the sentence processing system is to correctly analyze the structure of the sentence. It can be seen that the two are different but complementary to each other. The focus of text research is to analyze and extract the information of the article as a whole, which involves the structure of the article, the thought of the article, the division of meaning segments, the processing of language repetition, and the use of world knowledge for reasoning. Since text language machine processing involves the presentation, acquisition and utilization of abstract information such as central ideas, and the current technical level cannot handle this knowledge well, the research on text processing is still in its infancy.

2.3 Application of Big Data Analysis Technology in English Translation

(1) Story grammar

Story comprehension is an aspect of the application of natural language understanding. In story comprehension, in addition to sentence structure and semantics, the overall structure of the story should also be considered. Story grammar breaks down the story into different stages, and then refines the stages into actions, events, etc. Story grammar can establish the relationship between the various parts of the article, promote the program to deal with the omitted content, and realize some reasoning functions. This method focuses on the overall structure of the story description, and can be achieved using a top-down analysis method. The grammar of a story is context-free and cannot describe the interconnected plots in the story.

(2) Automatic abstract

Automatic summarization is to use a machine to synthesize the main content of the article or the content that the user is interested in, and output its summary in the same or different language as the original text. In language processing, there are generally the following steps: grammatical analysis, semantic analysis, pragmatic analysis and information extraction, and text generation. Because automatic summarization extracts the center from a chapter, it is more appropriate to use a plan or script-based format. But the disadvantage of this method is that the system is powerless for events that are not defined in advance, that is, they can only extract a summary of the news that they have an event frame.

(3) Central theory

In the language segment, use the center to emphasize the local consistency of the language and track the focus of the sentence. The definition of center is the entity in a sentence that connects this sentence with other sentences in the language segment in which it is located. The entity that appears in the sentence has a higher centrality than other entities. Each sentence in the language segment has a front-looking center set, except for the initial sentence of the language segment, each sentence has a back-looking center set. The center of the back look is the focus of the sentence, which is the link with other sentences in the language segment. The forward-looking centers are sorted by priority, and their positions in the sentence, such as subject position, object position, etc., determine their priority in the forward-looking centers.

Central theory is a promising technology to solve text problems. However, because it involves difficult problems such as text structure, language segment division, and central ideas, although there are many current researches, no breakthrough has been made.

2.4 Big Data Analysis Algorithm

(1) Cloud computing performance modeling

Assuming that the cloud computing center is composed of multiple heterogeneous components, the maximum number of heterogeneous components can be created on the i-th PM. Therefore, the total number of VMs in the cloud computing center can be expressed as:

$$\text{vm}_{\text{total}} = \sum_{i=1}^{M} m_i \quad (1)$$

(2) Time interval queuing system calculation

G/M/n queuing system. In this queuing system, the arrival of customers obeys the general distribution, and the subtasks (customers) obey the independent arrival mode of order quantity. Therefore, the time interval between two successively arriving subtasks is an independent and identically distributed random variable, which obeys the general distribution A. We use A(t) to represent the cumulative distribution function of the time interval, and a(t) to represent its probability density function. The Laplace transform of the time interval can be expressed as:

$$A^*(s) = \int_0^{\infty} e^{-st} a(t) dt \quad (2)$$

3. Experimental Research on the Application of Big Data Analysis Technology in English Translation

3.1 Experimental Subjects and Methods

Taking other intelligent English translation platforms as the research object, investigating their translation results under big data analysis technology, collecting their shortcomings in practical applications through questionnaire surveys, and providing corresponding countermeasures.

3.2 Data Collection

In the era of big data, more storage devices and faster I/O mechanisms are needed to access data. In order to achieve more convenient and rapid access to data in big data applications, academia and industry continue to propose new technologies and methods to shorten the gap between high-performance CPU and low-speed I/O. Secondly, although solid-state drives achieve efficient random I/O and alleviate storage difficulties, they cannot achieve efficient sequential I/O at the same time. High-efficiency and low-cost storage devices are still in the continuous development process.

4. Experimental Research and Analysis of the Application of Big Data Analysis Technology in English Translation

4.1 Application of Big Data Analysis Technology in English Translation Sentences

Take a sentence as the experimental object. This sentence is composed of subject, predicate, object, adverbial, etc., and use an English translation platform to translate it to study the most error-prone part. The experimental results are shown in Table 1:

Table 1 Translation Accuracy Rate of a Certain Platform

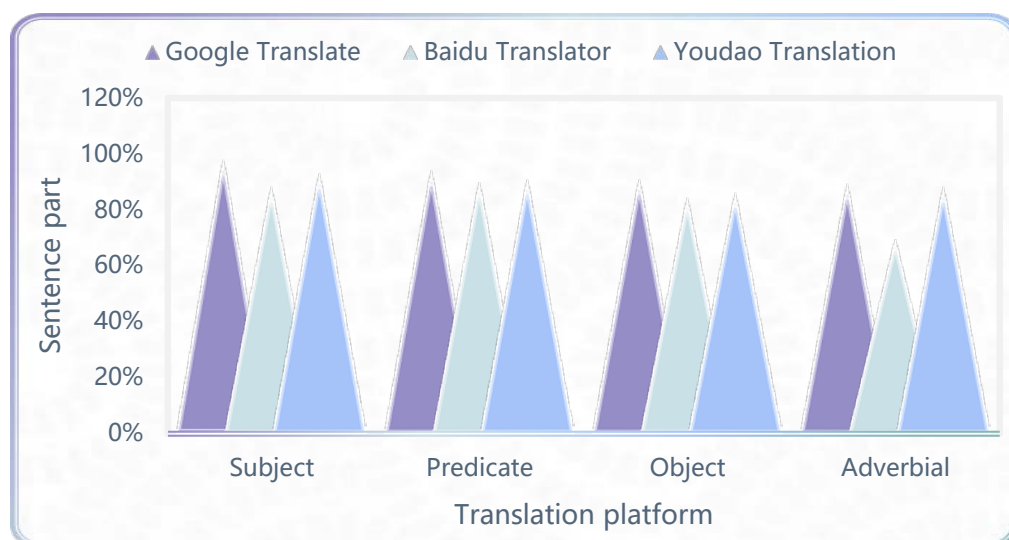|	Subject	Predicate	Object	Adverbial
Google Translate	99%	95%	92%	90%
Baidu Translator	89%	91%	85%	70%
Youdao Translation	94%	92%	87%	89%

Fig.1 Translation Accuracy Rate of a Certain Platform

As shown in Figure 1, the highest translation error rate is the adverbial part of Baidu's translation, which is 85%. Overall, Google Translate has the highest translation accuracy rate, with an average accuracy rate of 94%, which is higher than the accuracy rate of any part of other platforms. It can be seen that big data analysis technology is the best in the practical application of Google Translate.

4.2 Application of Big Data Analysis Technology in English Translation Chapter

Taking a certain article as the research object, this article is divided into three paragraphs, and it is translated using Google English translation platform to study its accuracy in word order, grammar, and context. The experimental results are shown in Table 2:

Table 2 Google Translated Article Accuracy Rate

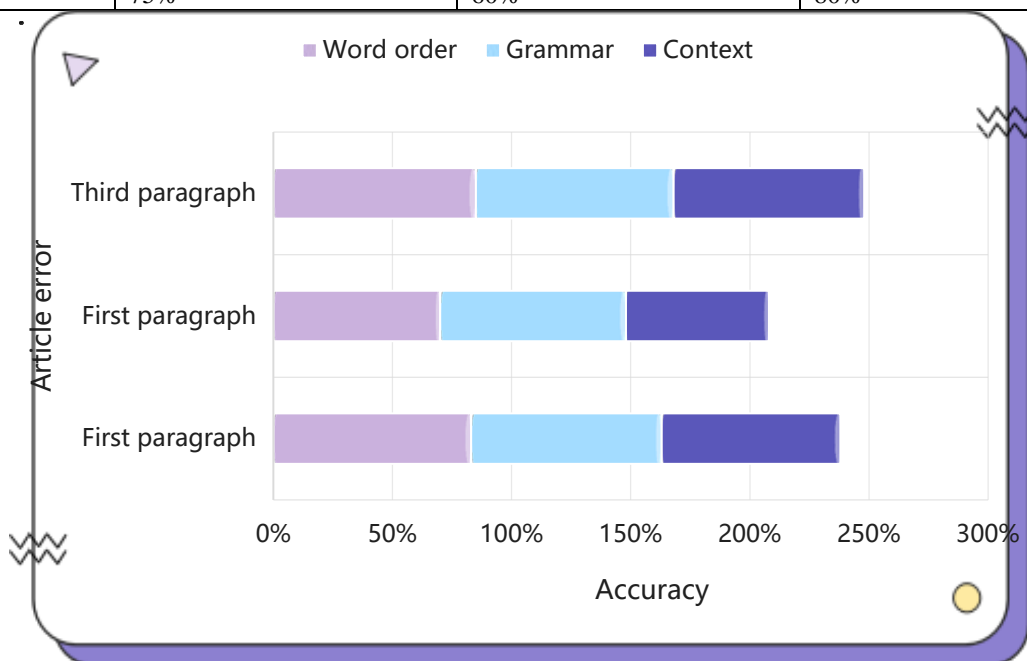|	First paragraph	First paragraph	Third paragraph
Word order	83%	70%	85%
Grammar	80%	78%	83%
Context	75%	60%	80%

Fig.2 Google Translated Article Accuracy Rate

As shown in Figure 2, the word order of the third paragraph has the highest accuracy rate, which

is 85%, and the context of the second paragraph has the highest error rate, which is as high as 40%. Generally speaking, even Google Translate, which has the highest translation accuracy, has certain problems in terms of word order, grammar, and context. The application of big data analysis technology in English translation still needs to be improved.

5. Conclusion

This article is based on big data analysis technology, and realizes a context analysis and processing system in practical application. With this system, some problems that cannot be solved by intelligent English translation systems can be dealt with. At the same time, the discovery of context is also a research field with many questions inconclusive. New theories and technologies will continue to appear, taking into account the development of existing technologies, but also in order to adapt to the development of future technologies. All in all, language processing in units of chapters is the development direction and ultimate destination of natural language processing. Text language research is becoming a research hotspot. The construction and use of context is a wide-ranging and difficult subject. There is a lot of research work to be done. Its development depends on the progress of basic subject research in big data analysis technology.

References

[1] Sciascia S, Radin M . AB1110What can google trends can tell us about a disease? big data trends analysis in systemic lupus erythematosus[J]. Annals of the Rheumatic Diseases, 2017, 76(Suppl 2):1443-.

[2] Beckmann J S, Lew D . Reconciling evidence-based medicine and precision medicine in the era of big data: challenges and opportunities[J]. Genome Medicine, 2016, 8(1):134.

[3] Béranger. [Chapter 1. From the study of risks to the translation of the ethical issues of Big Data in Health].[J]. Journal international de bioethique et d'ethique des sciences, 2017, 28(3):15-25.

[4] Peris A, Domingo M , Casacuberta F . Interactive neural machine translation[J]. Computer Speech & Language, 2016, 45(Sep.):201-220.

[5] Castilho S, Moorkens J , Gaspari F , et al. Is Neural Machine Translation the New State of the Art?[J]. The Prague Bulletin of Mathematical Linguistics, 2017, 108(108):109-120.

[6] Choi H, Cho K , Bengio Y . Context-Dependent Word Representation for Neural Machine Translation[J]. Computer Speech & Language, 2017, 45(Sep.):149-160.

[7] Fan G, Shen J . Application of Big Data Computing Model Based on Semantic Data Analysis Technology[J]. IOP Conference Series: Materials Science and Engineering, 2020, 750(1):012071 (5pp).

[8] Lin H Y, Yang S Y . A Smart Cloud-Based Energy Data Mining Agent Using Big Data Analysis Technology[J]. Smart Science, 2019, 7(JUN.):1-9.

[9] Katsuda T, Isshiki H . Contribution of medical big data analysis technology to medical innovations[J]. Fujitsu Scientific & Technical Journal, 2018, 54(2):59-65.

[10] Miyazaki T. Social Big-Data Analysis Technology to Support Program Production[J]. Broadcast Technology, 2017(68):19-19.