Research on the Application of Computer Big Data Technology in English Online Translation

Feiyu Shen *

Shaoyang University, Shaoyang 422000, China

* Corresponding Author Email: 3073683253@qq.com

Abstract. Aiming at the disadvantages of the traditional rule-based machine translation model that the English translation results are not accurate enough and it is difficult to accurately describe the relationship between words, the English machine translation model based on the semantic network is designed and improved. The algorithm analyses the English grammatical rules, then performs Gaussian marginalization on the semantics to obtain the rectangular window function, obtains the window feature vector, projects the semantic information entropy data, and adds the semantic correlation factors to the information entropy and information gain of the text., So as to obtain the semantic nonlinear spectral characteristics. Experimental results show that the designed English machine translation model has high translation accuracy and stability.

Keywords: Intelligent recognition technology; English translation; machine translation model; computer; big data.

1. Introduction

Due to the rapid development of globalization and the rapid flow of information between different countries, English has become the main language of international communication. At present, the application value of intelligent recognition technology in different fields is constantly improving. The English machine translation model based on modern intelligent recognition technology can improve the efficiency and accuracy of English machine translation and realize barrier-free communication [1]. However, the traditional English machine translation method based on syntactic analysis cannot solve the problem of partial structural ambiguity in the massive English language in intelligent recognition technology, and there is a problem of low accuracy of machine translation. Therefore, this paper designs an English machine translation model based on modern intelligent recognition technology, through the direct maximum entropy model, to obtain the best combination of different features in complex English sentences, eliminate some structural ambiguities, and improve the accuracy of English machine translation.

2. System Design

2.1. Design of the overall operation process of the system

The design of this part is mainly the specific process of the overall operation of the system, so as to ensure the stable operation of the system. In this regard, the system will first start the online translation system to determine whether the software is running for the first time. If it is the first time, it will directly load the thesaurus and put it into the specified folder. If not, then enter the main operating interface of the system and display four main function menus: word translation, word book, word query and management [2]. After selecting the specific menu, enter the relevant interface again. The specific process design is shown in Figure 1 (picture quoted from English-Mizo Machine Translation using Neural and Statistical Approaches).

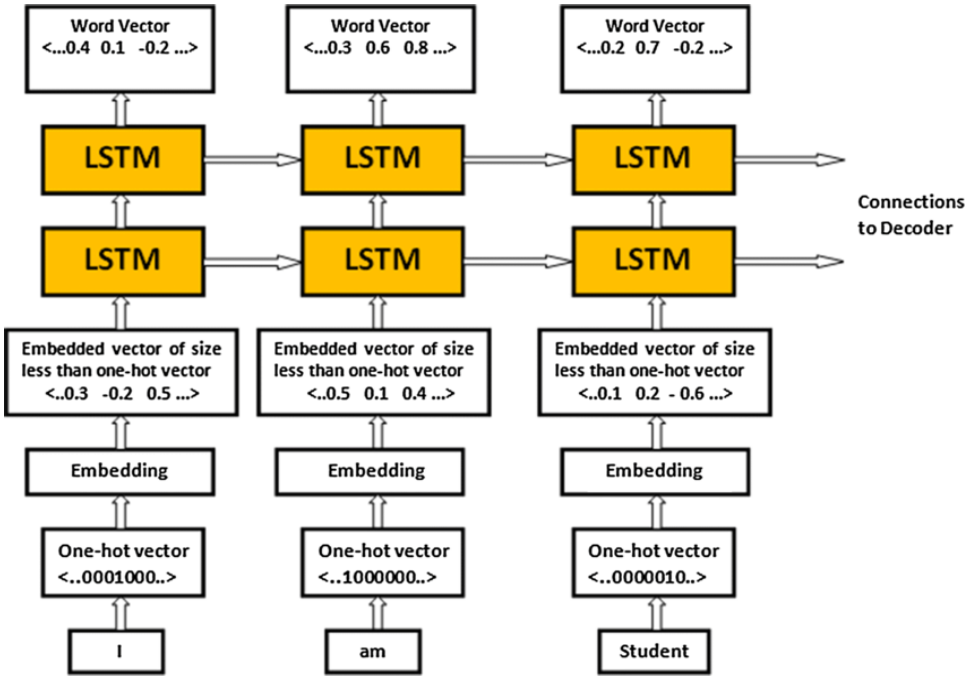


Figure 1. The overall operating process of the system.

2.2. System network topology design

The C/S structure is used to realize the design of the software system, that is, the user can use the Android collection client to realize the connection of the network interface, and then the word is translated online through the database server. In order to reduce the load of the server during operation, the translation program is installed on the query network server, so that the processing power of the server can be fully utilized. And the user can input the word information to be queried on the client of the application. After inputting, the client can realize the initial processing of the data. After the data exchange with the server, the server can realize the translation work according to the request. The system network topology is designed as shown in Figure 2 [3].

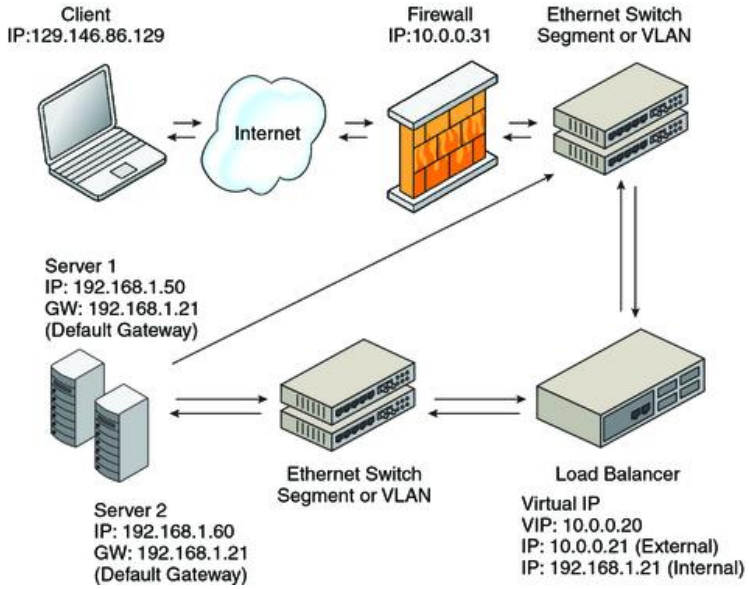


Figure 2. System network topology design.

2.3. Server-side design of English online translation system

The server side of English online translation platform mainly includes load balancing server, OCR processing server and translation processing server. The server side belongs to the core content of the

English online translation platform. Its main purpose is to realize the receiving and sending of image information, and the text information is extracted through the OCR engine, the image is recognized, etc., and then the application server is called to realize the text information. online translation [4]. The main purpose of the load balancing server is to receive the request sent by the client, realize the processing of the OCR node according to the corresponding principle, use the recognition image information, and extract the text information, so as to send the text information to the translation processing server. After the translation processing server receives the request, the application service is called, and then the online translation result feedback is realized on the client side.

When a large number of requests are used to implement server deployment using the traditional C/S architecture model, user access will be congested due to insufficient resource allocation, thereby improving the response time of system data, or because of the large amount of data access at the same time, the server will appear Paralyzed and crashed, users cannot get the response they want. When there are few user requests, the server will still be vacant for a long time, and the low efficiency of the server will waste a lot of resources. In order to avoid the above problems, it is possible to deploy back-end servers in the cloud platform to achieve efficient expansion, flexible services and dynamic control. For OCR processing servers, dynamic deployment and expansion can be achieved when user visits are large [5].

3. Principles of English Machine Translation

The phrase set is used for translation selection and combination, and constitutes the output sentence of the target language, and completes the English machine translation. The specific steps are described in detail as follows: Suppose that T represents the source sentence set, T_i represents the source sentence after segmentation, and j represents the translatable type of the source phrase sequence, then use formula (1) to obtain the source sentence phrase and the corresponding English phrase link

$$D_{ij} = \frac{R_*}{D_n} \bullet \left(T \times T_i\right)_j \tag{1}$$

In the formula, R_* represents the given Chinese-English grammar conversion rules, and D_n represents the semantic block.

$$S_{ww} = \frac{w_i \partial(t)}{p(o)} * SC(o) \tag{2}$$

In the formula, w_i represents the modifier of the word, and $SC(o)$ represents the number of semantic blocks constructed according to the grammar library.

$$WE_{\gamma} = \frac{\Phi''' \bullet H(F) S_{ww} \vartheta_*}{\gamma^* D_{ij}} \tag{3}$$

In the above formula, ϑ_* represents the redundancy of knowledge.

4. System test and simulation proof

4.1. Experiment 1

The experimental corpus used the 586,538 sentence-pair Chinese-English parallel corpus provided by Zhongjun Company, from which 1,000 sentence pairs were randomly collected as test corpus, 2000 sentence pairs were used as development corpus, and the rest were used as training corpus. The experiment divides the test corpus into three test sets of simple sentences, general sentences and complex sentences based on sentence length, as shown in Table 1.

Table 1. Test set classification results.

Sentence types	Word count	Number of sentences
Simple sentence	N<8	458
General sentence	8≤N<18	334
Complex sentence	N≥18	175

Using the method of this paper and the traditional syntactic analysis machine translation method, the BLEU value is calculated for the test set in Table 1, and the results are shown in Table 2. Among them, the BLUE value is an automatic evaluation method of machine translation. The higher the value, the better the quality of machine translation. Analyzing Table 2 shows that the BLEU value of the method in this paper is basically the same as that of the traditional syntactic analysis machine translation method for machine translation of simple sentences, while the BLEU value of this method is slightly higher than that of the syntactic analysis machine translation method for machine translation of general sentences, especially in the case of machine translation [6]. When machine translating complex sentences, the BLEU of the method in this paper is much higher than that of the syntactic analysis machine translation method. Therefore, it shows that the advantage of this method is machine translation of complex sentences. There are a large number of highest noun phrases in complex sentences. This method adopts statistical machine translation method based on the idea of maximum entropy, which can obtain the best combination of different English language features in complex sentences. Ways to eliminate some structural ambiguities and improve the accuracy of English machine translation.

Table 2. Comparison of machine translation performance based on test set classification.

	Simple sentence	General sentence
Machine translation method	33.58	32.85
Traditional syntactic analysis machine translation method text method	33.62	33.64

4.2. Experiment 2

The experiment analyzes the probability distribution of machine translation methods and actual machine translation results through two circular description syntax. The darker the color in the circle, the higher the probability of accurate machine translation results in the search range, and then the decoding performance of different machine translation methods can be analyzed. The probability results obtained by different methods are mapped to the actual machine translation results to realize the comparison of the retrieval range of different methods. The traditional syntactic analysis machine translation method, tree-based translation method and forest-based translation method are compared with the method proposed in this article. The word error rate and segmentation error rate of the translation results are shown in Figure 3. The final probability retrieved based on the traditional syntactic analysis machine translation method is in the lower left corner, while the actual machine translation result probability is in the upper right corner, and this method cannot be mapped to the actual result, which proves that the translation accuracy of the traditional method is poor.

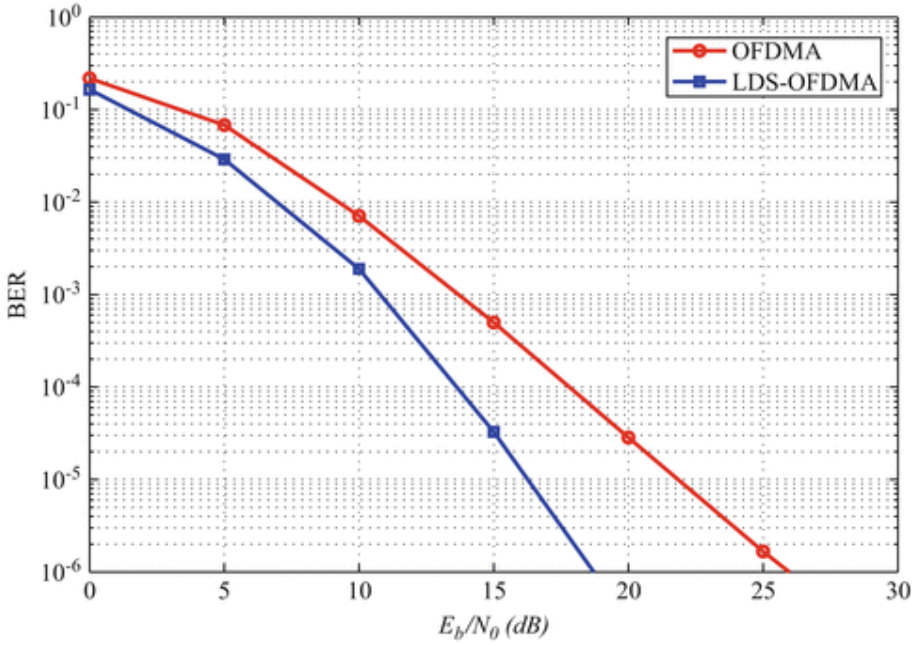


Figure 3. Word error rate.

It can be seen that the accuracy of using this algorithm to optimize English machine translation is higher than that of the literature algorithm. This is mainly because in the process of using this algorithm to optimize English machine translation, the English grammatical rules are analysed first, and the semantics are obtained [7]. The Gaussian marginalized rectangular window function projects a large amount of English semantic entropy data, which effectively guarantees the accuracy of the algorithm in this paper for English machine translation optimization.

5. Conclusion

The system implementation of the English online translation platform on the mobile platform mainly includes the realization of the pronunciation module and the realization of the translation module. Nowadays, the mobile devices basically have a special speech synthesis function configuration. For example, the TTS function in the Android system is to realize English. Function configuration of word pronunciation. The development and design of the word translation module is mainly realized by the IBM model and the EM algorithm. The process is mainly to create the word translation model, and then perform the translation process in different languages through the algorithm. In addition, if the English online translation platform is to be used in the professional translation field, its system implementation should also include a basic data maintenance module, a project data creation module, and a voice corpus management module.

References

[1] Xu Yuanyuan. Design and Implementation of English Translation System for College Webpages Based on the Horizon of Communication Studies. Automation Technology and Application, 299(05) (2020)178-181.

[2] Cui Dan. Design of Computer Intelligent Proofreading System for English Translation. Modern Electronic Technology, 42(4) (2019) 5-15.

[3] Sun Rui. Research on the Development of Computer Intelligent Proofreading System from the Perspective of English Translation Application. Microcomputer Application, 322(2) (2020) 149-152.

[4] Guo Lei. Design of an English Translation Computer Intelligent Scoring System Based on Natural Language Processing. Modern Electronic Technology, 2019, 42(4) (2019) 4-17.

[5] Gao Nan, Chen Wei. Design of a memory-assisted long-character English automatic translation system based on the B/S framework. Automation Technology and Application, 2020, 302(8) 40-42+59.

[6] Shi Jie. Research on the development and application of speech recognition system based on English translation. Modern Scientific Instruments, 2020, 1(3) (2020) 6-15.

[7] Zhang Xiaojuan. Design of an automatic scoring system for college students' English translation based on the idea of N-Gram co-occurrence statistics. Automation Technology and Application, 39(8) (2020)4-11.