Advances in Computer, Signals and Systems (2023)
Clausius Scientific Press, Canada

DOI: 10.23977/acss.2023.070906
ISSN 2371-8838 Vol. 7 Num. 9

Computer Intelligent Proofreading System of Translation Model Based on Improved GLR Algorithm

Huanhuan Gu*

Nanjing Sinovatio Technology Co., Ltd., Nanjing, Jiangsu, 210012, China
690554446@qq.com
**Corresponding author*

Keywords: Improved GLR Algorithm, Translation Model, Computer Intelligence, Proofreading System

Abstract: The intelligentization of translation computers refers to the use of modern computer science technology, network information technology and information processing theory to analyze and recognize massive texts and apply them to the translation process. This article intends to use the improved GLR algorithm to study the computerized intelligent proofreading system of translation models, and its purpose is to improve the translation accuracy of the computer-aided system. This article mainly uses experimental and comparative methods to test and study the computerized intelligent proofreading system for the translation model of the improved GLR algorithm. Experimental results show that the improved GLR algorithm machine translation's recognition accuracy rate can reach 95%. For this reason, the computer intelligent proofreading system can use the improved GLR algorithm to improve the accuracy of the system.

1. Introduction

Translation is a complex subject with strong practicality. With the rapid development of digital technology, as a data processing tool, computers have been used in scientific research and teaching in more and more fields. The main purpose of translation is to express that the author wants to express what he wants to achieve, while the translation is to hope that readers can understand the original content. Therefore, it needs to be intelligently calibrated through a computer.

There are many theoretical results in the research of the computerized intelligent proofreading system for translation models based on the improved GLR algorithm. For example, it is suggested to create a combination of semantic ontology model and sentence translation algorithm to create semantic ontology model degree [1-2]. Someone has developed an intelligent sentence recognition algorithm based on improved GLR algorithm [3-4]. In addition, it is also said that the English translation proofreading system based on sentences and grammar emphasizes sentences and calibrates syntactic accuracy, which does not solve the problem of poor contextual consistency in the language [5-6]. Therefore, in order to improve the coherence and accuracy of translation, this article needs to design with the aid of a translation model that improves the GLR algorithm.

This article first studies the relevant theoretical knowledge of computer-assisted translation. Secondly, the related description and analysis of the hierarchical phrase model fused with linguistic information. Then analyze the principles of the design of the intelligent proofreading system. Finally, an example is conducted through experiments, four algorithms are tested, and conclusions are drawn.

2. Computerized Intelligent Proofreading System for Translation Models based on Improved GLR Algorithm

2.1 Computer Aided Translation

With the development of translation, many computer-aided technologies have emerged, including text, pictures, and audio information. The data processing method used in traditional decoders is based on text as the core for judgment and classification. The language proofreading system designed based on the intelligent method also has its own outstanding characteristics: strong real-time interactivity and good portability [7-8].

For computer-assisted translation in the translation process, its main function is to realize the identification of translation content, semantics and related information. From a technical point of view, the traditional decoding methods based on text and sound language recognition have certain limitations. Three modes are mainly used in the translation process: manual verification, machine calibration and machine decoding. When manually inputting data, it is necessary to use network transmission equipment to transmit text information to the neural network for reading and manually convert it into recognizable language signals. Due to its own limitations and human factors, some uncertain information may affect the accuracy of the translation results, and even lead to a higher translation error rate. In the translation process, we need to perform basic operations on the computer, including text, pictures and other auxiliary information. These functions are to enable users to quickly and accurately find the required text when processing files [9-10].

Computer-aided translation is an automatic translation method based on text. It uses text as an information carrier to edit, modify and calibrate the translation. Its main tasks include: using the database to retrieve relevant data, obtaining the required parameters according to keywords, and generating the corresponding sentence instruction single-stream file format.

2.2 Hierarchical Phrase Model Incorporating Linguistic Information

In recent years, many scientists have applied statistical methods to machine translation, which has significantly improved the quality of translation. The sentence-based translation model has great advantages in processing local ordering, translation of fixed sentence expressions, and some insertion and deletion in response to local context. Sentence templates with the above capabilities provide a simple and efficient machine translation method. Compared with the word-based model, it expands the granularity of word-to-sentence translation, solves the local order problem, makes some idioms translated very real, and greatly improves the translation quality. But long-distance orders are still very few. The phrase-based model can correctly translate phrase segmentation and can also be used for partial order fitting, but the order fitting between expressions is very small. For long-term sequence adjustments, hierarchical expressions can solve this problem. Hierarchical sentences, as the name suggests, consist of a sentence itself containing smaller sentences, so that the rules of the hierarchical sentence itself can be used to solve the long-distance ordering problem between two sentences. In addition to the above-mentioned types of rules, hierarchical sentence model rules also include non-terminal and unsigned rules, that is, standard sentence model rules. The above rules are based on synchronous syntax without context. Therefore, the hierarchical sentence model is a synchronous formal grammar without context. Since the rules do not contain any linguistic knowledge, such as some marks and language relations, the hierarchical sentence model is just a formal grammar model [11-12].

Synchronous contextless grammar is a special form of contextless grammar. It plays a very important role in many fields, and one of its important uses is to express the recursive relationship between two languages. In the case of a source language sentence, there are many possibilities to derive it from a synchronous grammar without context, so that there is more than one possible translation in the target language segment. Then, this article created a model to determine which derivative is better than others.

The hierarchical syntactic translation system is called the formal syntactic translation system. On the one hand, it obtains the parse tree of the source language through analysis, and extracts the syntactic tree of the target language segment of the context-free synchronous grammar, so as to obtain the translation but does not use the syntactic information, only training Bilingual aligns the rules of the corpus for translation.

(1) GLR algorithm

This paper uses the GLR algorithm to identify and analyze one sentence in each paragraph. For large categories of sentences, the main structure of the sentence is extracted to make a reference for the generation and translation of the sentence.

(2) The analysis steps of the GLR algorithm are as follows:

1) Initialization. The analysis pointer points to the input symbol to be analyzed, and the termination flag is cleared.

2) Assignment of symbols. If there is no termination flag, the mapping function is used.

3) Consult the ACTION table to determine which action will be performed by the following actions:

When moving in, the current state and current symbol are pushed onto the stack, and the analysis pointer moves down. Observe that the GOTO table becomes the new state. If the conditions are not met, the termination flag is set. If it is a termination, refer to the "error report" on the termination of the crawl table.

4) Perform the following operations in order until the end of the analysis.

This article is an improvement on the basis of the classic GLR algorithm. The improved GLR algorithm is shown in formula (1):

$$G_a = (X_{\mathrm{b}}, X_{\mathrm{c}}, T, \eta) \tag{1}$$

In formula (1), X_{b} represents the cyclic symbol cluster, $X_{\mathrm{b}} \neq \phi$; X_{c} represents the termination symbol cluster, $X_{\mathrm{c}} \neq \phi$. T stands for the start symbol cluster.

Assuming that Q is any action in "and Q exists in X_{b}, we can get formula (2) after derivation:

$$Q \to \{\lambda, \mathrm{d}, m, \varepsilon\} \tag{2}$$

In formula (2), λ , d , m and ε represent the symbol on the right side of the action, the center symbol, the limit value and the marking method.

The process of intelligent recognition is shown in Figure 1:

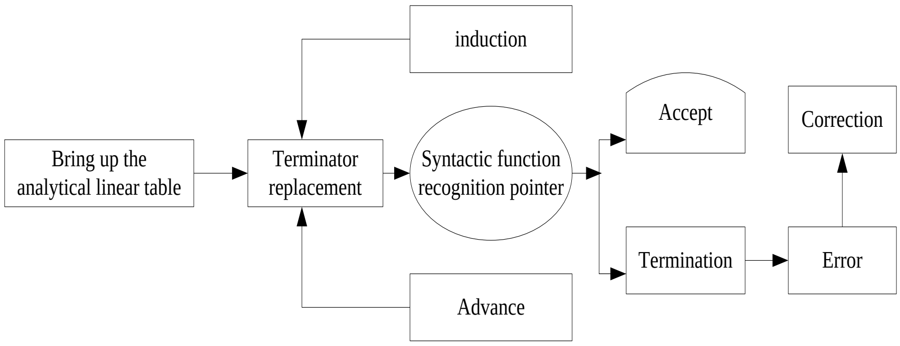

Figure 1: Intelligent Recognition Algorithm Correction Process

At present, in current English-Chinese machine translation algorithms, the final result of machine translation based on part of speech analysis on sentence corpus is often the result of matching the segmented sentence with the sentence corpus. For the GLR algorithm, it uses an analytic linear table to identify errors in the sentence recognition results, and performs a correction process by checking the highlighted content in the sentence body.

3. Model Verification Experiment

3.1 Verification Scheme

In order to test and evaluate the actual effect of the improved GLR algorithm for proofreading, the main performance indicators measured include: translation accuracy, translation speed, and update ability. Among them, the three kinds of English-Chinese translation engine parts of speech analysis level algorithms are statistical algorithm, dynamic storage algorithm, GLR algorithm and enhanced GLR algorithm.

3.2 Evaluation Process

This article gives 30 sentences to be translated by different machines. Through the system designed in this article, it is judged in terms of words, sentences, grammar and other issues. Then perform data analysis based on the verification results.

3.3 Identification Index

According to whether the expression is clear and the grammatical structure is correct, judges the recognition accuracy.

The total recognition time of the algorithm is multiplied by the weight, then added and divided by the number of sentence matches to judge the recognition speed.

The total algorithm update time is multiplied by the weight, then added, and then divided by the number of sentence matches to evaluate the possibility of update.

4. Analysis of Experimental Results

4.1 Evaluation Results of English-Chinese Translation Algorithms

According to the test results in Table 1, the improved GLR algorithm is the best in recognition accuracy, recognition speed, and update capability. Details as follows:

Table 1: Evaluation Results of Four English-Chinese Translation Algorithms

	Identify accuracy	Identify speed	Update ability
Statistical algorithm	76	73	62
Dynamic memory algorithm	91	80	66
GLR algorithm	91	84	70
Improved GLR algorithm	95	87	80

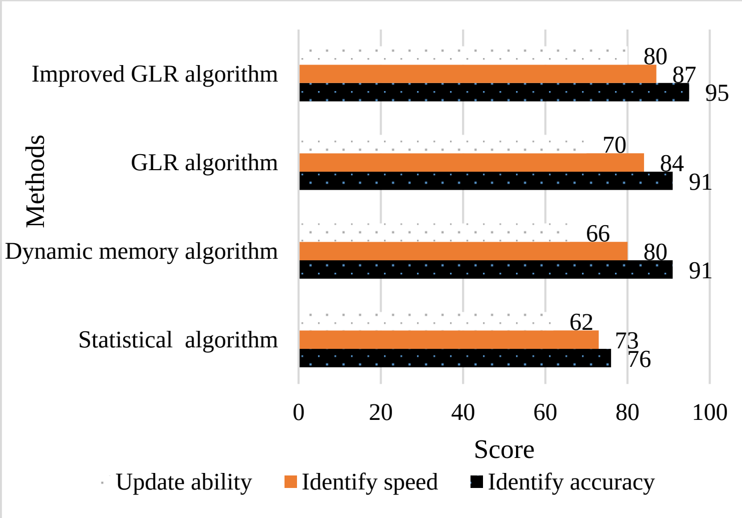

Figure 2: Evaluation Results of Four English-Chinese Translation Algorithms

As shown in Figure 2, we can see that the highest score based on the improved GLR algorithm is 95 points, and the lowest score based on the statistical algorithm is 76 points.

5. Conclusion

In the entire system development process, the requirements must first be analyzed, including functional modules, interface protocols, and database design. Then through the feasibility study, the overall structure of the system and the relationship between the various parts are determined. Next, establish the data collection process and real-time status monitoring interface required for the translation model computer intelligent calibration platform according to the user information content and specific requirements. Finally, the collected data is converted into standard text and stored in the upper computer code library to realize automatic identification, and generate target characters to control and analyze the translation process.

References

[1] Zheng Meng. Research on Intelligent English Translation Method Based on Improved Attention Mechanism Model. Electronic Science and Technology, 2020, v. 33; No. 374(11): 87-90.

[2] Cui Dan. Design of Computer Intelligent Proofreading System for English Translation. Modern Electronic Technology, 2019, 42(04): 179-182.

[3] Huang Dengxian. English translation software translation accuracy correction algorithm design. Modern Electronic Technology, 2018, 041(014): 170-172, 177.

[4] Hyatt J, Bienenstock E J, Tilan J U. A student guide to proofreading and writing in science. Ajp Advances in Physiology Education, 2017, 41(3): 324-331.

[5] Hu Bin. The transformation of proofreading objects and the revolution of proofreading models. East West, South and North, 2020, No. 558(10): 182-183.

[6] Krishnamoorthy D, Suwartadi E, Foss B, et al. Improving Scenario Decomposition for Multistage MPC Using a Sensitivity-Based Path-Following Algorithm. IEEE Control Systems Letters, 2018(99): 1-1.
[7] Zhang Meng. Chinese to Braille Translation Based on Braille Word Segmentation Using Statistical Model. Journal of Shanghai Jiaotong University (Science), 2017, 01(v.22): 84-88.
[8] Fejzer M, Przymus P, Stencel K. Profile based recommendation of code reviewers. Journal of Intelligent Information Systems, 2018, 50(3): 597-619.
[9] Sheng, Yanbin, et al. A two-scale attention model for intelligent evaluation of yarn surface qualities with computer vision. The Journal of the Textile Institute, Part 3. Technologies for a New Century, 2018, 109(6): 798-812.
[10] Daimi, Kevin. A Comparative Study of Neural Network Training Algorithms for the Intelligent Security Monitoring of Industrial Control Systems. Computer and Network Security Essentials. 2018, 10.1007/978-3-319-58424-9(Chapter 30): 521-538.
[11] Aroca R V, Burlamaqui AF, Gonçalves, Alves L. Method for Reading Sensors and Controlling Actuators Using Audio Interfaces of Mobile Devices. Sensors (Basel, Switzerland), 2012, 2(2): 1572-1593.
[12] Troussas C, Chrysafiadi K, Virvou M. An intelligent adaptive fuzzy-based inference system for computer-assisted language learning. Expert Systems with Applications, 2019, 127: 85-96.