2023 International Conference on Computer, Electronic and Materials Engineering (ICCEME 2023)

Application of Feature Extraction Algorithm in the Design of Interactive English-Chinese Translation System

Qingjuan Meng*

Jilin Animation Institute, Changchun, Jilin Province, China

Qingjuan_Meng@tom.com

* Corresponding author

Keywords: Automatic transfer; English-Chinese translation; Feature extraction algorithm

Abstract: In the English-Chinese translation system, the feature extraction algorithm is able to extract the feature semantics and optimal context, which leads to the improvement of translation accuracy. Although China has started to study the program development and architecture of translation systems, few results have been obtained and a perfect set of translation systems has not yet been formed. Feature extraction algorithm is a tool that can extract core information and serve as a mapping condition, which is more helpful to improve the quality of interactive English-Chinese translation. Therefore, this paper used a three-layer architecture to design the network architecture of the English-Chinese translation system, constructed a semantic mapping model, extracted semantic features of English-Chinese translated texts, clustered the graphical information of this feature and judged its attributes, so as to facilitate the implementation of the feature extraction algorithm part of the English-Chinese automatic machine-to-machine translation. Meanwhile, PLC bus technology and embedded program scheduling were used to design the software part of the translation system, including the output interface module for translation and conversion, cross-compilation module, IoT control protocol module and the underlying database module. In developing the English-Chinese automatic machine conversion translation system, the bus transfer automation control and the underlying adaptation development were used to better effect. Through line comparison experiments, the results showed that this system can obtain the optimal solution translation and has a high translation accuracy rate.

1. Introduction

In recent years, under the influence of globalization, China has become more and more closely connected with various foreign countries. Thus, to a certain extent, the importance of English is also reflected. Among English-Chinese communication tools, English-Chinese translation systems are a relatively important component and a hot spot for research [1]. Currently, there are two kinds of English-Chinese translation systems which are more common, namely, instance-based and rule-based. However, in actual translation work, these systems cannot extract feature semantics and optimal context, which reduces the accuracy of English-Chinese translation. In the face of more and more information, how to get the optimal translation from the huge amount of data is a key aspect of the English-Chinese system. It is found that by extracting the characteristics of machine translation, a deep neural network machine translation model can be constructed, which can identify translated texts, especially to meet the needs of translation with large access[2]. However, this system has limitations in grasping the optimal solution when performing machine translation. Feature extraction algorithm, as a method of extracting text features, can facilitate the extraction of text feature vectors by lexical blocking of text, a process of segmenting text sentences into meaningful caption sequences or blocks of words (token). In this paper, we designed an interactive English-Chinese translation system based on feature extraction algorithm. The article used the feature extraction algorithm to select feature semantics, constructed a semantic ontology mapping model, selected the optimal solution for the interactive English-Chinese translation process,

  DOI: 10.25236/icceme.2023.022

implemented the English-Chinese translation process using coding, and validated this system through comparative simulation experiments to prove its effectiveness[3].

Previous research has already achieved fruitful results in machine translation technology. In 2022, FU Xiaoxue introduced an English machine translation system based on a human-computer interaction algorithm[4], emphasizing the importance of user interaction. Concurrently, Zhao Juan and others optimized the online machine translation system using a deep convolutional neural network algorithm, enabling it to more effectively process complex language structures[5]. Additionally, the alignment template method by OCH and NEY[6] has achieved significant success in the field of statistical machine translation, laying a vital foundation for subsequent research. Google successfully narrowed the gap between human and machine translation with its neural machine translation system, bringing revolutionary changes to the entire field[7]. Mohammed and Aziz studied the English-Arabic machine translation based on a reordering algorithm, offering an effective solution for this specific language pair[8]. Lastly, Patel and Joshi devoted themselves to creating an automatic dataset for the specific Gujarati-English language pair, providing valuable data resources for optimizing machine translation systems[9].

2. Feature Extraction Algorithm

Feature extraction algorithm is a tool that can extract core information and serve as a mapping condition, which is extremely important to improve the quality of interactive English-Chinese translation [10]. In this study, this algorithm was used as a research tool to extract the best contexts in the original text. This study applied the mapping technique, implanted the contexts in the translated text, constructed the ontology mapping model. Meanwhile, it used this model to translate the utterances, so that the special contextual translation specimens can be improved. In this algorithm, the semantic translation kinds in the translated text were assumed to be W and the number of contexts was M. Then M_i=(i=1,2,... .W) was the number of contexts, where X_j={i=1,2... .W,j=1,2,... .Mi} denoted the directed vector result of dimension m; and X_i={Xi1, Xi2,... .Xm} denoted the probability of semantic translation using class W. The translation could be made to meet the standard translation context by performing qualified processing. The qualified translation context was assumed to be α_i, then the standard formula is as follows:

$$\alpha_i = \frac{1}{M_i}\sum_{j=1}^{M_i} x_j \tag{1}$$

The best context α was extracted from the original text and was calculated as follows:

$$\alpha = \frac{1}{W}\sum_{i=1}^{w_i} \alpha_{ij} \tag{2}$$

Meanwhile, the semantic translation context matrix was translated according to the best context criterion, denoted by S_B, and then the non-semantic translation context matrix was constructed, denoted by S_r. The equations for calculating the two matrices are as follows.

$$S_B = \sum_{i=1}^{W} (\alpha - \alpha_j)(\alpha - \alpha_j)^W \tag{3}$$

$$Sr = \sum_{i=1}^{W}\sum_{j=1}^{W_i} (\alpha_j - \alpha) \quad (\alpha j - \alpha)^W \tag{4}$$

Also, the optimal context of the two association matrices SrTSB was assumed to be η, and then f denoted the measure of association. The mapping of two association matrices was represented by the parameter value α, where W-1 denoted the upper limit of the optimal translation context of SrTSB, from which the optimal context was extracted, then the number of contexts could be obtained. that is, R (R ≤ W-1). Meanwhile, the optimal context was chosen as the translation criterion, and the characteristic utterance of the translated text, denoted by y, was expressed by the following equation:

$$Y = [\alpha_1, \alpha_2, ..., \alpha_R] \tag{5}$$

According to equation (5), the translation could be translated into the optimal contextual translation, and the semantics of the translation could be deeply restored, so as to achieve the translation purpose.

3. Designing the Interactive English-Chinese Translation System Based on Feature Extraction Algorithm

Firstly, the overall design of the system was framed in the study. The underlying database development module of the English-Chinese automatic machine translation system was constructed, and the English-Chinese automatic machine translation system was scheduled and the bus was developed using the embedded development method. According to the H.323 protocol, the application service adaptation layer and the development layer, the TETFSIP session protocol of this system was constructed, and then combined with the program initialization operation, the joint control parameter model of English-Chinese automatic machine conversion translation could be constructed. The signaling transmission control was used as the basic basis to construct the dialogue layer control protocol, and a three-layer architecture was used in developing the joint bus of the system. Meanwhile, when designing the system network architecture, the data link layer of the system was designed by constructing the session layer and application layer according to the layered idea, and then combined with the method of host port connection, whose overall structure is shown in Figure 1.

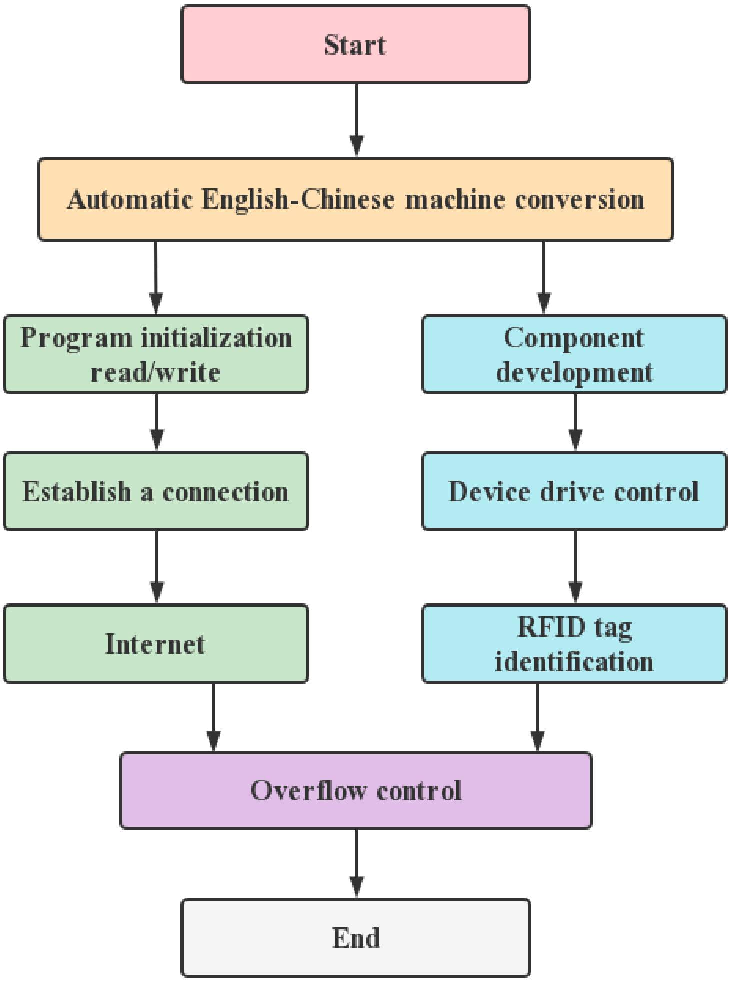

Figure 1. General Structure of English-Chinese Automatic Machine Translation System

According to the general structure of Figure 1, the QoS level control model of English-Chinese automatic machine translation control was constructed. The fusion processing of the underlying information base was carried out through the method of fusion management of the underlying database and semantic combination control [11]. The network topology of this system was obtained by applying GRPS grouping detection, as shown in Figure 2. Meanwhile, the feature extraction algorithm of English-Chinese automatic machine conversion translation was designed based on the overall system architecture, which can realize the accurate and optimal solution of machine translated text.

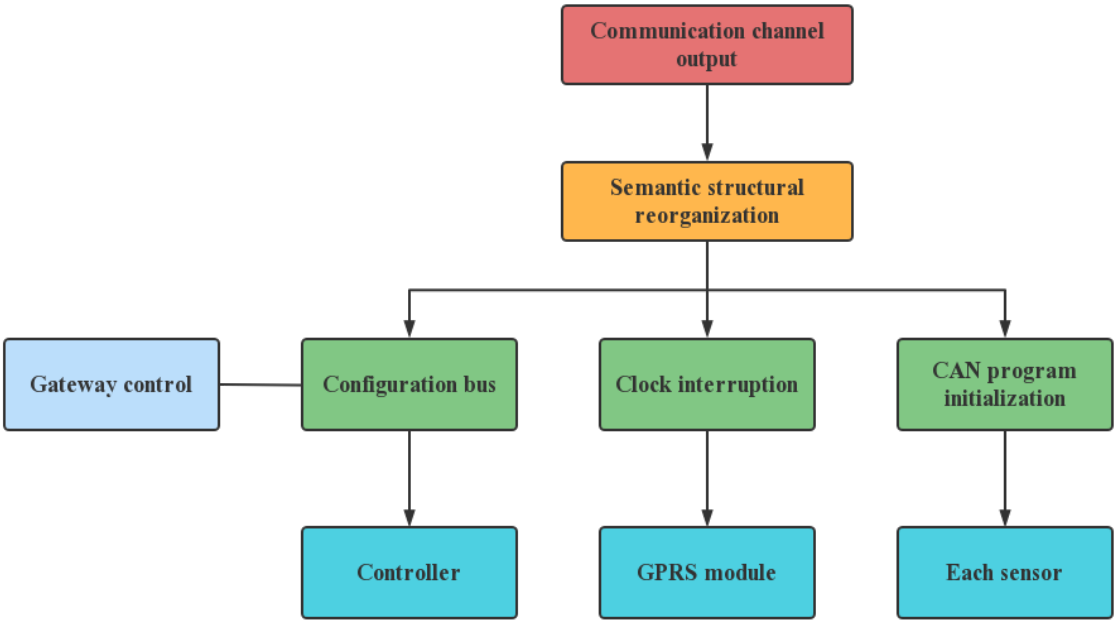

Figure 2. Network Topology Model of English-Chinese Automatic Machine Translation System

4. Extracting the Features of English-Chinese Automatic Machine Translation

4.1. Semantic ontology model for English-Chinese automatic machine translation conversion

In the analysis of the semantic features of English-Chinese automatic machine conversion translation, the overall structure of the model was taken as the basic basis, and then the semantic relatedness model expressions could be obtained through association rule reorganization and semantic relevance mapping [12], namely:

$$A = P(\frac{a}{2} + b)^2 \tag{6}$$

In the above equation, the set of entity probability density distributions was denoted by P; the adaptive resolution factor was denoted by b; and the equilibrium factor was denoted by a. By constructing a simulated information fusion model, the joint probability density space of n ordinary real word distributions could be obtained. By combining the mapping results of semantic information and applying linear weighting, the weighted control volume could be obtained, i.e.

$$B = \frac{(a+b) - A\sqrt{k}}{(a-b)^2} \tag{7}$$

In this formula, the translation text keyword weights were represented by k. Using the associated feature reorganization and semantic mapping, the features of the English-Chinese automatic machine conversion translation were extracted. Then, with the extraction results as the basic basis, the joint parameter distribution set of the semantic mapping could be obtained, and then combined with the fuzziness decomposition, so that the distribution mapping expressions could be obtained, namely:

$$C = \frac{\log_{k}(B + k) + 1}{\log_{k} B} \tag{8}$$

In obtaining the similarity fusion component of each subset of translations, the fuzzy decision of English-Chinese automatic machine conversion translation was used as the basic basis, i.e., {e1,e2,... , ei} 1≤i≤r, where the lexicon of the ith translated text is represented by i, and the subspace sequence fusion is used as the main basis when obtaining the template learning function, that is:

$$D = QN + \frac{C}{u} \tag{9}$$

In the above equation, the semantic relevance distribution set was represented by Q. N represented the joint feature parameter distribution set, which was able to obtain the feature clustering matrix. u represented the self-similarity fusion distribution coefficient. Meanwhile, combined with joint parameter analysis and distributed fusion of features, a joint feature analysis model could be constructed to facilitate the improvement of feature detection and analysis capability and output adaptiveness of this system.

4.2. Extracting the semantics of English-Chinese automatic machine conversion translation

In extracting the semantic context, an ontology fusion method of semantic units was applied to the construction of the joint feature distribution set by the semantic modification method [13], and then combined with the fuzzy synthesis decision method to the segmentation function of the automatic English-Chinese machine conversion, so as to obtain the semantic structure distribution, namely:

$$\sigma(t_1, t_2) = E^T + \frac{2(t_1 + t_2)}{D} \tag{10}$$

In this equation, the sampling time interval was represented by t2; the output time series was represented by t1, and the joint feature quantity was constructed, that is:

$$F = \delta(t_1, t_2) + uC \tag{11}$$

5. System Development Implementation

In the design of the software of the English-Chinese automatic machine translation system, the above algorithm was used as the basic basis for the embedded program scheduling and PLC bus technology. The content of this system includes the output interface module for translation and conversion, the IoT control protocol module and the underlying database module. In developing the bootloader for this system, the system software was designed by building a serial TWI storage method, especially in the underlying database [14]. The underlying database data loading was built, cross-compilation control and program loading was carried out in the embedded environment, and the Channel expansion bus technology was used for bus clocking and development. Also, when designing the underlying database, the AEE-NOTE68 system was selected, and the parameter configuration method of CAN was used to configure the PORT-MUX registers. Then cache control was performed to be able to obtain the initialization structure map of this program. Then the program control structure diagram of this system was used as the basic basis to design software development, and the software development implementation structure diagram could be obtained, as in Figure 3. In addition, according to the feature extraction algorithm, the semantic features of the translated text were extracted, and then the optimal solution of the translated text was grasped. The design of the English-Chinese automatic machine conversion translation system could be completed[15,16].

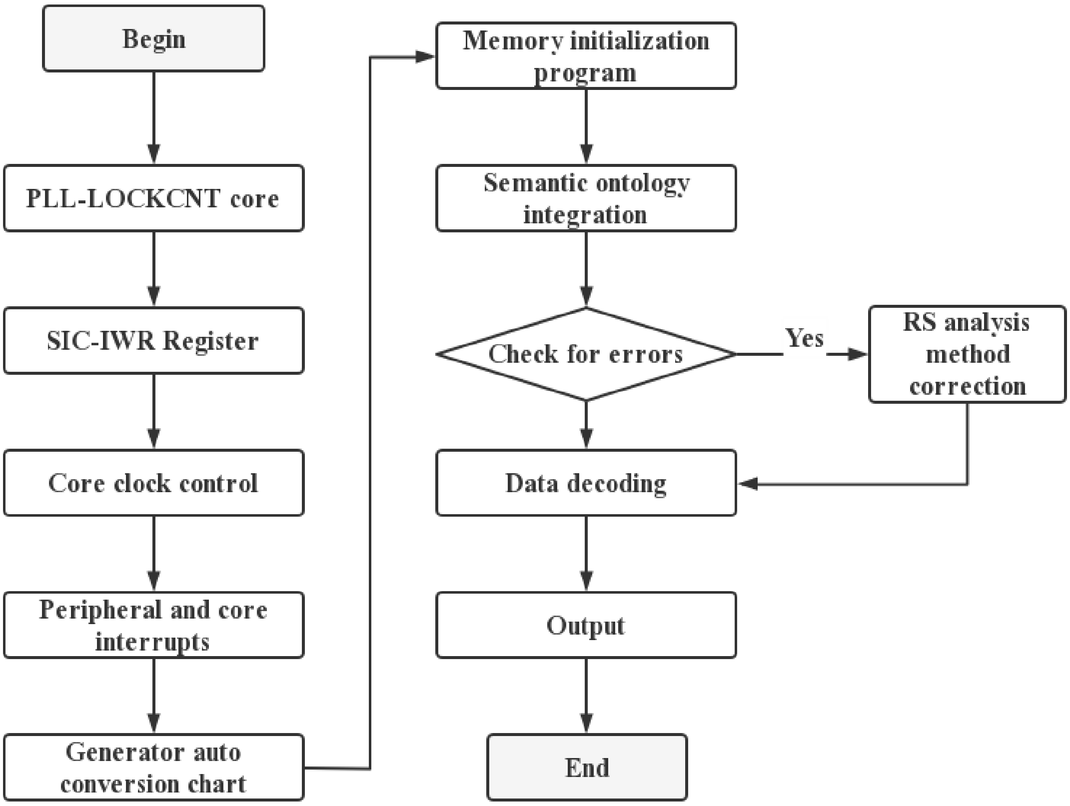

Figure 3. Software Development Implementation Structure

6. Simulation Test

In this study, when verifying the performance of the design system, the effectiveness of the results could be improved by selecting the simulation platform and setting the experimental parameters. Among them, the translation rate of English-Chinese machine automatic translation was 18KB /s, the extracted translation text was 800 characters, the length of the translation word was 18 bits, and the DA conversion frequency was 16.4khz. The sampling interval of interrupt mapping clock was 0.26ms. Meanwhile, this system was tested using the above parameters as the basic basis. The results showed that the semantic relevance and translation results can be fully represented according to the node-control distribution of the machine translated text, especially loose nodes suggesting that the keywords of the translated text do not fit with the semantic context. If the node-control points are closely distributed, it suggests that the translated text fits with the context and has strong coherence. At the same time, the analysis revealed that this designed system has a high accuracy in capturing keywords, fitting the semantics and translation, and has the characteristics of accuracy and coherence. In contrast, the two other systems have loose distribution of section control points, especially in capturing keywords, which cannot reflect the semantics of translated text, as shown in Table 1.

Table 1. Accuracy of English-Chinese Automatic Machine Translation Conversion

Number of iterations	The system used in this paper
100	0.935
200	0.985
300	0.992
400	0.995
500	1

By analyzing Table 1, it can be found that the accuracy of this system proposed in this study was as high as 93.5% at 100 iterations, while after 500 iterations, accurate translation could be achieved. This method has a higher accuracy compared with other methods. Meanwhile, in the experiment of practical applications, the number of iterations was set at 200, and its accuracy rate was 98.5%. The matching rate can fully reflect the accuracy of the automatic machine conversion translation system.

The interactive English-Chinese translation system proposed in this paper effectively integrates the feature extraction algorithm, significantly improving the quality and accuracy of translations. By adopting advanced feature extraction methods, the system can accurately capture and parse the semantic structure of the original text, ensuring that the translation remains true to the original intent while maintaining fluency. Furthermore, the interactive design allows users to directly participate in the translation process, catering better to their specific needs and preferences. This interactivity further enhances the personalization of the translation, offering users a more customized translation experience. In practical tests, the system demonstrated a high level of accuracy and a robust ability to capture semantics. Compared to other existing technologies, it showcased clear advantages, revealing the immense potential of feature extraction techniques in interactive system design.

7. Conclusion

In summary, this paper designed a set of English-Chinese automatic machine conversion translation system based on feature extraction algorithm. It took semantic mapping as the basic basis when extracting semantic features of translated text, then grasped the optimal solution, designed the software architecture, and verified its performance by using simulation platform. The results showed that this system has high accuracy and superior performance, which can meet the requirements of translation work. Clearly, the introduction of feature extraction not only streamlines the translation process but also ensures contextual accuracy and cultural relevance in translations. Looking forward, with the swift rise in user demand for accurate and rapid translations, this system holds the potential to establish a new benchmark in the English-Chinese translation field, presenting us with a novel amalgamation of accuracy, efficiency, and user interactivity.

Acknowledgement

Project Foundation: Research on English Movie Subtitle Translation from the Perspective of Multimodal Discourse Analysis, Key Project of Scientific Research of Jilin Animation Institute, No. KY22SZ14.

References

[1] Dongyan Li. Design of an automated English-Chinese translation system based on human-computer interaction. Automation & Instrumentation, 00(06), pp.181-185, 2022.

[2] Min Fu. Design of the automatic conversion and translation system based on feature extraction. Automation & Instrumentation, 00(05), pp.166-169, 2022.

[3] Wengai Wang. Cross-cultural perspective shift and translation skills in English-Chinese translation. Sinogram Culture, 00(10), pp.143-145, 2022.

[4] FU Xiaoxue. English Machine Translation System based on Human-Computer Interaction Algorithm. In: 2022 IEEE 2nd International Conference on Mobile Networks and Wireless Communications (ICMNWC). IEEE, 2022. p. 1-4.

[5] ZHAO Juan, et al. Optimization of machine online translation system based on deep convolution neural network algorithm. Computational Intelligence and Neuroscience, 2021, 2021.

[6] OCH Franz Josef; NEY Hermann. The alignment template approach to statistical machine translation. Computational linguistics, 2004, 30.4: 417-449.

[7] Mohammad Norouzi, et al. Google's neural machine translation system: Bridging the gap between human and machine translation. arXiv preprint arXiv:1609.08144, 2016.

[8] MOHAMMED Ehsan A.; AZIZ Mohd J. Ab. English to arabic machine translation based on reordring algorithm. Journal of Computer Science, 2011, 7.1: 120.

[9] PATEL Margi; JOSHI Brijendra Kumar. GEDset: Automatic Dataset Builder for Machine Translation System with Specific Reference to Gujarati-English. In: International Advanced Computing Conference. Cham: Springer International Publishing, 2021. p. 519-528.

[10] Li Zhao. Application of feature extraction algorithm in the design of interactive English-Chinese translation system. Information Technology & Informatization, 00(01), pp.138-141, 2022.

[11] Qizhen Zhang, Xianhong Sun. Design of interactive English-Chinese machine translation system based on B/S framework. Modern Electronics Technique, 44 (24), pp.69-73, 2021.

[12] Juling Zhang. Design of English-Chinese machine automatic conversion translation system based on feature extraction algorithm. Automation & Instrumentation, 00(07), pp.148-151, 2021.

[13] Jing Mi. Study of English-Chinese translation system based on feature extraction algorithm. Microcontrollers & Embedded Systems, 20 (03), pp.54-56, 2020.

[14] Lei Xing. Design of translation automatic generation system for English-Chinese machine translation. Modern Electronics Technique, 41 (24), pp.86-89, 2018.

[15] Qian Yu. Design of interactive English-Chinese translation system based on feature extraction algorithm. Modern Electronics Technique, 41 (04), pp.161-163+168, 2018.

[16] Lin Wang, Wuyin Liu, Xiaobo Liang. Design and implementation of English-Chinese bidirectional term dictionary system on philosophy and social science. China Terminology, 16 (02), pp.18-21, 2014.