Study of the Establishment of a Reliable English-Chinese Machine Translation System based on Artificial Intelligence

Fu Xuan, Lu Wei
Hebei Jiaotong Vocational and Technical College
Shijiazhuang 050035, Hebei, China
FXFXFX3358@sina.com

Zhu Ligang
School of Foreign Language of Tongji University
Shanghai 200092, China

Zhou Shuai
Hebei College of Industry and Technology
Shijiazhuang 050091, Hebei, China

ABSTRACT: *Since twenty-first century, the increase in communication among different countries has made the need for the language translation of the enterprises and individuals more and more. Artificial translation is accurate, but the cost is too high and time-consuming; while the cost of the machine translation is not only low, but the speed is fast. However, the accuracy of machine translation has been criticized by users, therefore, how to build a new generation of machine translation system to improve the accuracy has been imminent. Based on this status, a reliable English-Chinese machine translation system based on artificial intelligence is established in this paper, and the principles that should be followed in the process of establishing the system are described in detail. The overall framework, the translation algorithm and the working flow of the system are discussed, and the sentence alignment method based on the translation is proposed. The research results show that the reliable English-Chinese machine translation system based on artificial intelligence designed in this paper can improve the credibility and accuracy of machine translation.*

Keywords: Artificial Intelligence, Reliability, English-Chinese Machine Translation

Received: 10 February 2017, Revised 14 March 2017, Accepted 19 March 2017

1. Introduction

As a cross discipline, machine translation is a high degree of knowledge intensive technology, which needs to be completed by the linguists and computing technology experts [1]. The perfection in the research and development of the machine translation

is not yet achieved to meet the demands of the users which is very high and accuracy needs to be ensured in the translation operation [2]. However, even if the machine can help people to complete some intelligent activities, such as neural computing, but the ability of intelligent translation is different from the traditional intelligent computing[3]. The regularity of the operation is strong and the technical staff can sum up the rules to use the computers to simulate the process [4]. The characteristics of the translation work are that the normalization is weak, and its process is difficult to be simulated [5]. So far, the study of the mechanism and the way of the brain processing language has been still very little, so machine translation effect has been still not satisfactory, especially the translation of complex sentences and texts [6]. Therefore, it is scientific and essential to build a trusted English-Chinese machine translation system based on artificial intelligence [7].

2. Background

After the advent of the first electronic computer [8], in 1947, engineer Weaver officially proposed the concept of machine translation in his book “Translation”, and in the decades after that, the machine translation has also experienced a tortuous process of development. In 1954, the first machine translation system was proposed, which proved the feasibility of the machine translation system to the world [9]. In 1966, the ALPAC report proposed that it was unnecessary to add more investment to the MT (Transfer System Machine) [10]. The low tide period of machine translation system research was coming [11]. From 1975 to 1989, the second generation of the machine translation systems have been actively developed [12]. This generation of machine translation system has combined many technologies, such as knowledge and separation algorithms, modular design, syntax analysis and semantic analysis of a variety of strategies, and most of them have applied artificial intelligence technology. After 1989, the third generation of machine translation system was born and developed rapidly; this generation of translation system was both based on statistics and the actual case, and combined with the advantages of the former two generations of machine translation.

In twenty-first Century, machine translation system also should progress with the times, therefore as an initial step, this paper presents a reliable English-Chinese machine translation system based on artificial intelligence. In the third section, the established principles, the overall framework of the translation algorithm, and the work flow and the automatic alignment method of the bilingual sentences of the machine translation system based on the artificial intelligence are introduced. In the fourth section, the performance of the English-Chinese machine translation system based on artificial intelligence is analyzed, and finally the conclusion is made (as shown in Figure 1).

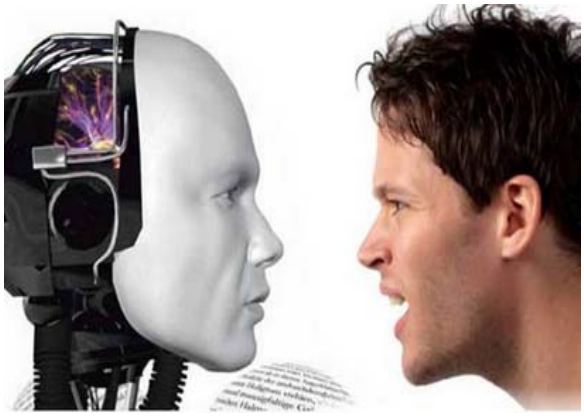

Figure 1. English-Chinese machine translation system based on artificial intelligence and reliability degree

3. Methods

3.1 Principles of the establishment of the English-Chinese machine translation system based on artificial intelligence

It is a very difficult project to build a reliable English-Chinese machine translation system based on artificial intelligence, which not only needs to guarantee the intelligence of the system, but also maintains the credibility of the system. The following important principles are adopted in the establishment of this system:

Data and procedures are independent of each other. The dictionary used by the machine and the system data rule is represented

separately in the form of a database file on separate disks. The independent of the translation program is not only conducive to data management, maintenance, update, but also conducive to the upgrading and expansion of the translation process. From the point of view of the process control mechanism, “control” and “implementation” are the two different mechanisms, where the execution mechanism is responsible for the data analysis and operation, and the control mechanism is responsible for the control of the whole system at a higher level. The improvement of the control mechanism will improve the execution mechanism, but will not affect the logical performance of the execution mechanism, and the change of the implementation mechanism will not fundamentally affect the control mechanism, because the control mechanism is only implemented in the underlying implementation mechanism.

Designing a more specialized machine dictionary will help to improve the quality and efficiency of the English-Chinese machine translation system based on artificial intelligence. The form of “the basic entry professional entry” is used in this system. The literary translation of professional fields is based on dictionary translation and professional entry. Such as the literary translation of the automatic control technology, the dictionary only needs to have basic entry and computer professional entry, when the professional field is changed, it only needs to change a part of the professional items, and maintain the basic items. The dictionary with this structure not only has the inheritance, but also can be removed and replaced, which is conducive to the expansion and improvement of the dictionary.

Paying attention to the order and classification rules of translation: The English-Chinese machine translation system based on the artificial intelligence not only needs to have a suitable machine dictionary, but also needs the appropriate rules, and follows the principles of the first special after the general. The English-Chinese machine translation system based on artificial intelligence should carry on the scientific classification on the basis of the principles of syntax analysis and semantic analysis.

3.2 Overall framework design of the English-Chinese machine translation system based on artificial intelligence

The framework of an English-Chinese machine translation system based on artificial intelligence can be roughly divided into the following three parts (as shown in Figure 2):

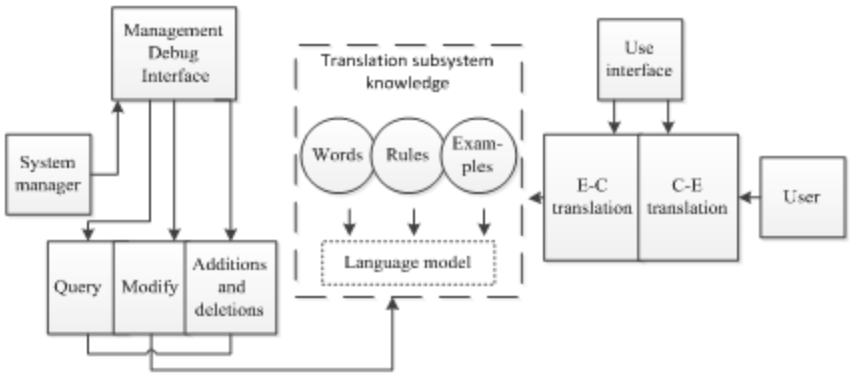

Figure 2. The framework of the English-Chinese MT System

Translation system knowledge base: The translation process of the English-Chinese machine translation system based on artificial intelligence can be regarded as an application and knowledge of the process of reasoning. Knowledge representation is the basis of this process, and the knowledge representation methods used in the English-Chinese machine translation system based on artificial intelligence is divided into two categories: internal and external knowledge. The external knowledge is stored in the knowledge base, which is managed by the language knowledge, researchers, such as the dictionary and all kinds of rules. The internal knowledge is the sentence to describe the syntax and semantic features of knowledge in the process of translating, such as the tree graph, the structural feature and the semantic network. The knowledge base of the translation system includes a language model, a dictionary, a lot of rule bases and a case base. The dictionary is divided into basic bilingual dictionaries and bilingual dictionaries. Rule base stores the phrase rules, sentence patterns, sentence pattern matching rules, English-Chinese translation rules, and so on. All the rules have the same data structure. Case library stores the English -Chinese bilingual examples and related information.

Part of the processing system: English word processing includes English automatic segmentation and fuzzy word processing,

where this part is the basis of the combination of phrases and sentence matching, which automatically uses the maximum matching word segmentation algorithm, and the system uses the rules and statistical methods to eliminate and merge. In the English analysis part, there are two parts in the system, the combination of phrases, and the sentence matching. The task of phrase merging is to use a variety of phrase rules and methods to divide the class words into five basic phrases and the sentence matching is based on the combination of phrases. In the Chinese- English translation and the Chinese generation part, there are three levels of transformation and generation, including words, phrases and sentences, the tasks and algorithms of each layer are different.

Operation interface: The interface includes user interface and administrator management debug interface. For an actual machine translation system, debugging language rules and dictionaries are very important. In the debugged good machine translation system management interface, the system administrator can conveniently and intuitively debug the language rules and dictionaries, where it can improve the efficiency of debugging and improve the quality of the language knowledge base. The management and debugging interface of the English-Chinese machine translation system based on the artificial intelligence is responsible for the maintenance of the knowledge base management and translation and debugging. Knowledge base management function helps the system administrator to create and manage each knowledge base and inquire the debugging operation. For example, the system administrator can carry out debugging through the observation of the specific examples of the translation process of the language knowledge base. The system administrator can manage the production process of any syntax component and the characteristics of the corresponding node attributes and values.

3.3 Translation algorithm and work flow of the English-Chinese machine translation system based on the artificial intelligence

People often solve the new sentence through the past learning experience of learning a language. Based on this idea, the examples from the corpus can be used to translate new sentences, that is to say, the experience of the past can be aroused by imitating human analogical thinking, and the grammatical structure, the semantic choice and the current translation of the target language are obtained. The translation algorithm of the English-Chinese machine translation system based on the artificial intelligence is based on the corpus as a priority, which is the rule of the auxiliary translation throughout. For each sentence, first the pre-treatment is carried out, the most similar sentences and phrases in corpus based translation template are found out, if the size of the corpus is not large enough, then the translation is achieved by the rules concluded by the linguists, including the segmentation rules, the rules of the merger, the phrase rules, and the target language rules and so on. The translation process of the trusted English-Chinese machine translation system based on the artificial intelligence is shown in Figure 3.

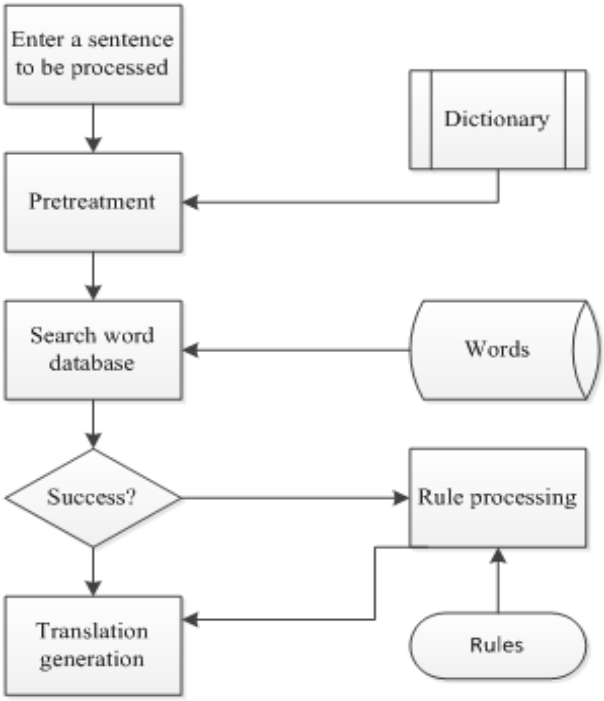

Figure 3. General procedure of MT System

Among them, the translation method is based on the principle of conversion, follows the edge analysis, and simultaneously generates the design principles. The specific translation algorithm is shown in Figure 4:

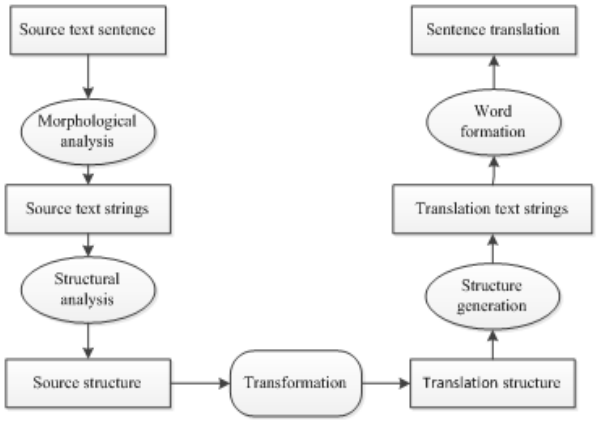

Figure 4. Procedure on MT System based on rules

Text files, keyboard input and scan input of the three ways can be used to input the original text.

The English morpheme analysis stage is divided into overlapping word processing and segmentation of the two steps, English word segmentation uses the bidirectional maximum matching algorithm. The rule base is used to eliminate ambiguity in the processing of the text word segmentation. When a word segmentation ambiguity occurs, it is unnecessary to make a judgment, and it is needed to be maintained to the structural analysis phase to be processed. If there are no disambiguation rules, the Chart Parsing algorithm can be used to transfer the default phase structure, and the phase analysis adopts a top-down and local sub tree structure generation. The phase transformation algorithm uses the method of the combination of partial sub tree transform algorithm and top-down global sub tree displacement algorithm.

3.4 Automatic alignment method for bilingual sentences in the English-Chinese machine translation system based on the artificial intelligence

In the existing machine translation systems, the applications of statistics and case method are more and more extensive, and the bilingual corpus is more and more important. The bilingual corpus has a variety of forms, such as text level, sentence level and lexical level. It is very easy to get to the level of text, but it is not very useful. Automatic translation of bilingual texts from the textual level is a process of automatic translation and alignment of bilingual sentences. It can further find the relationship between the corresponding vocabulary, and obtain the basis of other translation knowledge.

A large number of studies have been made on the translation by the foreign scholars in India and Europe, and a good result has been achieved. Their approach is summed up in two ways: -Based on the length and alignment method, and based on the word alignment method. But there is a big difference between Chinese and India, although these methods can also be used directly, but the effect is not obvious. And due to changes in the language environment, the use of these methods has also a lot of restrictions. On this basis, Wang Bin proposed a sentence alignment method based on comprehensive information, but the accuracy is not high.

Therefore, the establishment of the English-Chinese machine translation system based on the artificial intelligence needs to propose a new alignment method. In this paper, a new method is proposed, which is based on the translation method. The specific contents are as follows:

The formal representation of sentence alignment is: sentence alignment is to regard the translation of a sentence as a sentence with two languages together, and its formal representation is: Setting M and P are the translation the target text of each other, and can be expressed as a combination of sentences, namely, $M = m_1 m_2 ... m_n$, $P = p_1 p_2 ... p_n$. The Setting U is the minimum alignment of S and T, it consists of x sentences m_{i1}, ..., m_{i+x-1} in M and y sentence p_{i1}, ..., p_{ix-1} in P. So the alignment of M and P can be expressed as a sequence of all U : $u_1 u_2 ... u_k$, the task of Ui ($i = 1 \sim i$) is to find a minimum alignment sequence.

The discussions based on the translation of the English-Chinese bilingual sentence alignment are as follows. The choice of word evaluation function is based on the choice of the translation of sentence alignment method, which involves not only the English sentences, but also the Chinese sentences. In addition, in order to reduce the influence of the dictionary on alignment accuracy, the part of speech tagging and additional processing are not introduced. As long as there is a translation and Chinese-English sentence in a string matching, it is considered that the English word is the translation in Chinese sentences. This occurs in most cases. Although the nature of the English words is different, but its meanings in Chinese are similar, and are not always included in the dictionary.

For the dynamic programming algorithm and overall evaluation function, the dynamic programming algorithm is used in this paper in the process of choosing alignment. The Setting E_1 is the number of sentences in the English segment, and E_2 is the number of sentences in the Chinese segment. The value of each possible sentence in the English and Chinese texts is evaluated, and according to the consideration of the system, which includes the following.

The number of evaluation values of 1 : 1 is:

$$E_1 * E_2 \quad (1)$$

The number of evaluation values of 1 : 0 is:

$$E_1 * (E_2 + 1) \quad (2)$$

The number of evaluation values of 0 : 1 is:

$$(E_1 + 1) * E_2 \quad (3)$$

The number of evaluation values of 2 : 1 is:

$$(E_1 - 1) * E_2 \quad (4)$$

The number of evaluation values of 1 : 2 is:

$$E_1 * (E_2 - 1) \quad (5)$$

The number of evaluation values of 2 : 2 is:

$$(E_1 - 1) * (E_2 - 1) \quad (6)$$

Therefore, the sum of the values of the total is the sum of the values above, namely, $B = 6* (E_1 * E_2) - E_1 - E_2 + 1$. According to this value, the dynamic programming algorithm is used, and the opposite of the output is in the path of the word, which is the segment of the aligned sentence. In the specific process of solving, the node is the best path from start to the current node, the predecessor value of the node is evaluated, and the two values are used to select the best path. The tag in the path includes the alignment of the string. The overall evaluation function value of each path is the value of the review and the path of all alignments.

According to this value, the dynamic programming algorithm is used to find the path of the sentence which is the largest of the total evaluation value from the beginning to the end of the paragraph. In the process of solving the specific problem, the node from the starting point to the current node is first obtained, and the weight of the optimal path is recorded, then, the evaluation value of the start node is obtained, and the two values are used to select the optimal path. The marks on the path make up the alignment of the sentences are on the string. The value of the overall evaluation function of each path is the value of the evaluation of all alignment sentences on the path.

4. Results, analysis and discussion

For the establishment of the English-Chinese machine translation system based on the artificial intelligence, the automatic alignment method of bilingual sentences and the accuracy of the system is very important for the whole system. Based on the above mentioned bilingual alignment method based on translation, this section conducts experiments to test the accuracy and reliability of the English-Chinese machine translation system based on the artificial intelligence. Six forms are considered in the experiment: 1:1, 1:0, 0:1, 2:1, 1:2, 2:2, and before the experiment, the English and Chinese texts are carried out preprocessing, and are divided into the form of a sentence. The whole alignment results are summarized in Table 1.

Matching form	Number of aligned sentences	Correct number	Correct rate
1 : 1	1988	1982	99.7%
1 : 2	93	84	90.3%
2 : 1	172	166	
2 : 2	16	2	12.5%
1 : 0	0		
0 : 1	0		

Table 1. The Experiment results of the experimental group

From the above table, it can be seen that the sentence alignment method based on the translation fundamentally eliminates the mistakes caused by text spread in the text in the based method for length. In addition, the method doesn't involve the issue of training text.

In addition, we carried out the same experiment with ordinary machine translation system, and the experimental results are shown in Table 2. The comparison of the two groups is shown in Figure 5.

Matching form	Number of aligned sentences	Correct number	Correct rate
1 : 1	1988	1766	88.8%
1 : 2	93	67	72.0%
2 : 1	172	142	77.3%
2 : 2	16	1	6.3%
1 : 0	0		
0 : 1	0		

Table 2. The Experiment results of the control group

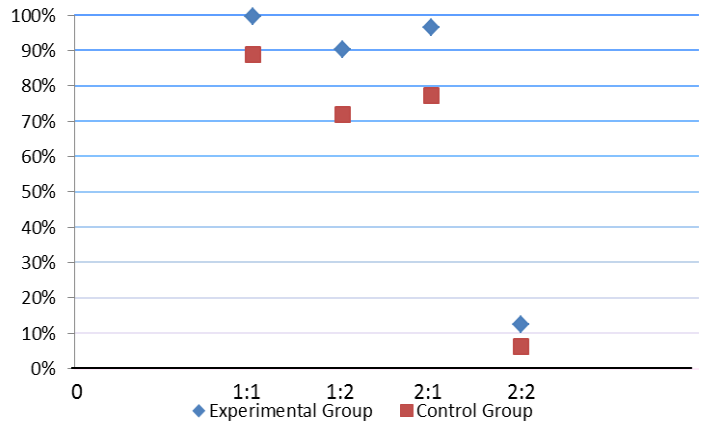

Figure 5. Comparison of experimental results

From the above experimental results, it can be seen that, for the 1 : 1 form, the accuracy of the translation of the English-Chinese machine translation systems based on artificial intelligence designed in this paper increases from 88.8% to 99.7%; for the 1 : 0 form, which increases from 72% to 90.3%; for the 0 : 1 form, which increases from 72% to 90.3%; and for the 2 : 1 form, which increases from 72% to 90.3%; Thus it can be obtained that the accuracy of the translation of the English-Chinese machine translation systems based on artificial intelligence is significantly higher than the ordinary machine translation, where the progress confirms the feasibility and the scientific nature of this machine translation system based on the artificial intelligence.

5. Conclusions

In order to further improve the scientific accuracy of the machine translation, the machine translation is tested with more accurate services. Based on the literature review and dynamic programming and other methods, a kind of reliable English-Chinese machine translation system based on artificial intelligence is established in this paper, and the principles that should be followed in the process of establishing the system are described in detail, the overall framework, the translation algorithm and the working flow of the system.

The research results show that the reliable English-Chinese machine translation system based on artificial intelligence successfully combines many features of the previous machine translation systems, and on this basis, it is further upgraded. The key point is that the new sentence alignment method is put forward, which is the automatic alignment method based on the target text. The experiment proves the feasibility and accuracy of this method and the translation system. However, the research on semantic analysis and translation of web pages is still not perfect, where it is needed to pay more efforts.

Acknowledgment

The work presented in this paper is supported by the study of Educational Bureau of Hebei Province with the title, “Study on Translation of Urban Rail Transit Guiding Signs System”, and the research number is, SZ1623.

References

[1] Zhang, D M., Zhu, Y., Jin, Y H. (2014). Recognition and Transformation of Normal Sub-Sentence in Chinese-English Patent Machine Translation. *Applied Mechanics & Materials*. 513, 4610-4616.

[2] Chen, H., Yu, Y S. (2013). Automatic Translation in Chinese and English Based on Mixed Strategy. *Advanced Materials Research*. 760, 1942-1946.

[3] Silva, W R L D., Temberk, P. (2013). Shooting-inspired fuzzy logic expert system for ready-mixed concrete plants. *Journal of Intelligent & Fuzzy Systems Applications in Engineering & Technology.* 25 (2) 481-491.

[4] Liu, X., Zhu, Y., Jin, Y. (2014). Recognizing and Reordering the Translation Units in a Long NP for Chinese-English Patent Machine Translation. *Communications in Computer & Information Science*. 493 (21) 33-48.

[5] Ke, X., Ma, Q. (2014). Study on an impersonal evaluation system for English-Chinese translation based on semantic understanding. *Perspectives Studies in Translatology*. 22 (2) 242-254.

[6] Shi, Y B., Jian, Y F., Qiu, S Y. (2012). Test System for Airborne ILS Navigation Apparatus Based on Artificial Intelligence and Virtual Instrument. *Advanced Materials Research*. 588. 1602-1605.

[7] Kardan, A A., Aziz, M., Shahpasand, M. (2015). Adaptive systems: a content analysis on technical side for e-learning environments. *Artificial Intelligence Review.* 44 (3) 365-391.

[8] Gao S, Yang X, Yu Z. (2015). Chinese-Naxi machine translation method based on Naxi dependency language model. *International Journal of Machine Learning & Cybernetics*. 15, 1-10.

[9] Sajadi, A., Borujerdi, M R M. (2013). Machine translation based on unification link grammar. *Artificial Intelligence Review*. 39 (2) 109-132.

[10] Wu, X., Zhuo, S. (2014). Chinese Text Sentiment Analysis Utilizing Emotion Degree Lexicon and Fuzzy Semantic Model. *International Journal of Software Science & Computational Intelligence*. 6 (4) 20-32

[11] Nguyen, Q., Nguyen, A., Dinh, D. (2012). An Approach to Word Sense Disambiguation in English-Vietnamese-English

Statistical Machine Translation. Rivf International Conference on Computing and Communication Technologies, *Research, Innovation, and Vision for the Future*. 21 (1) 1-4

[12] Espla-Gomis, M., Sanchez-Martinez, F., Forcada, M L. (2015). Using Machine Translation to Provide Target-Language Edit Hints in Computer Aided Translation Based on Translation Memories. *Journal of Artificial Intelligence Research.* 53, 169-222