Scholar Publishing Group
Machine Learning Theory and Practice
https://doi.org/10.38007/ML.2021.020101
ISSN 2790-0983 Vol. 2, Issue 1: 1-8

AI machine Translation + Post Editing in English Chinese Translation under the Background of Big Data

Yun Li[*]

Nanchang Institute of Technology, Jiangxi 330099, China

lily110807@163.com

[*]*corresponding author*

Keywords: Big Data, Artificial Intelligence, Machine Translation, Classification Decision Tree, Post Editing

Abstract: In today's era of big data, due to the rapid development of software technology, new computing methods are constantly emerging, and computers are constantly improving their performance to meet everyone's needs. At the same time, the progress of machine translation technology has been recognized by the industry and customers. Important aid and efficient means. Machine translation is more and more widely used in today's translation activities, and its role and influence can not be underestimated. Some experts even predict that it will replace manual translation in the future. This paper aims to analyze and study the components of AI MT+Post editor in English-Chinese translation. This paper analyzes the results of intelligent translation and concludes that machine translation has the characteristics of randomness and time-varying. By applying graph theory and ID3 algorithm, it is established that AI MT+ post-editing is used in EC translation, and corresponding technologies are applied to prune classification decision trees. The classification rules are generated, and the classification model construction in EC translation of the above-mentioned post-editing is completed. The experimental results show that the translation + post-editing algorithm designed after the error correction of the GPC algorithm is applied, and the response time in the English-Chinese translation system is less, the overshoot is small, and the control effect is improved by 20% compared with the GPC before error correction.

1. Introduction

Machine translation, a part of computer linguistics, is a process of transforming a source language into a target language. Up to now, machine translation systems can be roughly divided into rule-based, corpus based and artificial neural network based systems according to the technologies they use [1-2]. The direct translation system generally replaces the words or sentences in the source language with those in the target language, i.e. literal translation, such as various electronic dictionary software; the rule-based machine translation system sets a rule between the source language and the target language to express the relationship between them; the corpus based translation system achieves a certain development in computer technology, machine translation system developed on the basis of large-scale bilingual corpus created by language researchers [3-4]. The machine translation system used in this study is a Chinese English online translation tool supported by Google neural machine translation, which was launched by Google at the end of 2016 [5-6].

Post-translation editing is the editing of the output text of machine translation to test whether the text is accurate and fluent, thereby improving the text, facilitating understanding and correcting errors [7-8]. There are two main types of post-editing: comprehensive and simple; according to the purpose of translation of the text or the client's request, select the appropriate use method. Zhao, a Chinese scholar, believes that international standards also point out that the requirements of the comprehensive type are more accurate, easy to understand, and the style is appropriate, and the language, usage, and symbols cannot be wrong. In order to make the translation quality of post-editing output comparable to that of human translation output [9]. The purpose of the simple type is to capture the main points of information, not to publish, that is, to provide the main idea and general idea of the article. Post-editing output at this level requires accurate content, ease of understanding, and appropriate style. In summary, post-editing is generally used in translation subjects where translation quality, efficiency, and cost are important. Colakoglu, a foreign scholar, believes that no matter what form of post-translation editing, it is necessary to ensure that the information is complete without adding or reducing; editing inappropriate content; reorganizing sentences with ambiguous or incorrect meaning; the grammar, syntax and semantics of the target language content. Correct; and the terms required by Client and Field are consistent [10].

This paper analyzes the professional content of translation processing projects, usually copies and pastes a large section of the original text into the translation machine, and then on the basis of the translation, either completely adopts or modifies it to form its own initial translation. In the process of using translation machine, it is found that the common feature of the two texts processed by the translator is the rapid processing of simple and professional texts, including basic structure and professional terminology. Allow translators to quickly apply professional terms, reduce review time, and become familiar with articles.Then, according to the situation, the professional dictionary or search engine can be consulted for zero post translation editing or simple post translation editing.

2. Agorithm of AI Machine Translation + Post Editing in English Chinese Translation

2.1. Application of Artificial Intelligence Algorithm

According to the architecture model of artificial intelligence system application, we divide the system into the following three main functional modules: English text preprocessing functional module, feature item processing functional module, similarity calculation functional module. Each module is an important process in the similarity calculation process, and the specific work of each module is introduced according to the method:

(1) English text preprocessing module.

This section deals specifically with the customization of English wording and word extraction for comparing two English words. The accuracy of text sharing directly affects the accuracy of similarity. With the continuous research of English grammar by many scholars, English grammar technology is also developing continuously. At present, some open source partitioning software has appeared, among which the most commonly used basic software partitioning methods are SCWS simple English partitioning system and NLPIR partitioning system. The SCWS English subtitle system is developed using PHP. Its basic structure is the same as the English grammar method based on dictionary frequency. Based on a personal word frequency dictionary, English words can be divided into word patents to remove words from English texts. It supports English text encoded by scripts such as UTF-8, GBK, BIG5, etc. It is a popular English subtitle program. NLPIR English word segmentation system is an open source English language writing system developed by the Institute of Technology, Chinese Academy of Sciences, also known as ICTCLAS English word segmentation system. You can split English text into individual words, also highlight sections of partial text, and support user comment dictionaries to update dictionaries regularly. It runs on basic operating systems like Linux and Windows. It has strong scalability and accuracy.

(2) Feature item processing module.

This section mainly calculates feature data and object weights for pre-rendered English text. Features are words that can express specific and unique content of an English text. The selection of attributes is an important part of creating a vector field model, which directly affects the relative value of English words. After the features are removed, the density of the features must be calculated. To replace the words corresponding to the vector field model, it is necessary to add an appropriate weight value to each feature, and add an appropriate weight to each feature. Words and small weight values set for features of the topic indicated by the text are less important.

(3) Similarity calculation module.

The main function of this section is to calculate the similarity of English words and the method to obtain the similarity of English words. First, after using the TF-IDF weight calculation method to calculate the weight value of a feature, we can learn the characteristics of English words, and can measure the similarity of the field vector models of two English words through cosine value calculation. In English text between the two-part vectors of . Then calculate the average similarity of the two English words, and finally measure the average similarity of the two English words and the similarity of the vector field model. Because, to get its size is the average similarity of the two English words and the similarity of the vector field model, that is, the relative value of the two English words.

2.2. Sememe Similarity

During the process of language description, semes are connected to each other to form a tree of positions, in this system, the depth between different points. It is calculated from the depth of each sememe in the tree system and the space between them. The following are two typical sememe similarity analysis methods:

(1)The first sememe similarity calculation formula:

$$Sim(S_1,S_2)=\frac{\alpha}{dis\tan ce(S_1,S_2)+\alpha} \qquad (1)$$

Among them, S1, S2 represent the sememe; distance (S1, S2) is the path length of the two sememes; α is the adjustment parameter, generally 1.6.

(2)The second sememe similarity calculation formula:

$$\text{〖}\mathrm{Sim}(\mathrm{S}_1,\mathrm{S}_2)=\frac{\alpha*\min(\mathrm{depth}_{\mathrm{s}1},\mathrm{depth}_{\mathrm{s}2}}{\alpha*\min(\mathrm{depth}_{\mathrm{s}1},\mathrm{depth}_{\mathrm{s}2})+\mathrm{distance}(\mathrm{S}_1,\mathrm{S}_2)}\text{〗} \qquad (2)$$

Among them: S1, S2 represent the two sememes; depthS1, depthS2 are the depths of the sememes respectively; distance(S1, S2) is the path length; min(depthS1, depthS2) represents the depth of the two sememes; α is a variable Adjustment parameter, generally take 0.5.

2.3. Other similarity algorithms

(1) Concept similarity

The concept of function words is simple to calculate. When calculating, only the grammatical and semantic similarity of the relationship between them is calculated. The real words are more complex and are divided into four parts. The overall similarity formula is as follows:

$$\mathrm{Sim}(C_1, C_2) = \sum_{i=1}^{4} \beta_i \prod_{j=1}^{i} \mathrm{Sim}_j(S_1, S_2) \qquad (3)$$

Among them, C1 and C2 represent two concepts; βi (1 ⩽i ⩽ 4) is an adjustable parameter, generally specified according to experience, and there are β1 + β2 + β3 + β4 = 1 and β1 ⩾β2 ⩾ β3 ⩾ β4 , the conceptual The main feature is described by the first independent sememe descriptor, so its weight is generally above 0.5.

(2) Word similarity

A word can be represented by multiple concepts, and the concept similarity can be obtained by weighting. Assuming that W1 and W2 are two Chinese words, word W1 has m concepts: C11, C12, … , C1m; word W2 has n concepts: C11, C12, …, C1n; then the similarity value of the two words is It can be represented by the maximum value of the similarity in all combinations of C1i and C2j, and the specific formula is as follows:

$$\mathrm{Sim}(W_1, W_2) = \mathrm{Max}(\mathrm{Sim}(C_{1i}, C_{2j}) \qquad (3)$$

$Among\ them{:}\ i = 1, 2, \ldots, n;\ j = 1, 2, \ldots, m;\ Sim(W1, W2) is\ the\ similarity\ value\ between\ the\ two, Sim\ (C1i, C2j)\ is\ the\ similarity\ value\ between\ the\ two\ concepts\ .$

3. Experimental Study on the Application of artificial intelligence machine translation + Post editing in English Chinese Translation

3.1. Simulation of Real Data Collection Experiments

The receiving system signal is a rectangular signal. The signal driver detection phase is timed, receives fixed-period motor acceleration samples with a 10ms sampling time, and sends a digital signal to the control panel to receive the event's driving action. After receiving the information for the search session, the control value is calculated and the result is sent over the network to the active session.

3.2. Predictive Control for Modification in a Broad Sense

The general prediction is corrected by taking the same BP network prediction error as ye (k + j) as compensation

$$y(k + j) = y_m(k + j) + y_e(k + j) \qquad (4)$$

As a traditional general prediction algorithm, y_M(K+J) is responsible for predicting the data at time k. Substitute it into the above formula

$$y(k+j) = G_j(z^{-1})\Delta\mu(k+j-1) + F_j(z^{-1})y(k) + y_e(k+j) \quad (5)$$

Compensation is done, and the best solution can be obtained, as shown below

$$\Delta U = (G^T G + \lambda I)^{-1} G^T (Y - F - Y_e) \quad (6)$$

The dynamic BP network can predict and adjust the GPC error online, changing the load setting state after traditional BP network training.

4. Experimental analysis of teaching transformation of computer application technology specialty based on Artificial Intelligence

4.1. Experimental Model Analysis of the Smith Forecasted Control System

This document connects the default Smith control to the default PID control. and the default ruler is recommended PID. Designed system default Smith Predictive Fuzzy PID control image using default PID control instead of normal control PID. Display shows the test results in the table 1 below.

Table 1. Analysis of smith predictive controller system imitating real experiments

Time(s)	The step response of the system	the PID controller for predicting accurate graphics	the PID control introduced in this paper
1	0	0	0
3	0.42	0.56	0.35
5	0.73	0.98	0.61
7	0.84	1.35	0.71
9	0.91	0.89	0.93
11	0.99	0.97	0.98

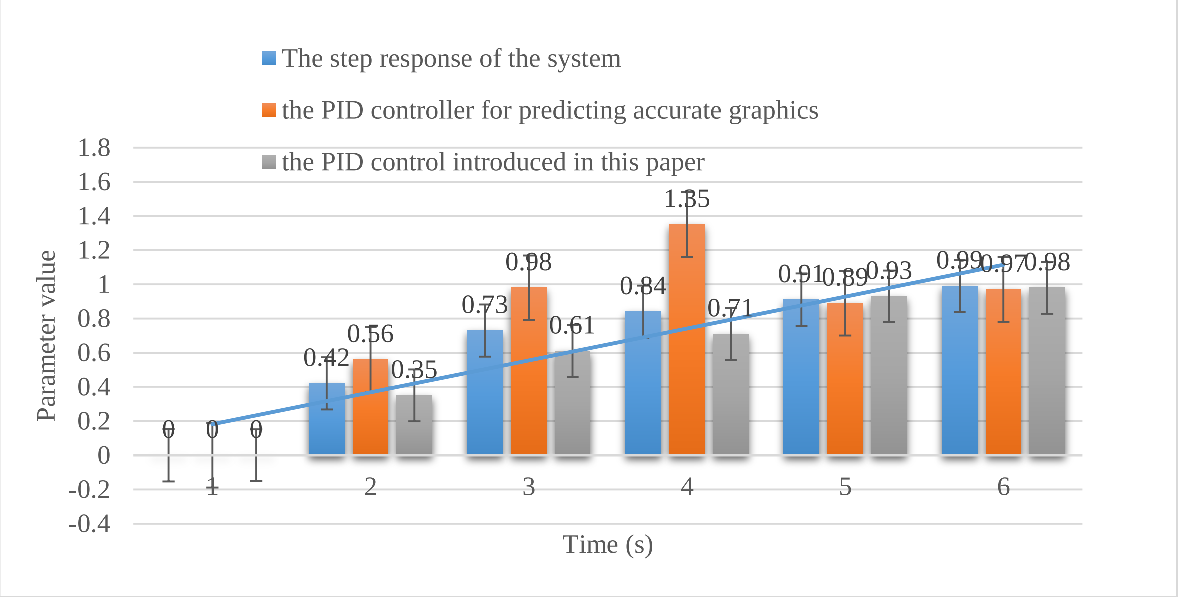

Figure 1. Analysis of smith predictive controller system imitating real experiments

As shown in the figure above, experience has shown that when the system performs well and is clear and accurate, Smith PID predictive control can make the system react more quickly. Smith's preview model is not detailed enough. At present, the former can make the performance of the system more flexible and dynamic, and has strong robustness.

4.2. Dynamic BP Network Simulation Experiment Analysis

In the similar system, the second-order objective transfer function model is selected. On the basis of correcting the dynamic BP network error of the controller, the default general control algorithm is adopted. The time mode of the driver is accepted in the detector, and the contact mode of the driver is accepted in the controller and the controller. The image display test results are shown in Table 2.

Table 2. Data analysis of traditional GPC error and dynamic BP error correction GPC error

Time (s)	Uncorrected GPC	Error-corrected GPC
0	0	0
5	0.46	0.36
10	0.83	0.72
15	1.24	1.21
20	0.91	0.95
25	0.97	1.01

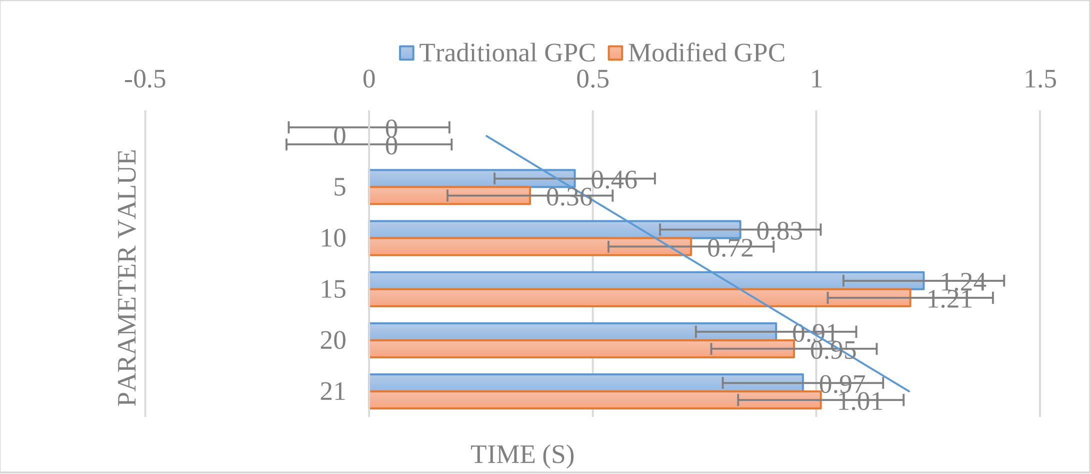

Figure 2. Data analysis of traditional GPC error and dynamic BP error correction GPC error

Specifically, as shown in Figure 2, when the system accesses the network late, the traditional GPC algorithm and the network dynamic BP error correction method based on the GPC algorithm are used. The results show that the control system based on GPC algorithm has faster response time and error correction ability of BP. The dynamic error of the network is improved by 20% compared with the GPC before error correction.

5. Conclusion

In this paper, mainly studies the influence of computer application technology on teaching reform on the basis of artificial intelligence. Machine translation with its translation speed helps translators save a lot of time. However, compared with manual translation, there are still many imperfections in machine translation, which requires manual post editing to varying degrees. This paper first introduces the characteristics of intelligent translation, and then summarizes its related characteristics and application scope,classification and general principles of post translation editing. Then through the study of the application of translation in the project of maintenance instructions and the project of new media text translation, it is found that the common feature of translation in dealing with the two kinds of text is that it can quickly process simple professional information for the lack of professional knowledge Background translators save time to understand the original text. However, there are still some deficiencies in the processing of polysemous words, logical relations and information in machine translation. When fully editing after translation, translators need to refer to professional dictionaries, use network resources and common sense to speculate the commonly used words to replace the improper words in machine translation; analyze the grammatical structure of the original text, understand the differences between Chinese and English expressions, and adjust the sentence structure In order to convey the original meaning, professionalism and readability of the original text, we should fully understand the information of the original text and express it completely in the translation.

Funding

2018 Humanities and Social Sciences Research Program of Jiangxi Province. The Construction and Application of Bilingual Corpus on Jiangxi Tourism Resource under the Background of “Belt and Road Initiative” (YY18220).

Data Availability

Data sharing is not applicable to this article as no new data were created or analysed in this study.

Conflict of Interest

The author states that this article has no conflict of interest.

References

[1] *Fordyce, R. Ewan. Royal Society Te Apārangi and the Pursuit of Research Excellence. Journal of the Royal Society of New Zealand, 2018, 48(2-3):63-63. DOI:10.1080/03036758.2018.1469515*

[2] *Yang, Dong, Jiang, Lihui. Simulation Study on the Natural Ventilation of College Student’ Dormitory. Procedia Engineering, 2017, 205(4):1279-1285. DOI:10.1016/j.proeng.2017.10.378*

[3] *Hoshino, Katsuharu, Inoue, Masafumi. Research on the Role of Carbon Offsets in the Building Construction Work. Transactions of the Materials Research Society of Japan, 2015, 40(1):1-6. DOI:10.14723/tmrsj.40.1*

[4] *Denning, Jeffrey T. College on the Cheap: Consequences of Community College Tuition Reductions. American Economic Journal: Economic Policy, 2017, 9(2):155-188. DOI:10.1257/pol.20150374*

[5] *O, Akhavan, R, et al. Hydrogen-Rich Water for Green Reduction of Graphene Oxide Suspensions – Science Direct. International Journal of Hydrogen Energy, 2015,*

40(16):5553-5560. DOI:10.1016/j.ijhydene.2015.02.106

[6] Gui, Herong, Lin, Manli, Song, Xiaomei. Research on Pore Water and Disaster Prevention in China Coalmines. Water Practice and Technology, 2016, 11(3):531-539. DOI:10.2166/wpt.2016.056

[7] Zhu, Zicheng, Zhang, Xuejun, Wang, Qiang, Chu, Weijun. Research and Experiment of Thermal Water De-Icing Device. Transactions of the Canadian Society for Mechanical Engineering, 2015, 39(4):783-788. DOI:10.1139/tcsme-2015-0062

[8] He, Zhen. A New Era of Water Environment Research. Water Environment Research, 2019, 91(1):3-4. DOI:10.1002/wer.1025

[9] Zhao, Yong, Zhu, Yongnan, Lin, Zhaohui, Wang, Jianhua, He, Guohua, Li, Haihong, Li, Lei, Wang, Hao, Jiang, Shan, He, Fan, Zhai, Jiaqi, Wang, Lizhen, Wang, Qingming. Energy Reduction Effect of the South-to-North Water Diversion Project in China. Scientific Reports, 2017, 7(1):15956. DOI:10.1038/s41598-017-16157-z

[10] Colakoglu, Mert, Tanbay, Tayfun, Durmayaz, Ahmet, Sogut, Oguz Salim. Effect of Heat Leakage on the Performance of a Twin-Spool Turbofan Engine. International Journal of Exergy, 2016, 19(2): 173. DOI:10.1504/IJEX.2016.075604