Research on the Design of Computer Scoring System for Chinese College Students' English Translation

Qin Wang¹, Fang Pan¹,*

¹Guilin University of Electronic Technology (Beihai Campus)  Beihai, Guangxi 536000

*Corresponding author e-mail: panfang@gliet.edu.cn

Abstract. In the current field of Chinese institute students' English computer translation, for the subjective questions in the form of question and answer, the answer is often open and non unique, which makes it difficult for the computer system to accurately identify and judge the answer. The score result still has a large drift from the manual score, so the fitting degree and prediction rate need to be further improved. Based on this, this paper first analyzes the research status and progress of English translation computer scoring, and then studies the construction of the computer scoring system of English translation grammar for Chinese institute students, and gives the logical linear model, the algorithm of feature extraction of translated text and the construction process of the model.

Keywords: Computer Scoring System, Institute Students, English Translation

1. Introduction

With the iterative development of computer tech represented by AI algorithm, it has been widely and deeply studied and applied in many fields, especially in the field of translation represented by English translation, which greatly promotes the double improvement of translation efficiency and translation quality [1]. At present, English, as a necessary bridge and tool for cross-border communication and trade, plays an important role in promoting the process of global integration. In this context, how to improve the accuracy and efficiency of English translation has become an important part of current institute English translation teaching and research. As an important part of the evaluation of Institute Students' English translation results, the accuracy of their translation scoring system has become the focus of improving the efficiency of Institute English translation teaching.

Institute English translation computer scoring system can significantly reduce the workload of teachers' review and correction, so that teachers have more energy to carry out academic research and improve the teaching curriculum. Moreover, the computer automatic scoring system can also significantly reduce the interference of human factors and the influence of supervisor factors, so as to improve the objectivity and scientificity of Institute Students' English translation scoring. Therefore, the Institute English translation computer scoring system has been gradually promoted and applied, and has been applied in some large-scale examination scoring. The current English translation computer scoring systems are mainly as shown in Figure 1. Although these computer scoring systems
have their own characteristics and applicable scenarios, the consistency and efficiency of these systems still need to be further improved.

![Diagram of current English translation computer scoring systems.](image)

**Figure 1.** Current English translation computer scoring systems.

In addition, with the continuous iteration of computer tech represented by data processing and AI algorithm, the stability and accuracy of its application in institute students' English translation scoring are also constantly improved [2]. The computer scoring system reduces teachers' low price, high repetitive labor intensity and dependence on the participation of teachers' supervisors. The current computer scoring system has higher recognition and accuracy for the objective questions which are mainly choice. However, for the subjective questions in the form of question and answer, due to the openness and non uniqueness of the answers, it is difficult for the computer system to accurately identify and judge the answers. AI algorithm and deep learning are needed to help the computer simulate the operation process of human brain, so as to improve the efficiency and accuracy of recognition and judgment.

The scoring of Institute Students' English translation belongs to the category of natural language processing, so its research and design are in many disciplines, such as language processing, computer, mathematics and data analysis. Because the computer is difficult to analyze and understand the real scene, there are big problems in the process of natural language processing, resulting in a large error rate in the scoring results. In the field of Institute Students' English computer translation, the score results still have a large drift from the manual score, so it is necessary to further improve the fitting degree and prediction rate. Therefore, it is of great practical value to study the computer scoring of Institute Students' English translation.

2. Research status and development of computer scoring in English translation

2.1. Research status of computer scoring in English translation

At present, English translation computer scoring analyzer based on Chinese syntax depicts the relationship between each node in the tree, especially the dominating and dominated relationship between the head word and its adjacent nodes [3]. Dependency tree based parsing can provide useful structural information for machine translation. The translation scoring system of skeleton dependency tree only analyzes the overall syntactic structure of a sentence, which is represented by the head word and its direct dominating components. Skeleton dependency tree considers the whole structure information and lexical semantic information of the sentence, and avoids complex analysis of the sentence.

In addition, at present, there is no effective method to automatically evaluate and grade long sentences in the automatic scoring system for English translation. The evaluation and processing of long sentences is still focused on the segmentation of long sentences. The scoring framework of English translation based on backbone association analysis is shown in Figure 2 below.
Figure 2. The scoring framework of English translation based on backbone association.

In the recognition level of base NP, as the basic task of natural language processing, base NP contains more abundant language information and less ambiguity than single word; compared with parsing, base NP has less workload and more practical application value. Base NP is a simple, non-nested noun phrase, which does not contain other subitem phrases [4]. Therefore, it is very convenient to be recognized and scored by computer. Based on computer tech, a relatively complete large-scale base NP training test corpus is established to facilitate the computer to compare with the training test corpus in the process of translation scoring, so as to further improve the accuracy of scoring. The scoring results based on the process of base NP algorithm are shown in Figure 3 below.

Figure 3. Base NP chunking model.

2.2. Research progress of computer scoring in English translation

At present, the research directions of Chinese institute students’ English translation computer scoring are mainly focused on parsing / syntax based SMT, long sentence analysis / backbone association analysis, Chinese base NP and Chinese named entity [5]. In this process, the module of Chinese word segmentation and named entity recognition is updated to reduce the noise and improve the module of phrase extraction and probability calculation. In the translation scoring system based on Chinese syntactic analyzer, the head driven parser and the translation based on skeleton dependency concept are implemented. In the automatic recognition level of digital time, WFST recognition is used to solve
the establishment of an effective digital time phrase state machine and the matching algorithm of
digital time string under a given finite state machine.

3. Computer scoring system for English translation grammar

3.1. Logical linear model
For a given Chinese source sentence $f$, students translate the target language into English $e$, then the
maximum probability of all possible target sentences is:

$$
\hat{e}_1 = \arg \max_{e_1} \{ P_r(e_1^{f1}, f_1^{f1}) \}
$$

According to the maximum entropy theory, the direct probability model is:

$$
P_r(e_1^{f1}, f_1^{f1}) = \frac{\exp \left[ \sum_{a=1}^{u} \lambda_a h_a(e_1^{f1}, f_1^{f1}) \right]}{\sum_{e_1'} \exp \left[ \sum_{a=1}^{u} \lambda_a h_a(e_1', f_1') \right]}
$$

3.1.1. Application of GMM in force sensing field
The deep processing syntactic feature function is formed by two models: the feature function with the
best alignment and the feature function of tree to string [6]. Among them, the feature function of tree
to string, Chinese sentence and English analysis tree, and the feature function of alignment probability
sum are as follows:

$$
h_{\text{TreeToStringSum}}(e, f) = \log \sum_{\Theta} \prod_{\Theta(e_{i,j})} p(\Theta_i^{e_j} \mid e_i^{f_j})
$$

$$
h_{\text{TreeToStringViterbi}}(e, f) = \log (\max_{\Theta} \prod_{\Theta(e_{i,j})} p(\Theta_i^{e_j} \mid e_i^{f_j}))
$$

3.1.2. Characteristic function of tree to tree
The analysis trees of two languages with the same meaning have different structures. Some of them
can be automatically converted and aligned, while others can not be automatically aligned and
transformed. In this paper, we give a special feature function at each level, and get the theoretical
framework of computer scoring of English translation.

3.2. Feature extraction algorithm of translated text
First, in the level of data collection and scoring standards, it mainly includes data collection and
processing, the determination of scoring standards, text characteristics classification and other major
components. Secondly, in the semantic basic feature extraction level, the Bleu value extraction
algorithm is mainly included to achieve better results in the translation computer scoring. The
algorithm of extracting the semantic similarity of segmented translation is also included, so that the
relationship between similar fractional translation can be fully utilized. In addition, at the level of
comparison between Bleu value and sub translation, we need to find the text similarity, through
comparing the student translation with the reference translation, and the sub translation latent semantic
similarity. At the level of feature extraction of matching rate of test points, it is necessary to
automatically get the meaning of words that have not been found. Finally, for the extraction of
coherence features, it is necessary to collect the coherence words and phrases to form the coherent vocabulary at the algorithm design level.

3.3. Construction of computer scoring model for English translation

First of all, at the level of constructing corpus and formulating scoring criteria, we need to construct corpus and formulate scoring criteria, and define the probability of its occurrence in the corpus as an indicator of grammatical errors. Secondly, at the level of similarity and distance, we need to measure the similarity of strings, especially the length of the longest common subsequence. In addition, in the process of constructing the n-gram scoring model, it is necessary to obtain the coefficient value from the training set, and then verify the Bleu scoring method on the test set.

In the process of constructing the computer scoring model for English translation, the corpus input module is used to input the translation of the training set collected, and then the operational corpus file is output after text processing. Then, the text feature values of each training set translation are calculated for the corpus files. Finally, all text eigenvalues and scores are formatted and output to the file as the input data of neural network. Based on the input student translation, the score of student translation is obtained by using the stored model data and feedback module.

4. Conclusion

In summary, the computer scoring system reduces the low price and high repetitive labor intensity of teachers in the process of English translation scoring. The computer scoring system of Chinese institute students' English translation can significantly reduce the workload of teachers' review and correction, thus reducing the dependence on the participation of teachers' supervisors. In the field of Chinese institute students' English computer translation, the score results still have a large drift from the manual score, so it is necessary to further improve the fitting degree and prediction rate. This paper analyzes the current situation and development of computer scoring system for English translation. And then constructs a scoring model of English translation, and finally constructs a computer-based translation model to analyze the translation process.

Acknowledgment

1. Undergraduate Education Reform Project of Guangxi Higher Education in 2020 (2020JGA182);
2. The second batch of Production-learning-Researching Cooperative Education Project of the Ministry of Education in 2019 (201902079011).

References

[1] Chen Guimin, Jia Jianyuan, Han Qi. Research on inertia weight decreasing strategy of particle swarm optimization algorithm in Xi'an Jiaotong[J]. Journal of Xian Jiaotong University, 2016, 27 (5): 24 - 27.
[2] Liang Maocheng, Wen Qiufang. Review and Enlightenment of foreign automatic scoring system [J]. Audio visual foreign language teaching, 2017 (5): 18-24.
[3] Wang Jinfu. Development and application of automatic translation scoring system at home and abroad in the past 50 years[J]. Journal of Yangzhou University, 2016, 20 (2): 123-128.
[4] Wang Jinquan, Wen Qiufang. A review of automatic scoring systems at home and abroad -- and the enlightenment for Chinese students [J]. Foreign languages. 2010 (1): 75-81.
[5] Wang Jinquan, Wen Qiufang. Formulation of analytical scoring rules for learners' Chinese English translation. Foreign language teaching, 2019, 30 (4): 15 - 18.
[6] Wen Qiufang. Application of bilingual alignment tech in automatic scoring of English test translation [J]. Audio visual foreign language teaching, 2019 (1): 3-8.