Design of English Translation Computer Intelligent Scoring System Based on Natural Language Processing

Hong Yang¹, Yan Yang²

¹Foreign Language School, Xichang University, China, 615000
²Organization Department, Xichang University, China, 615000

*Corresponding author e-mail: 416853478@qq.com

Abstract. With the development and maturity of automatic scoring technology of English composition, the computer-aided composition marking system based on automatic scoring technology has begun to enter colleges and universities to assist English writing teaching. However, there are still many problems in the current scoring system, such as long running time and large deviation of scoring results, which requires us to design an English translation computer intelligent scoring system based on natural language processing. Through the system, we can reduce the workload of manual scoring, which will improve the efficiency of scoring. Therefore, we need to construct the structure of English translation scoring system, including translation data collection module, information feature extraction module, analysis model construction module and result feedback scoring module. By building a language model, the system can translate the probability distribution of specific sentences or word sequences. As an adaptive learning model, BP neural network has more advantages in dealing with the relationship between complex variables. Through BP network model, the system can extract the feature information of translation translation and translation training set. Through fitting calculation, the system will realize the intelligent scoring of English translation.

Keywords: Natural Language Processing, English Translation, Intelligent Scoring System, BP Network Model

1. Introduction

The process of economic globalization further brings about the globalization of language. English has become the only leading role of language globalization, and English learning has gradually become the content of people's attention. In college curriculum, English education is a very important part. In order to test the stage level of College Students' English learning, there will be corresponding examinations at the end of the University. The Ministry of education also regularly holds large-scale CET-4 and CET-6¹¹. These examinations can promote English teaching and provide feedback, but the examination paper marking work consumes a lot of manpower and material resources. In many cases, through computer technology, machine automation process has replaced the tedious and meaningless repetitive labor. In
the field of education, in order to reduce unnecessary marking work, the research of computer automatic marking is becoming more and more mature\cite{2}.

Natural language processing (NLP) is an important field formed by the integration of computer science and artificial intelligence, which involves many disciplines, including linguistics, computer science and mathematics. Therefore, NLP makes it possible for computers to process human languages. The fundamental reason for the difficulty of natural language processing is that the computer can’t understand the language in combination with the scene, which leads to a large number of polysemy or ambiguity. In recent years, natural language processing based on statistics, rules and corpus has achieved good results.

Based on the great progress made in natural language processing, many automatic scoring systems for English composition have been developed abroad, such as peg, tea, e-rater, which have been applied to GRE, GMAT and other large-scale examinations. By extracting the text feature values, we can use the linear regression method to construct the composition scoring model, which makes the machine scoring and manual scoring achieve a high correlation\cite{3}.

2. Basic theory of artificial neural networks

2.1. Overview of neuron model

The basic structure of the nervous system is the neuron nerve cell, which is the basic unit to process human information transmission. The brain can handle extremely complex analysis and reasoning. On the one hand, the number of neurons is huge, on the other hand, neurons can process input signals non-linearly. Therefore, the establishment of engineering neuron mathematical model can better study engineering problems, as shown in Figure 1. The model is a multi-input single-output nonlinear element. The weights represent the connection strength between neurons.

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{neuron_model.png}
\caption{Mathematical model of neurons}
\end{figure}

Mathematical model expression of neuron:

\[ y_j = f(\sum_{i=1}^{p} w_{jp} x_p - \theta) \]  \hspace{1cm} (1)

Among them, \( w_{jp} \) is Connection Weight Coefficient; \( x_p \) is the input signal at the input end; \( \theta \) is the neuron threshold; \( p \) is the number of input signals; \( y_j \) is the output of neuron at the j-time; \( f(\cdot) \) is the excitation function\cite{4}.

2.2. BP model
BP model is also called Back Propagation Network and Multilayer Forward Network. The structure of BP model is shown in Figure 2.

![BP model structure](image)

**Figure 2.** The BP model structure

2.3. *Algorithm process based on BP network model*

The algorithm process based on BP network model is shown in Figure 3.

![Algorithm process](image)

**Figure 3.** The algorithm process based on BP network model

Feasibility analysis of computer intelligent scoring system for English translation.

3. *Feasibility analysis of computer intelligent scoring system for English translation*

3.1. *Technical feasibility*

Based on ASP. Net online examination, the key technology lies in the dynamic display and management of the web page, which shows the corresponding test paper data from the database. By getting the user's input data, we can control the whole examination process. Through the latest ASP. Net and SQL technology development, we can use the large database SQL Server 2005, which provides the ability of
database management. Therefore, the technical scheme of English translation computer intelligent scoring system is mature and feasible\cite{5}.

3.2. Economic feasibility

The traditional theory examination needs a large number of paper printed test questions, which should be supervised by the teaching affairs office and broadcast listening materials. After the examination, teachers spend a lot of time on marking papers and counting scores. Generally, the development time of online examination system is about 2-3 months. The current computer configuration can bear the development of the required hardware and software facilities, and the cost is not high. At present, most schools have high-performance computers and local area networks. The installation, deployment, operation and maintenance of the computer intelligent scoring system is feasible, which will not increase the cost of the school.

3.3. Operational feasibility

At present, most PC and LAN can run the system, the original computer system settings and network layout will not change because of the system installation, debugging or operation, and most users do not need any training can easily operate the software. Therefore, the operation of English translation computer intelligent scoring system is feasible.

4. System design

4.1. System structure level

The whole translation scoring system includes four modules: corpus input module, feature extraction module, model building module and feedback scoring module\cite{6}. In the early stage of the system, a corpus input module is needed to input the student's translation. After the module is processed, a formatted corpus file is generated. The feature extraction module analyzes the corpus file and extracts the features from the corpus by using the corresponding extraction algorithm. The model building module uses these features to build scoring model and generate corresponding model files. With the corresponding model, the feedback evaluation module can score the translation and comment feedback. The relationship is shown in Figure 4.

![Figure 4. System structure level](image)

4.2. Feature extraction module

This module further extracts the features of the corpus files input by the corpus input module. First, the corpus file is parsed and loaded into the memory space according to the specified format. Second, the feature extraction of language basic points is divided into word error number, grammar error number and article length. Thirdly, feature extraction is carried out on semantic feature points. Semantic feature
points are divided into surface semantics and hidden layer semantics. Surface semantics is realized by Bleu algorithm, while hidden layer semantics requires knowledge related to latent semantics. Fourth, establish test point file, summarize the semantic points and show them in the form of files. Fifthly, the obtained eigenvalues and teacher's score are stored in the form of file, which is prepared for the model building module.

4.3. Translation system flow

The flow of the translation system is shown in Figure 5.

![Figure 5. The flow of the translation system](image)

5. Conclusion

This paper designs a computer intelligent scoring system for English translation based on natural language processing. After building the hierarchical structure of the system, this paper designs the modules of each level. Combined with the operation process of the system module, we get the overall framework of the translation scoring system. By adopting language model and other related algorithms, this paper realizes the utility of the corresponding module.

References

[1] Qu Weiguang. Summary of AMR research on abstract semantic representation of natural language sentences [J]. Data acquisition and processing, 2017, 32: 26-36.

[2] Chen Zhiming, Li Maoxi, Wang Mingming. Sentence level translation quality estimation based on neural network features [J]. Computer research and development, 2017, 54: 1804-1812.

[3] Xie Di fan, Du Zifang. Improvement of Chinese film online rating system [J]. Computer applications, 2018, 38 (4): 1218-1222.

[4] Wen Qufang, Qin Ying, Jiang Jinlin. Application of bilingual alignment technology in automatic scoring of English test translation [J]. Foreign language audio visual teaching, 2019, 1:3-8.

[5] Pang Xiuhua. The clever use of famous aphorisms in English composition [J]. English for middle school students: foreign language teaching and research, 2014, 5: 122-122.

[6] Liang Maocheng. A review of foreign automatic scoring system for composition and Its Enlightenment [J]. Audio visual foreign language teaching, 2017, 5: 18-24.