

Article

The Comprehensive Training Effect of Translation Ability of College English Majors Based on Machine Learning

Donghua Li^{1,*}

¹ School of English Language and Culture, Xi'an Fanyi University, Xi'an 710105, Shaanxi province, China

* **Correspondence:** lidonghua8010@163.com

Abstract: With the increasingly frequent exchanges between countries, my country's demand for high-quality English translators has greatly increased. However, an important problem we are currently facing is that China's translation talents are far behind the demand. An important reason for this phenomenon is that the traditional translation teaching is difficult to cultivate translators who can meet the market demand. Therefore, it is necessary to improve the traditional translation teaching mode. Translation teaching for English majors is an important part of translation teaching. Therefore, after evaluating the speech characteristics and speech data, this document first proposes a translation classification error detection model based on mfcc-rf. The acoustic function of the extracted 39 dimensional Mel inverse spectral coefficient is the input of the random forest classifier, and a classification error detection model is established. By analyzing the experimental results, the MFCC radio frequency translation error detection model has achieved high classification error detection accuracy under three types of errors (rising, falling and shortening). The experimental results show that, with semantic similarity as the design principle of distractors, using the word vector training method of the context word prediction model to automatically generate distractors can ultimately improve the comprehensive training efficiency of college English majors' translation ability.

Keywords: Machine learning, College English, Translation ability, Intelligent algorithm optimization

1. Introduction

The motivation for learning translation ability of college English majors has always been an important topic in the field of second language acquisition, which has a significant impact on the learning effect of learners, and can provide learners with motivation and maintain long-term learning behavior [1,2]. With the acceleration of globalization, the traditional motivation theory has undergone continuous iteration and improvement. In this developmental process, the concepts of future self-image, ideal second-language self and vision were proposed from the perspectives of personality psychology and motivational psychology [3,4]. Oriented motivational flow is a super-motivational state that emerges during a specific period of time in order to achieve a given goal and vision. After the learner activates the directional motivation flow, it helps the learner to overcome the possible emotional fluctuations by experiencing positive emotions in the process of completing the goal, so that the learner continues to produce high-intensity motivational behaviors [5]. In classroom teaching, activating

students' directional motivation flow can help improve students' learning effect, while teaching intervention can generate strong directional motivation flow, which is conducive to improving the effect of second language classroom teaching. Output-oriented method can be used as a teaching intervention method to help Students activate directed motivational flow [6]. It has always been a common problem for English learners that they dare not speak, and they are afraid of making mistakes. Teachers cannot accurately correct students' one-to-many teaching. English translation [7, 8].

On the translation problem of learners, the influence of the native phonetic system is mainly manifested in the non-standard translation-related organs or translation actions, and there will be large differences in phoneme discrimination [9, 10]. Learning itself can only be translated through the tradition of blind obedience, and can not understand the correct translation methods, so it can not find pronunciation errors in time. Therefore, sometimes learners do not even know whether translation is standard. In order to ensure accurate translation, many people pay high fees and require foreign teachers to correct translation errors [11]. In addition to learning languages online, automatic translation can also find errors and correct system errors at historical moments. At present, there are few translation products in the market, most of which are relatively simple work [12, 13].

To sum up, how learners can discover their own English translation problems freely, in real time and efficiently and correct them in a targeted manner becomes the research significance of the machine learning automatic translation error detection and correction system.

The translation for automatic error detection comes from the computer supported language training system (Canada) [14]. Computer supported translation is an important part of the computer supported language training system. The main components of the cap system are to evaluate English, find translation errors and correct feedback errors. The two main components of cap are the discovery of automatic translation errors and the correction of feedback errors. However, the historical development is not long, mainly after the 1990s [15]. Research on machine translation errors found in some literatures, supplement to VQ carrier and Research on synchronous Mechanics (DTA) [16].

Used to detect possible translation errors when a learner is learning Norwegian given L1. Similarly, in 2005, for native Vietnamese learners, a research team studied common translation phonemes and translation substitution errors between Vietnamese and English [17]. For example, /TH/ is incorrectly replaced by /t'/; /S/ is replaced by /SH/; /JH/ is replaced by /Z/, /CH/, /T/, /Y/, /ZH/, /S /; /CH/ is replaced by /S/, /Z/, /ZH/; and /ZH/ is replaced by /Z/, /CH/, /JH/.

In this study, we find that the translation error detection method based on language and features has a good advantage in finding typical errors among language learners in different families. As mentioned above, this is mainly due to the differences in translation between regions and languages, which is a typical mistake of second language learners [18]. This type of error is very limited, so most of the work using this method to detect translation problems needs to create an error type library to merge the possible error types in the library. However, due to unstable translation, translation errors based on unique advantages cannot handle all types of translation errors. If the type of translation error does not belong to the translation category specified by the enterprise, there will be library type translation errors, and the system cannot detect serious translation errors [17]. Therefore, translation error detection based on language and features is not universally applicable to all public translation errors. Therefore, it is more appropriate to identify the wrong translation identifier in the list than the identification method.

In China, as a leading country in the field of information technology, artificial intelligence is the most active in theoretical research and educational practice, laying a theoretical and practical foundation for China's digital education and information management. Especially in the field of educational technology in China, scientists have gradually applied artificial intelligence to education, and the research from theoretical research to technological development has become the focus. Therefore, the research results in recent years are very important. The most important achievement is the establishment of an "intelligent education system". Its main purpose is to help math education. Use

information technology as a tool to educate students on the sense of three-dimensional geometric space and improve mathematics teaching.

In short, the method of using logistic regression model on learning machine, the method of strengthening decision tree, the method of combining two student prediction models into English translation, and the analysis of factors affecting students' English translation ability according to the prediction results are explained. At the same time, this has become the focus of this study.

2. Machine Learning and Corpus Coverage

With the development of automatic speech recognition technology, CAPT using automatic speech recognition (ASR) in language learning has received considerable attention. Automatic translation error detection at the phoneme level can usually be done by computing confidences derived from hidden Markov model-based speech recognition systems. A Hidden Markov Model is used to train the native language translation data, and enforced alignment is performed on each phoneme level, and then the GOP score of the translation at the phoneme level is obtained. The GOP score of each phoneme corresponds to the normalized phoneme and the recognized phoneme frame Log-likelihood ratio. GOP is specifically designed to score and detect translation errors in spoken language translations. The learner's translation is first aligned to the phoneme level, and the posterior probability of the phoneme x is calculated by observing the sample O , which can be approximated as the following formula.

$$\text{GOP}(x, O) = P(x, O) \approx \log \left(\frac{p(olx)}{\max_{y \in Q} P(oly)} \right), \quad (1)$$

where Q is the total number of phonemes in the phoneme bank. Distribution of correct translations and incorrect GOP scores obtained by observation. GOP thresholds for error detection can be obtained.

2.1. Corpus Coverage

The work and research related to natural language processing has gradually increased, such as speech recognition, speech evaluation, speaker recognition, etc., and the accuracy rate of recognition technology is also increasing compared to the past. The higher the level, as the basis of these technologies, the demand for corpus is also increasing, and a good corpus can make work or research more effective. The corpus usually not only contains voice files and text files corresponding to voices, but also has a lot of work on the storage of information such as gender, age, and dialect of the recorder. In order to ensure the quality of the speech data stored in the corpus, the construction of a corpus generally needs to meet certain specifications: including the specifications of the recorder, data collection, data transcription, data screening, and corpus annotation. The characteristics of vocabulary, syntax, semantics and psycholinguistics can reflect the difficulty of sentences in different degrees. Sentence difficulty assessment shows the following characteristics in feature selection. First of all, due to different research purposes, service objects, sentence languages and fields, the selected features of sentence difficulty assessment are different, but the vocabulary and syntactic features are basically taken into account. Second, the number of secondary features used was different from study to study. Some studies use as many as 151 features, and some only use one feature and can achieve better difficulty prediction results. Finally, feature selection shows a trend of combining some theories and features of other disciplines such as psychology and linguistics as shown in Table 1.

Corpus	863 corpus	THCHS30	863+ THCHS30
Number of sentences	1500	1000	2500
Biphonic coverage	58.5%	71.6%	73.5%
Tri phoneme coverage	7.0%	14.4%	16.7%

Table 1. Common Distributed and Parallel Machine Learning Algorithms and their Characteristics

As shown in Figure 1, the density of the low-frequency distribution is higher than that of the high-frequency distribution. This is different from the sensitivity and frequency of human hearing and is plotted on cepstrum.

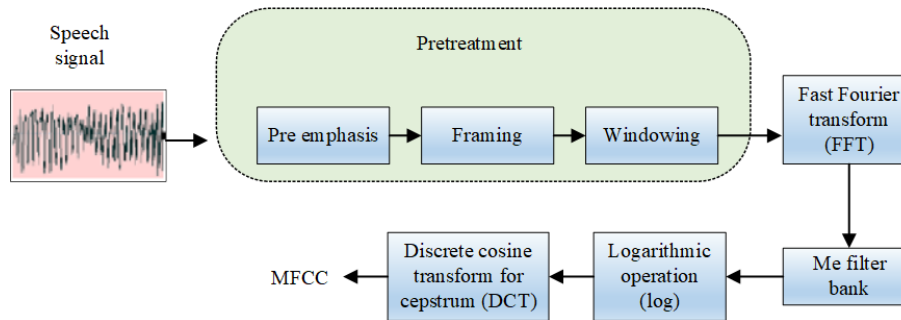


Figure 1. Flow Chart of Mel Cepstral Coefficient Extraction

Language is a continuous signal of time, so the two frames before and after can better distinguish the static state of MFC. Mechanics is a function that can exist at the same time, which better reflects the continuity of language. Generally, the feature dimension is differentiated. to continuously the speech information of the previous frame and the next frame. The research on automatic translation error detection system in this paper is based on the acoustic characteristics of MFCC and its first- and second-order difference coefficients.

To sum up, in the study of sentence difficulty assessment, there are more studies on relative difficulty assessment, but less on absolute difficulty assessment. Most of the studies are carried out in the context of text simplification. There is a lack of sentence difficulty assessment research that can reflect the needs of different users and reflect personalized application scenarios. The method based on classification is the most used method at present, followed by the method based on ranking. With the integration of disciplines, many researchers have begun to try to use the method of psycholinguistic features. Feature selection is very important to the performance of sentence difficulty evaluation. There is no absolute value in the number of features. The selection and number of features should be considered in combination with the research purpose, application scenario and sentence language.

2.2. Deep Belief Networks

The deep belief network used in this paper is a network structure composed of restricted Boltzmann machines proposed by Dr. Fisher, first proposed by Geoffrey Hinton of the University of Toronto. At present, DBN has been widely used in the research of image, speech, medical and other industries. However, there are few precedents in translation error detection, which is the main reason why this paper chooses DBN for translation classification error detection modeling.

Suppose the visible unit has V neurons and the hidden unit has H neurons. Its energy function can be defined as:

$$E(v, h, \theta) = \sum_{i=1}^V \sum_{j=1}^H w_{ij}v_ih_j - \sum_{i=1}^V b_i v_i - \sum_{j=1}^H a_j h_j. \tag{2}$$

The marginal probability of the node set V in the visible layer of the model can be obtained as:

$$P(v; \theta) = \frac{\sum_h \exp(v, h, \theta)}{2} \quad (3)$$

Neural networks can learn by themselves and compute the joint probability distribution density between features. The resulting features can be used as input to learn a classifier, as shown in Figure 2.

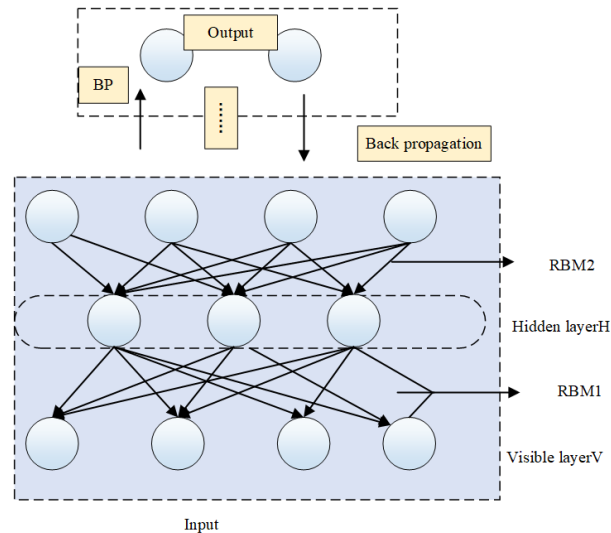


Figure 2. Example of Deep Belief Network Structure

Through the pre-training of the first step, the initial parameter matrix is obtained, but since the forward training of the deep belief network is unsupervised, the parameter matrix at this time is not optimal. In order to make the training model have better learning ability, in the second step, supervised error learning is performed from the top to the bottom, and the parameter matrix of the model is adjusted. During the reverse fine-tuning process, the RBM generation weight is adjusted by the gradient descent algorithm. According to the top-level state and the generation weight, the bottom-level state is generated, and the cognitive weight of each RBM is adjusted at the same time, and error learning is carried out in two directions, and the weight is optimized to obtain the optimal model.

Machine learning is expressed in the simplest language, that is, the machine calls the existing instance in the current execution process, gives the last result, and records each running process and result, so as to quickly give the current execution result. As shown in the figure is the basic structure diagram of a learning system, which mainly represents the relationship between the main elements in the learning system. The four elements of environment, learning, knowledge base and execution form a relationship.

In this system, the system environment affects the learning function. After machine learning, the existing knowledge base will be changed, and the execution is based on the knowledge base. When the machine is executed, this system operation is an experience and a learning process of the system. Through the circular guidance of learning and execution, the system will have a more powerful knowledge base, thereby enhancing the self-learning ability and judgment ability of the system. As shown in Figure 3.

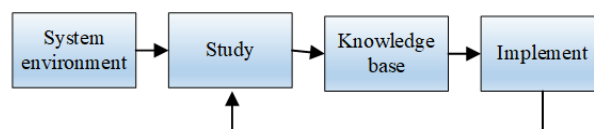


Figure 3. Basic Structure of Machine Learning System

Therefore, in the field of language application, why is the focus of this document not on applying English education in secondary schools or other subjects, but on exploring English schools? (William Shakespeare, Winston, language, language, language, language, language, language, language, language, language, language, language) the most important factor is that the dilemma of high quality and high efficiency faced by English schools requires the technical innovation force to lead the maze. The intelligence in language learning can make up for many defects of English schools, not only curriculum English schools. It can also change the classroom teaching methods of English schools and improve the students' learning styles.

3. Methods

College English classrooms have an urgent need for an interactive teaching environment, because the college English learning environment is not comparable to that of universities, and many universities have more conditions and opportunities to create an interactive English environment through English corners, foreign teachers, exchange students, foreign company internships and many other channels, but these interactive environments are not available inside or outside the university classroom. In the traditional classroom of context, it is difficult for students to understand English, and English dialogue is smooth. English is also the core of the problem. However, the technical support of the conversational English college can provide a learning environment. You can manually process information through computer technology, text, graphics, images, audio and other media. In an intelligent system, you can logically connect the whole. The biggest advantage of this intelligent system is that it can provide a variety of Interactive English teaching methods. You can perform interactive English training methods at the same time, including voice interaction, text and graphical interface, so as to obtain appropriate results in the training. With the help of human-computer interface, students can interact with intelligent machines and position themselves in an almost real English environment. Even if these errors can be corrected timely and accurately, it can also give students the motivation to speak boldly and correct errors, and change the current situation of "English voice". The interactive teaching provided is shown in Figure 4.

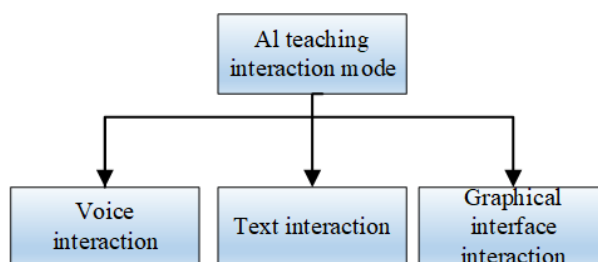


Figure 4. Interaction Mode of College English Translation

English teacher training classes can book courses and upload videos through the system interface. Teachers can use mining technology to evaluate students in English tests, analyze and evaluate students' knowledge system and overall language ability, provide teaching reference materials, etc., and evaluate the overall situation of English tests. Its SAIES module relationship diagram is shown in Figure 5.

Second, students can view knowledge points through video or text, review and apply the acquired knowledge. Third, students can set up units in the system to conduct self-test and learn their own English knowledge system. The purpose of mastering or mastering the basic knowledge including four kinds of English reading, listening, writing and dialogue is to comprehensively learn the students' listening, dialogue, reading and skills. Finally, after the student completes the practice, relevant scores and learning suggestions will be automatically generated. In addition, teachers can use this unit to conduct written tests for students and evaluate their practical knowledge. Because the

examination system is a database based on a large number of scientific questions. The research on the automatic generation of test questions is an effective way to solve the problem of human and financial consumption of manual proposition. The personalized test in this study means that the test is "tailored" for learners, that is, the test can meet the needs of learners with different language levels. The measurement of text readability is of great significance for intelligent adaptation, text simplification, computer-aided language learning system and other applications. To realize the automatic generation of personalized vocabulary test questions, it is necessary to measure the difficulty of sentence level first. On the basis of previous studies, we extracted six features (sentence length, minimum word frequency, syntactic depth, word meaning sum, dependency chain length, and language model probability) from three levels of vocabulary, syntax and semantics to measure the difficulty of sentences, and verified the effectiveness of these features through experiments. During this process, the generalized architecture is shown in Figure 6.

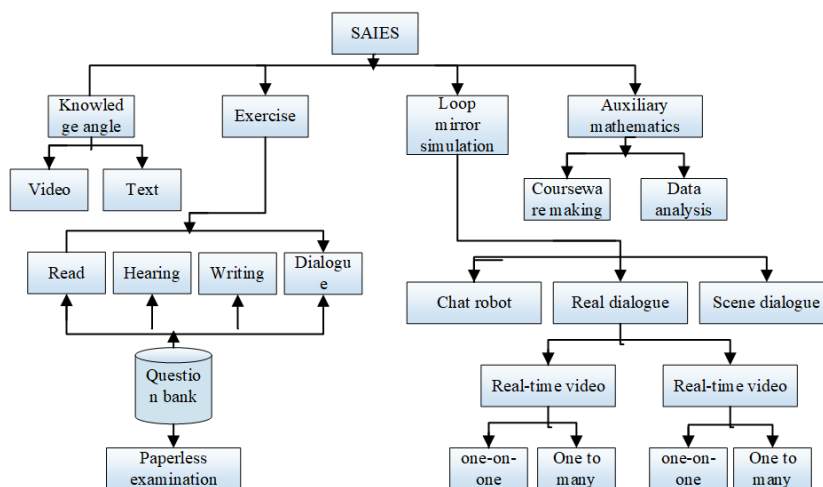


Figure 5. SAIES Module Relationship Diagram

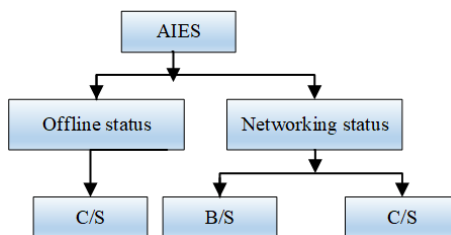


Figure 6. Aies Architecture

4. Case study

At this stage, the entire data mining work has been completed, and the generated models can be used online or used to solve more complex problems. The five stages in the data mining process are shown in Figure 7.

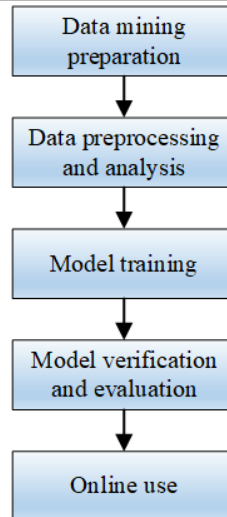


Figure 7. Five Stages of Data Mining of Translation Ability of College English Majors

4.1. Logistic regression ability improvement

Logistics regression is to arrange the amount obtained by linear regression as a function to obtain a digit (0,1) below the decimal point. If the number is above the specified threshold, it is considered positive and negative when the threshold is low.

$$g(z) = \frac{1}{1 + e^{-wx}}. \quad (4)$$

They are model parameters or slopes. The result training mode is the appearance of input point samples and s-image functions as shown in Figure 8.

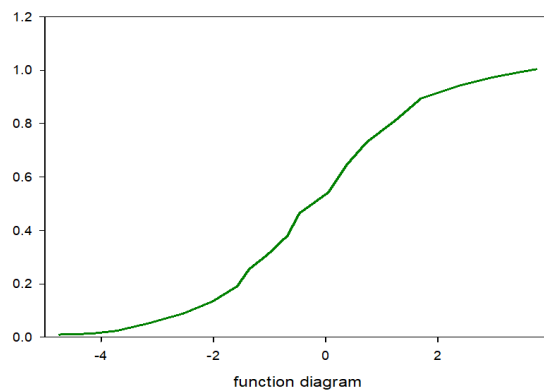


Figure 8. Sigmoid Function Diagram

After the logistic regression model is derived, the analytical solution cannot be obtained. The gradient-based iterative optimization algorithm can be used to solve the parameters. The most commonly used optimization algorithms are the stochastic gradient descent method and the batch gradient descent method. By derivation of the objective function, the direction of the derivative is also the direction of the gradient, that is, the current point is the initial point, at this position along the steepest descent direction of the terrain, and the gradient descent is to find the function along the direction of the gradient line. The optimal value of , stochastic gradient descent is that only one sample participates in the calculation every time the parameter w is updated, this optimization algorithm is easy to fall into the local optimal solution; batch gradient descent is that every time the parameter w is updated, all the training samples are To participate in the calculation, when the amount of data is large, the amount of calculation is large.

The experimental results of the AUC value are shown in Table 2.

Number of iterations	AUC
1	0.5443790485
2	0.5447356268
3	0.5459663297
4	0.5475944763
5	0.5478530141
6	0.54780117338
7	0.5480953722
8	0.5481455638
9	0.5481964282
10	0.5482134782

Table 2. AUC Value Distribution Table

4.2. Improvement of Algorithm Optimization Capability

In this experiment, we enhanced the linear regression model, that is, the decision tree of the linear regression model. The last expected result is likely to be the logistic regression model and the improved decision tree model. The two regression models are the ninth regression model, which replicates the logistics regression model and strengthens the model decision tree. 0.3 in case of error. In this study, we performed experiments with different numerical values, and the specific AUC experimental results are shown in Table 3.

<i>a</i>	0.1	0.2	0.4	0.6	0.8	0.9
AUC	0.615325	0.621744	0.632115	0.624388	0.619374	0.618257

Table 3. AUC Values of Logistic Regression and Enhanced Decision Fusion Training Results

The flowchart of the translation error detection model algorithm based on MFCC-RF is shown in Figure 9.

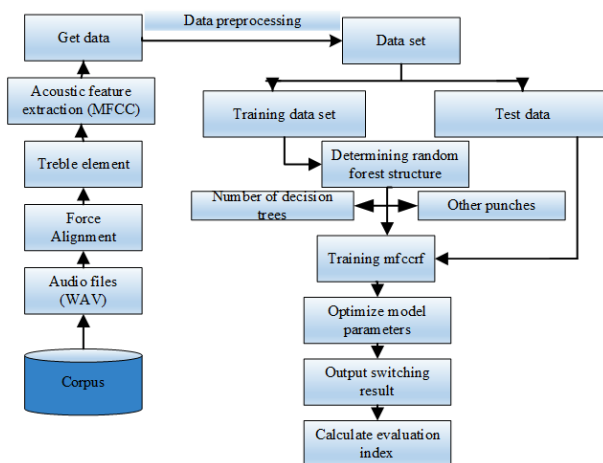


Figure 9. Flow Chart of Translation Error Detection Model Based on MFCC-RF

In this paper, we have realized the automatic generation of preposition oriented multiple-choice questions. The Euclidean distance similarity measure is used to automatically extract sentences from the corpus that are similar to the average sentence difficulty of the target textbook; Taking semantic similarity as the design principle of interference items, word2vec method is used to automatically generate interference items. We evaluated the difficulty of the stem sentence, the plausibility and reliability of the interference items objectively and manually. The evaluation results show that the

proposed method can initially generate test questions with different difficulties to meet the needs of learners with different English levels, and can effectively generate interference items with high reliability and plausibility. Another finding of the experimental results is that the difficulty of the whole test can be flexibly changed by constructing the interference item selection model from the perspective of interference items.

Figure 10 shows the tagging function / me / please at the audio translation level. MFC of different dimensions represents 13 lines of different colors, X represents the number of frames of samples, and linear represents cepstrum.

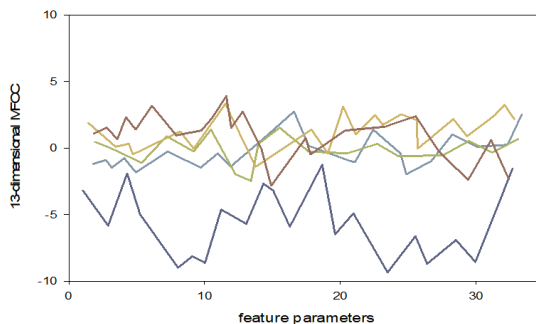


Figure 10. 13-Dimensional MFCC feature Parameters

This document distinguishes 13 static FCC sizes, increases the front and back sizes of information, and combines static and dynamic characteristics to better reflect continuous speech. In this file, 13 sizes 1 and 2 sizes 13 MFC sizes are self vector inputs of the model.

5. Conclusion

The core of computer supported translation training is to find translation errors and correct feedback. After discovering the previous translation errors, the training mainly focuses on several common errors such as input audio, reading errors and incomplete reading, which directly improves the translation quality after the training, and is equipped with an algorithm learner and a template for discovering translation errors. We are doing our best to promote this model in industrial applications. With the rapid development of network technology, we combine network and online learning. This mode is part of the Internet system and aims to build an error correction system that can be used by all people in real time. At the same time, in order to solve the problem of unbalanced distribution of sample data based on mfc-rf model, this document provides a translation error detection model based on profound belief and support based on network router (SVM). The results show that DBN SVM not only overcomes the defects of mfc-rf model samples, but also expands the scope of translation error detection. In addition, the accuracy of translation errors is slightly improved. The depth of neural network is used to detect errors, which lays a foundation for obtaining audio information for translation errors. Finally, in order to reflect the practicality of the model, an automatic translation system network is designed to correct errors, which effectively improves the translation quality and the learning efficiency of English learners. It is very important to speed up reform and improve English teaching methods.

6. Consent for publication

The author reviewed the results, approved the final version of the manuscript and agreed to publish it.

7. Conflicts of Interest

The author declares no conflicts of interest regarding this work.

References

1. Tang, J., 2020, September. Research on cultivation of college english translation abilities assisted by the internet. In *Journal of Physics: Conference Series* (Vol. 1634, No. 1, p. 012002). IOP Publishing.
2. Li, Q., 2021. Analysis and practice on the training of key ability of students majoring in electronic information in higher vocational education. *Procedia Computer Science*, 183, pp.791-793.
3. Huo, Y., Huang, A., Yang, Z., Xu, C., Yang, F., Shi, Y., He, L., Wang, G. and Wu, M., 2021. Development of practical liberal courses for students with non-biology majors under the "Double First-class" background. *Sheng wu Gong Cheng xue bao= Chinese Journal of Biotechnology*, 37(8), pp.2976-2983.
4. Goodlin, S.J., 2021. RESPONSE: Comprehensive Training for Heart Failure Clinicians: The Integration of Palliative Care. *Journal of the American College of Cardiology*, 77(4), pp.504-505.
5. Ni, C., 2021. Research on the training path of professional core literacy matching ability of private universities under the background of new liberal arts. *Open Journal of Social Sciences*, 9(10), pp.163-172.
6. Gao, Y., 2021. A survey study on the application of modern educational technology in English major college teaching in the age of 5G communication. *Theory and Practice in Language Studies*, 11(2), pp.202-209.
7. Xu, Q., 2021. Moral Education and College English Teaching—Taking "A Space Tourist Makes His Case" Text Study as an Example. *Open Access Library Journal*, 8(7), pp.1-6.
8. Mai, T.T., Wiest, G. and Nguyen, N., 2020. Asynchronous video-based discussion for the enhancement of intercultural competence among Vietnamese non-English majors. *Computer-Assisted Language Learning Electronic Journal*, 21(3), pp.159-174.
9. Tingting, S., 2021, May. Research on the construction of comprehensive training system for business administration majors in colleges based on computer. In *Journal of Physics: Conference Series* (Vol. 1915, No. 3, p. 032056). IOP Publishing.
10. Zhang, S. and Shi, X., 2020. Simulation training effectiveness evaluation based on fuzzy comprehensive evaluation. *Journal of Physics: Conference Series*, 1624(2), 022064 (4pp).
11. Li, G., 2020, July. The training skills of college students' oral english based on the computer-aided language learning environment. In *Journal of Physics: Conference Series* (Vol. 1578, No. 1, p. 012040). IOP Publishing.
12. Wang, Y. and Ma, X., 2020. Research on blended teaching ability of college English teachers. *Open Journal of Social Sciences*, 8(12), pp.308-319.
13. Zhengyu, P.A.N. and Kongkerd, S., 2022. Research on the marketing strategy of training institutions of guangxi dongfang yixue education group. *Psychology*, 12(6), pp.384-397.
14. Kuchma, V.R., Makarova, A.Yu. and Tikashkina, O.V., 2020. Comprehensive hygienic assessment of modern technologies of specialized medical training of high school students. *Hygiene and Sanitation*, 99(12), pp.1431-1437.
15. Zhou, Q., Yang, H., Zhou, Q. and Pan, H., 2021. Effects of cognitive motor dual-task training on stroke patients: A RCT-based meta-analysis. *Journal of Clinical Neuroscience*, 92, pp.175-182.

16. Panchal, M., Singh, S. and Rodriguez-Villegas, E., 2021. Analysis of the factors affecting the adoption and compliance of the NHS COVID-19 mobile application: a national cross-sectional survey in England. *BMJ open*, 11(8), p.e053395.
17. Xu, Y., Tao, Y., Zhang, C., Xie, M., Li, W. and Tai, J., 2022. Review of digital economy research in China: a framework analysis based on bibliometrics. *Computational Intelligence and Neuroscience*, 2022(1), p.2427034.
18. Akpaca, D., 2020. New trends in translation and interpretation teaching/training. *English Language Teaching and Linguistics Studies*, 2(2), pp.5-12.



©2024 the Author(s), licensee Combinatorial Press.
This is an open access article distributed under the
terms of the Creative Commons Attribution License
(<http://creativecommons.org/licenses/by/4.0>)