Application of Artificial Intelligence Software based on Semantic Web Technology in English Learning and Teaching

Yan Dong*
School of Foreign Languages, Chaohu University, China
dy20210918@126.com

Abstract

English teaching and learning in the information age need to be carried out with the help of intelligent system software to change the disadvantages of traditional English teaching as well as learning. According to the actual circumstances of teaching besides learning, this paper combines semantic Web technology and artificial intelligence technology to construct an English learning and teaching system. Moreover, this paper divides the entire speech data into frames in the speech processing, performs subsequent analysis in units of frames, and uses the autocorrelation function method in the time domain to extract the pitch of each frame of data corresponding to the English sentence. In addition, this paper combines the actual needs of English learning and teaching to construct system function modules, and designs experiments to analyze system performance and make statistics on user satisfaction. From the research results, it can be seen that the system constructed in this paper basically meets the actual needs of English autonomous learning and English teaching.

Keywords: Semantic web technology, Artificial intelligence, English learning, intelligent teaching

1. Introduction

The enrollment expansion policy of universities in recent years has resulted in an increasing number of graduates each year. At the same time, due to changes in the supply and demand of jobs, the employment situation of college students is not optimistic, which has attracted widespread attention from all walks of life. Therefore, we must adhere to the "employment-oriented" teaching philosophy. An excellent instructional passage indicates that you appreciate learning and have considered carefully your methods of teaching and evaluation tactics, as well as the goals you intend to achieve in your classes as well as the connection among research and education. A self-reflective expression of your ideas regarding education and learning is your process improvement. It's a one- to two-page storyline that expresses your essential beliefs about what it takes to be a good teacher in your field. This is not only a return to the basic attributes of college education, but also the objective requirements of the society for talents under the current new situation. After joining the World Trade Organization, China's manufacturing industry has developed rapidly and has gradually become a global manufacturing center. This requires not only tens of millions of professional research talents, but also hundreds of millions of applied talents with strong practical ability. The applied talents mainly come from college graduates. Students are the main body of higher vocational education. The prerequisite to ensure that every graduate from colleges and universities can better adapt to market-oriented competition is to ensure and improve the level of teaching management and the quality of education [1].

For this reason, the development and innovation of college English education are the first to bear the brunt. As a highly specialized language subject, English has become very common in all stages of education in various fields in our country. In particular, certain achievements have been made in college English education. However, there are still many problems, such as poor oral ability in actual use, which hinders the further improvement of teaching quality. Although teachers know how to screen out suitable English learning materials, due to limited sharing methods, it is difficult to share English resources in the hands of each school year in a timely manner. It can only be shared and taught in English classes once every few days, which makes good English learning materials appear "oversized and underutilized" within a certain range. The situation that students want to learn but cannot find English resources and teachers want to teach but cannot share English resources makes it difficult for English teaching management to proceed smoothly. In this regard, similar problems exist in English teaching at home and abroad. Based on this background, it is particularly meaningful to study and analyze English teaching management systems [2].

In today's highly developed information technology and the Internet, by using modern software design methods to model English teaching work and educational administration management, building an Internet resource sharing platform, and promoting the overall transformation of English teaching from the teaching mode, it is a new driving force and boost for English teaching. The construction of modern teaching technology and network management technology needs to be based on the actual background of English teaching and the actual situation of colleges and universities. At the same time, it needs to combine the English learning foundation of college students and the characteristics of the physical development of young people, and the own resources and teaching theories of college English teaching [3]. Artificial intelligence has been utilized in education, especially in the form of skill development tools besides test systems. It can enhance productivity, personalization & administrative responsibility, giving teachers more power as well as opportunities to concentrate on understanding and adaptation, which are unique qualities of human beings. In the field of education, artificial intelligence (AI) enables schools to create personalized learning opportunities for students.AI can determine a student’s learning rate and needs based on their data [16].

This article combines intelligent voice Web technology and artificial intelligence software to construct an English

*Corresponding Author: Yan Dong; E-mail: dy20210918@126.com
DOI: 10.53106/160792642022012301015

learning and teaching system, thereby effectively improving the effects of English learning and English teaching.

2. Related work

Experts and scholars have carried out relevant research on the evaluation of pronunciation quality and have achieved corresponding results. The literature [4] added prosodic factors to the original monophonic and triphonic models, and constructed a prosody model method to improve the performance of pronunciation quality evaluation. The literature [5] solved the problem of confusion between the probability space and the target pronunciation acoustic model by studying the frame-regulated logarithmic posterior probability and its transformation related to phonemes, so that the pronunciation quality evaluation performance has been significantly improved. The literature [6] proposed a new algorithm that introduces the GMM-UBM model in the phoneme pronunciation quality evaluation, and built a feature distribution model that is independent of phonemes. The scoring impact is superior than that of other algorithms, and it is comparable to expert scoring correlation. The literature [7] proposed a new calculation strategy, that is, applying linguistic rules in the logarithmic posterior probability algorithm. The literature [8] proposed a comprehensive evaluation algorithm for pronunciation quality depends on MFCC as well as LSP factors in addition to an impartial scoring algorithm depends on the ellipse technique, which greatly improved the objectiveness and rationality of pronunciation quality evaluation. The literature [9] proposed a new pronunciation quality evaluation algorithm, and successfully applied to the CALL system for English learners. Through the verification of the non-native language voice database collected and finely labeled by the laboratory, it is found to be superior to other scoring algorithms. The literature [10] comprehensively evaluated the pronunciation quality as compared with the intonation, velocity, rhythm, stress, as well as intonation of the sentence to be evaluated and the standard sentences of the corpus, and obtains good results. These research results provide strong support for the research and application of CALL.

Semantic WEB technology establishes a mathematical speech signal time series structure statistical model, which may be thought of as a double random process. One method is to mimic the implicit random process of changes in the statistical features of the speech signal using a Markov chain with a finite number of states. A Markov chain is a randomized process that shows a series of possible occurrences where the likelihood of each occurrence is exclusively resolute through the state attained in the previous event. A discrete-time Markov chain is a finitely infinite series in which the chain shifts state at time varying increments. [11]. The other is the arbitrary cycle of the perception succession related with each condition of the Markov chain. The previous is showed through the last mentioned, yet the particular boundaries of the previous are immense [12]. Indeed, the human discourse measure is additionally a twofold irregular interaction. The discourse signal is a discernible time-fluctuating arrangement, which is a flood of phoneme boundaries discharged by the cerebrum as per discourse needs as well as linguistic information (inconspicuous state). It can be seen that the semantic WEB technology reasonably imitates this process besides well describes the overall non-stationarity & local stationarity of the speech signal [13]. Semantic technology employs formal semantics to assist Intelligent systems in comprehending and processing information in the same way that individuals accomplish. As a result, they can store, handle, and retrieve data based on its significance and logical connections. Semantic Technology is a branch of computer science that develops languages to represent rich, personality data interrelationships in a machine-processable format [18]. Furthermore, its model library is not a pre-stored template, but an optimal model formed through repeated training. Moreover, in the recognition process, the best state sequence corresponding to the maximum likelihood probability between the speech sequence to be recognized and the semantic WEB technology model is output as the recognition result. Therefore, it is an ideal speech recognition model. In short, the semantic WEB technology model reasonably describes the acoustic model of speech, and the statistical training method is used in the speech recognition search algorithm to organically combine the underlying acoustic model and the upper language model, so better results can be obtained [14]. However, semantic WEB technology also has certain limitations. First, the method based on semantic WEB technology does not consider the impact of perception. Secondly, a large-scale speech corpus needs to be collected to train standard speech semantic WEB technology templates to obtain robust semantic WEB technology [15]. Furthermore, since CALL is an aid to second language learning, it involves more recognition of non-native language speech. When recognizing non-native-language speech, the recognition performance of semantic WEB technology usually trained by native-language speech will be greatly reduced, so the non-native-language speech needs to be adapted. Even so, the adaptive semantic WEB technology is still difficult to achieve good results in the recognition of non-native language speech [17]. Semantic WEB technology also has the following problems: it requires prior statistical knowledge of the speech signal, has weak classification decision-making ability, and has a large amount of calculation for the Viterbi recognition algorithm and the probability of the mixed Gaussian distribution. The Viterbi algorithm is a deterministic methodology for calculating the upper limit a scientific theory reliable prediction of most probable sequential order of hidden units, known as the Viterbi path, which ultimately resulted in a sequential order of observations, specifically in the context of Markov data sources and hidden Markov models. Both in perspective of storage and computation time, the Viterbi algorithm is costly [19]. For a succession of fixed length, genetic algorithm uses memory proportionate to event duration and time proportionate to side height to discover the optimum path via a structure with s stages and e edges. Problems with mixture distributions, on the other hand, are related to deriving the attributes of the general population from that of the subpopulations that are intended to produce inferential statistics based on their identification information [21]. These shortcomings make it difficult to further improve the performance of the semantic WEB technology model. For English speech recognition with large amounts of data and complex pronunciation changes, the deficiencies of semantic WEB technology are more obvious, making speech recognition take longer. Therefore, the speech recognition method based on semantic WEB technology has encountered a larger development bottleneck [20].

Based on the above analysis, this paper combines semantic WEB technology and artificial intelligence software to

construct English learning and teaching software, so as to overcome the shortcomings of traditional semantic WEB technology.

3. Mel frequency cepstrum coefficient

Mel Frequency Cepstral Coefficient (MFCC) is based on the human hearing mechanism, which simulates the human ear's response to speech signals of different frequencies. In fact, the human ear's response sensitivity to different frequencies of speech signals is different, and is similar to a special nonlinear system, and is basically a logarithmic relationship [22-24].

The extraction process of MFCC feature parameters is shown in Figure 1.

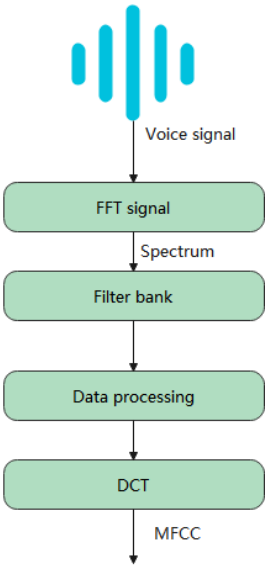


Figure 1. The extraction process diagram of MFCC feature parameters

MFCC is amongst the most extensively utilized speech processing techniques. It is derived from the audio signal, which acts as a hub and accurate starting point for voice recognition. The voice frame is sent through a quantizer windows after preprocessing, and the energy level is computed using a rapid Fourier transformation. To eliminate the influence of overtones, a Mel filter bank is utilized. The spatial domain conversion is the final stage. Despite the fact that many speech aspects may be retrieved, these still are singular features that are unrelated to the remainder of the voice stream [25].

The calculation of MFCC speech feature parameter extraction is as follows.

1. The fast Fourier transform (FFT) is shown in formula (1):

$$X[k]=\sum_{n=0}^{N-1}x[n]e^{-j\frac{2\pi}{N}nk},k=0,1,2,\cdots,N-1 \tag{1}$$

Where,

$x[n],n=0,1,2,\cdots,N-1$ - frame of discrete speech sequence attained next to sampling,

N - frame length.

$X[k]$ - complex sequence of N points,

Then $X[k]$ is modulated to obtain the signal amplitude spectrum $|X[k]|$. The great feature of Fourier analysis would be that the input loses very little data throughout the conversion. The Fourier transform preserves intensity, harmonic, and planning while translating the information into the frequency response using all elements of the waveforms.

2. The authentic frequency scale is converted to Mel frequency scale as follows:

$$Mel(f)=2595\lg\left(1+\frac{f}{700}\right) \tag{2}$$

Where,

$Mel(f)$ - Mel frequency,

f - actual frequency, its unit is Hz.

3. The triangular filter bank is configured and the output of every triangular filter next to filtering the signal amplitude spectrum $|X[k]|$ is calculated:

$$F(l)=\sum_{k=f_o(l)}^{f_h(l)}w_l(k),l=1,2,\cdots,L \tag{3}$$

Among them,

$$w_l(k)=\begin{cases}\dfrac{k-f_o(l)}{f_c(l)-f_o(l)} & f_o(l)\le k\le f_c(l)\\ \dfrac{f_h(l)-k}{f_h(l)-f_c(l)} & f_c(l)\le k\le f_h(l)\end{cases} \tag{4}$$

$$\begin{aligned}f_o(l)&=\frac{o(l)}{\left[\frac{f_s}{N}\right]}\\ f_h(l)&=\frac{h(l)}{\left[\frac{f_s}{N}\right]}\\ f_c(l)&=\frac{c(l)}{\left[\frac{f_s}{N}\right]}\end{aligned} \tag{5}$$

Where,

$w_l(k)$ - filter coefficient of the corresponding filter,

$o(l)$、$c(l)$、$h(l)$ - lower limit frequency, center frequency besides upper limit frequency of the corresponding filter on the actual frequency axis,

f_s - sampling rate,

L - number of filters, and

$f(l)$ - filter output.

4. Logarithm calculation is performed on all filter outputs, besides Discrete Cosine Transform (DTC) is further performed to obtain MFCC characteristics, such as:

$$M(i)=\sqrt{\frac{2}{N}}\sum_{l=1}^{L}\log F(l)\cos\left[\left(l-\frac{1}{2}\right)\frac{i\pi}{L}\right],i=1,2,\cdots,Q \tag{6}$$

Where,

Q - order of MFCC parameters, this paper takes 13, and

$M(i)$ - obtained MFCC parameters.

The goal of employing MFCCs is to characterize the randomized input image with characteristics that are resistant to measurements mistakes. The MFCCs generated from the Discrete cosine transform of the input improve standard errors resilience. Figure 2 is a two-dimensional diagram of the MFCC parameters of the sentence that It will be in the place where we al-ways put it.

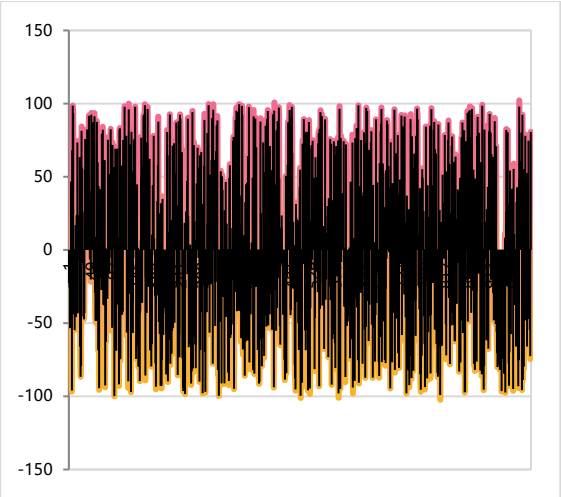


Figure 2. Two-dimensional MFCC diagram of the sentence that It will be in the place where we always put it

Chinese English learners have made great breakthroughs in the overall level of pronunciation in lately, especially in the aspect of monophonic (vowel and consonant), such as the pronunciation of vowels is late and becomes fuller. This is largely due to the country's investment in education and teaching resources and the enthusiasm of English learners. Therefore, for most English learners, they can better master the pronunciation of English words. However, in real life, sentence rhythm is the focus of natural and fluent English communication. Numerous learners encounter a bottleneck next to they have mastered the pronunciation of monophonic sounds, since they find that they can pronounce every sound supplementary standardly, then the English they express still has a robust Chinese flavor. The stated collection contains beginners, intermediate learners, English majors as well as certain individual who use English for a long time. The intensity of the Frequency Domain simply indicates how much each Logo black is present in any given device. Whenever the time domain waveforms are written as a combination of cumulative density rates, the amplitude of each bin is the intensity of such a carrier frequency for the that waveform throughout the spatial domain [26]. The prosody of phonetics is a very important factor in sentences. Every language has its own features in terms of prosody, besides sentences that cannot grasp the prosody of the language will appear unnatural.

The evaluation of pronunciation quality is mainly to comprehensively evaluate the intonation, length, and rhythm of the speech. In the evaluation process, phonemes and words are mainly evaluated based on intonation; while sentences and paragraphs not only examine the content expressed by the pronunciation itself, but their prosodic characteristics determine the true meaning of the sentence to a large extent. When evaluating the pronunciation quality of sentences and paragraphs, it is necessary to comprehensively consider the prosodic information such as whether the speaker can grasp the key information of the sentence relatively accurately, whether the unimportant information of the relatively weak sentence is relatively weak, and whether the sound length is appropriate. In other words, in the evaluation of the pronunciation quality of English sentences, a good pronunciation quality requires not only complete & precise content, strong as well as fluent elocution, absence of obvious elocution errors, but also reasonable speaking velocity, precise accent elocution, strong sense of rhythm, and accurate and natural intonation. Therefore, this article uses two major indicators of intonation and prosody to evaluate pronunciation quality.

The inflection assessment fundamentally looks at whether the substance data of the articulation sentence is finished besides precise, regardless of whether the elocution is clear & familiar, in addition to whether there are articulation blunders. Here, the MFCC coefficients dependent on the human hearing model are utilized as the assessment boundaries of inflection, as well as the discourse acknowledgment model is set up through the profound conviction network for discourse acknowledgment to decide if the substance is finished besides right. Simultaneously, the connection coefficient between the standard sentence as well as the MFCC element of the information sentence is determined to pass judgment on the elocution Is it clear & familiar.

The superiority of English articulation is evaluated through intonation and feedback, as shown in Figure 3.

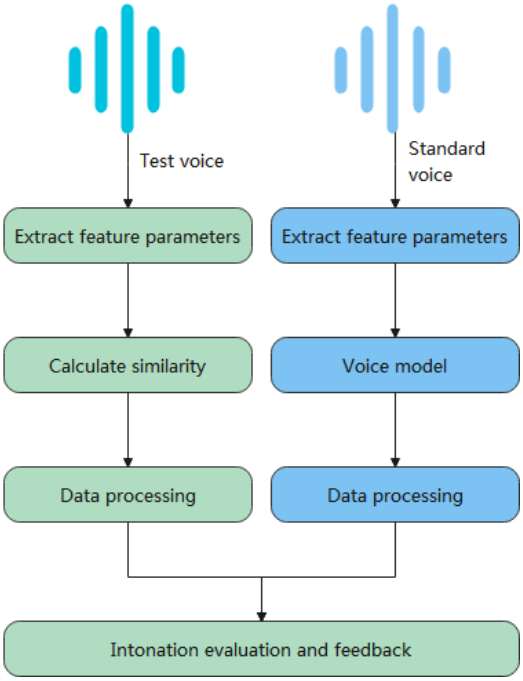


Figure 3. Intonation evaluation

Speaking rate for the most part alludes to the speed of elocution, which is a proportion of the speed of the speaker's articulation. It tends to be reflected through computing the quantity of syllables N spoken in a unit of time T, in addition to it very well may be generally estimated through the complete discourse span including stops. Since various speakers have certain distinctions in talking rate, the way to express a similar sentence is diverse in the sentence length of various individuals. Likewise, the speaker's enthusiastic state will likewise influence the pace of discourse. For instance, in irate besides glad expresses, the discourse rate is by and large

somewhat quicker than in the quiet state, while in bitterness, the discourse rate is for the most part slower.

This article adopts the speech rate estimation dependent on the speech period to calculate the time ratio φ of the test sentence as well as the standard sentence, as shown in the following formula:

$$\varphi = \frac{Len_{Std}}{Len_{Test}} \tag{7}$$

Among them,

Len_{Std} - period of the standard sentence, Len_{Test} - period of the test sentence.

Further, φ is compared with the set speech rate threshold, as shown in Figure 4. It should be noted that the duration is preprocessed by the dual-threshold endpoint detection technique of short-term energy as well as short-term average zero-crossing rate, which can effectively eliminate the noise interference of the silent section.

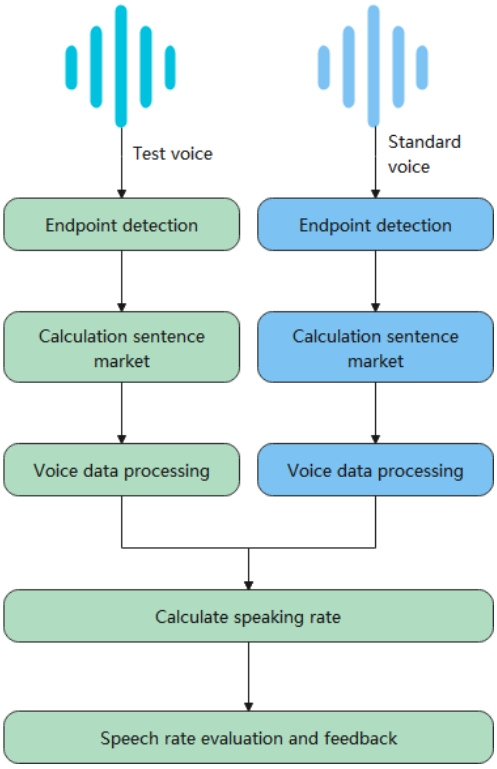


Figure 4. Speech rate evaluation

The rhythm of the language is the similarities and alterations of the height, severity, length, and urgency of the phonetic, and it appears regularly as well as interchangeably with a confident sort of phonetic unit fragments. It is categorized into three sorts: fully accented, incomplete accented, and emphasized accented. When reading and speaking, the rhythm groups formed by different combinations appear alternately, and its meaning function is to enhance the melody and sense of music. Standard voice focuses on community engagement, entertainment, and way of life, which includes advertising in our media via our platform. As a measurement of ability to hear, a test of a person's personal perception of sound a murmured either spoken keyword from a certain distance. An audition involving a vocal test and examination of a recording specimen of a voice to identify if it corresponds to a specific person.

English is a typical stress-based language, that is, the basis and main body of each sentence in English are stressed syllables, and the quantity of stressed syllables determines the beat of the sentence. This is different from Chinese—a language that is timed by syllables, that is, the number of syllables determines the beat of the sentence.

English sentences have the following three characteristics: (1) Generally speaking, the higher the occurrence of harassed syllables in the sentence, the slower the speaking speed, and the clearer the syllables will sound; (2) The unstressed syllables appear crowded between the stressed syllables The syllable sounds brisk and vague; (3) The length of time needed to talk a sentence doesn't rely upon the quantity of words or syllables in the sentence, but more importantly, it depends on the quantity of stressed syllables in the sentence. Stressed syllables play a role of emphasis as well as dissimilarity in sentence organization besides semantic expression, then have the following three characteristics: (1) loud; (2) long elocution; (3) clear as well as easy to differentiate.

The rhythm evaluation mechanism is shown in Figure 5, which specifically includes the following steps:

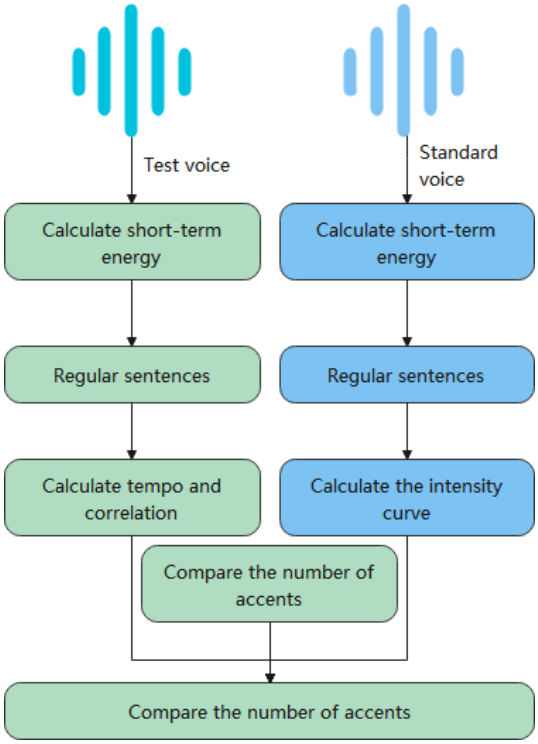


Figure 5. Rhythm evaluation

(1) The short-term energy value of speech is extracted to form a graph of speech intensity.

The loudness of stressed syllables in a sentence is directly reflected in the energy concentration in the period realm, that is, the speech energy intensity of the stressed syllable is large. The short-term energy of the speech signal $s(n)$ is defined as follows:

$$E_n = \sum_{m=-\infty}^{\infty} [s(n)\omega(n-m)]^2 \tag{8}$$

The short-term energy value is extracted from the speech sentence to form an intensity graph.

(2) Regularize sentences.

Due to the dissimilarity in speech speed betwixt dissimilar speakers, the pronunciation of the same sentence will have different sentence lengths for different speakers. However, its pronunciation follows a certain rule, that is, the ratio of the duration of the stressed syllable in the sentence to the duration of the entire sentence is relatively fixed. Therefore, in order to facilitate data processing and obtain more objective evaluation results, before evaluating the test sentence, the duration of the test sentence needs to be scaled to a degree similar to the standard sentence.

(3) The improved dynamic time rounding (DTW) algorithm is utilized to compute the matching degree of the intensity curve betwixt the standard sentence as well as the input sentence.

The basic principle of the DTW algorithm is dynamic time warping, which matches the originally mismatched time length among the test template besides the reference template. To compute the similarity, we utilize traditional Euclidean distance. Let the reference template as well as the test template be R and T. The smaller the distance $D[T,R]$, the higher the similarity. The drawback of the existing DTW algorithm is that when template matching is performed, the weights of all frames are the same, and all templates must be matched. DTW is a series data realignment technique that was initially designed for voice recognition. Its ability to align 2 models of extracted features by bending the temporal vector repeatedly till the diverse set are perfectly aligned. The amount of calculation is relatively large, especially when the number of templates increases quickly, the amount of calculation increases very quickly.

As shown in Figure 6, this paper limits the intersection to be calculated within the parallelogram by setting the matching boundary. R and T are divided into N and M frames in equal time division, and can be separated into 3 sections of path $(1,X_a)$、(X_a+1,X_b)、(X_b+1,N) to calculate the distance. According to the coordinate calculation, we can get:

$$X_a=\frac{1}{3}(2M-N) \tag{9}$$

$$X_b=\frac{2}{3}(2N-M) \tag{10}$$

X_a、X_b is set to the nearest integer. When the restriction condition $2M-N\geq 3, 2N-M\geq 2$ is not met, dynamic matching is not performed, which reduces system expenses.

Every frame on the X axis matches the frame betwixt $[y_{\min},y_{\max}]$ on the Y axis, and the formula for calculating $y_{\min}$、$y_{\max}$ is as follows:

$$y_{\min}=\begin{cases}\frac{1}{2}x & x\in[0,X_b]\\ 2x+(M-2N) & x\in[X_b,N]\end{cases} \tag{11}$$

$$y_{\max}=\begin{cases}2x & x\in[0,X_a]\\ \frac{1}{2}x+\left(M-\frac{1}{2}N\right) & x\in[X_a,N]\end{cases} \tag{12}$$

If $X_a>X_b$, the matched path could be divided into $(1,X_b)$、(X_b+1,X_a)、(X_a+1,N).

When the X coordinate axis is forwarded by one frame, although the number of frames corresponding to the Y coordinate axis is different, the regularity characteristics are the same, and the cumulative distance is:

$$D(x,y)=d(x,y)+\min\begin{Bmatrix}D(x-1,y)\\ D(x-1,y-1)\\ D(x-1,y-2)\end{Bmatrix} \tag{13}$$

Among them, D and d represent the cumulative distance and the frame matching distance respectively.

(4) The stress threshold as well as the non- stress threshold are set as the characteristic double threshold and the accented vowel duration, and the stress unit is divided to determine the number of accented pronunciations.

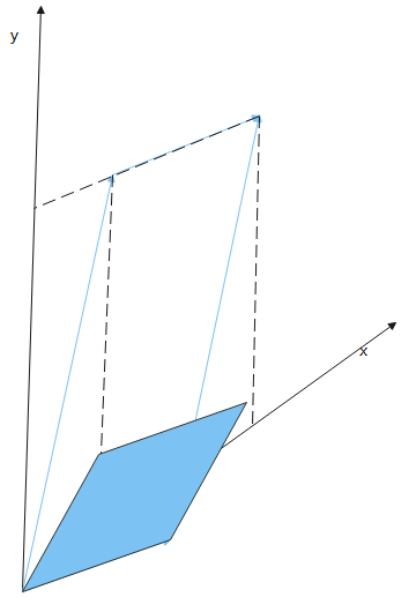


Figure 6. Schematic diagram of matching path constraints

In this paper, the double-threshold comparison technique is utilized for accent endpoint detection. The accomplishment of endpoint identification immediately enhances voice coding, voice commands, automatic speech, and social contact quality and productivity. In the presence of low signal-to-noise ratios and complicated noise, the resilience and classification accuracy of methods are always hot subjects for many researchers. After a lot of experimental verification, the threshold is set as follows

The stress threshold is:

$$T_u=\left(\max(sig\_in)+\min(sig\_in)\right)/2.5 \tag{14}$$

The non-stress threshold is:

$$T_l=\left(\max(sig\_in)+\min(sig\_in)\right)/10 \tag{15}$$

In the double-threshold comparison method, the maximum speech energy value $S_{\max}$

T_u

S_l and S_r equal to the unstress threshold T_l are searched to the left and right of $S_{\max}$, then the sentence

stress signal can be set to S_l and S_r. At the same time, the energy value between S_l and S_r is set to 0 to avoid repeated searching between S_l and S_r. Because the pronunciation of the stressed syllable in the sentence is too long, the stressed syllable unit searched in the first step may have a large energy value, that is, the auditory performance is bright, but the duration is very short. These units do not constitute stressed syllables. They may be short vowels or interference from signal spikes. Therefore, the stressed syllable units need to be further screened according to the characteristics of long pronunciation of accented syllables.

This article uses an improved dPVI parameter calculation formula to compare the length of the syllable unit segment of the average sentence as well as the test sentence correspondingly, then utilizes the converted factors used for the system evaluation basis, as shown in the following formula:

$$dPVI = 100 \times \left(\sum_{k=1}^{m-1} |d1_k - d2_k| + |d1_t - d2_t| \right) \Big/ Len \tag{16}$$

Among them, d is the period of the phonetic unit segment divided by the sentence (for example, d_k is the duration of the k-th phonetic unit segment), $m = \min \begin{pmatrix} StdNumber\ of\ units, \\ TestNumber\ of\ units \end{pmatrix}$, and Len is the period of the standard sentence. Since the length of the test sentence has been adjusted to be equivalent to the length of the standard sentence before the PVI operation, the calculation can only use Len as the calculation unit.

(6) The number of accents, intensity curve matching and dPVI parameters of test sentences and standard sentences are comprehensively compared to estimate besides feedback the superiority of English elocution.

In the investigation of sound, pitch is the most fundamental and significant part of inflection. According to the actual perspective of sound, the pitch is dictated by the vibration recurrence of the article. The recurrence of the vibration of an item is corresponding to the pitch: the more the quantity of vibrations, the higher the sound, while the less the quantity of vibrations, the lower the sound. In speech, the ups and downs of the sound are expressed as pitch. The voice material form of pitch is manifested as the fundamental frequency change of the vocal cords. From the change of the fundamental frequency, the different modes of intonation change can be determined, that is, the pitch can determine the different modes of intonation. Therefore, the key to intonation evaluation is to extract the pitch corresponding to each frame of the speech signal in the sentence.

The whole pitch is smoothed through setting the middle channel, lastly the DTW calculation is utilized to compute the level of sound fit between the standard sentence as well as the information sentence, to assess besides input the English elocution quality, as displayed in Figure 7.

The autocorrelation function method uses the autocorrelation function to calculate the similarity between a sound frame $s(i), i = 0,1,2,\cdots,n-1$ and itself, as shown in the following formula:

$$acf(\tau) = \sum_{i=0}^{n-1-\tau} s(i)s(i+\tau) \tag{17}$$

Among them, n refers to the length of one frame of voice data, and τ is the amount of time delay, which is in units of sampling points. The pitch of the frame can be calculated by first finding the value of τ that can make $acf(\tau)$ in a certain reasonable interval.

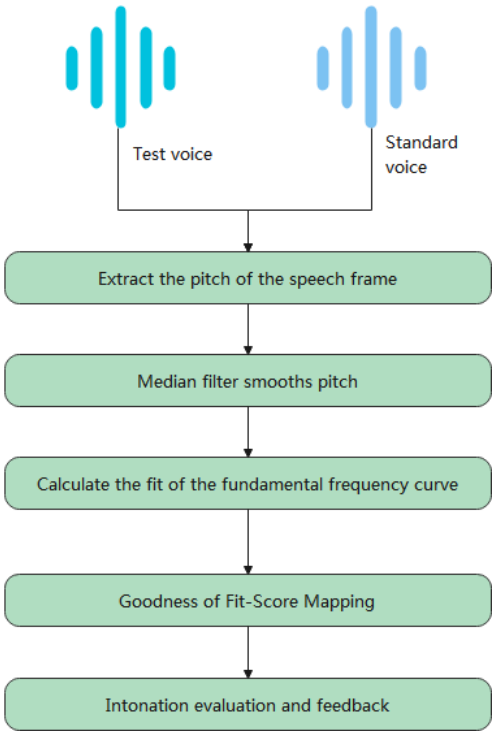


Figure 7. Analysis of Intonation evaluation

Different groups (such as elementary school students, middle school students, college students, business people, etc.) have different requirements for learning spoken English, and their English pronunciation quality evaluation standards are also different, as shown in Figure 8.

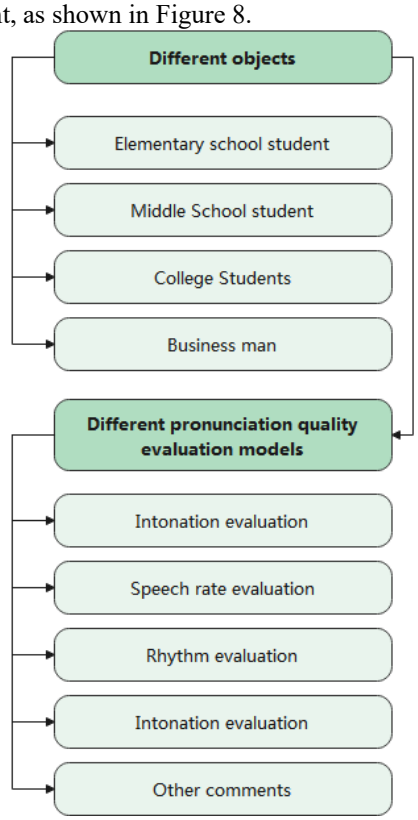


Figure 8. Multi-parameter pronunciation quality evaluation model for different objects

This article takes college students' spoken English as the research object, and comprehensively analyzes the relationship between the various indicators based on the evaluation of multiple pronunciation quality indicators such as intonation, speech speed, rhythm and intonation. Moreover, this paper focuses on considering the weight of each indicator in the overall pronunciation quality evaluation, and establishes a multi-parameter English pronunciation quality evaluation model and method for college students to conduct a reasonable and objective comprehensive evaluation of pronunciation quality, as shown in Figure 9.

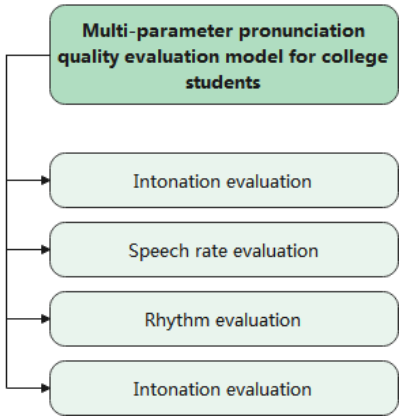


Figure 9. Multi-parameter pronunciation quality evaluation model for college students

In communications, intonation is particularly essential since it provides information further than the simple meanings of words. It can both express and provide semantic meaning, and also convey the presenter's opinion or feeling about someone else. The rate of speech is frequently indicated in words per minute. You'll have to record oneself conversing for several minutes and then tally increasing the total of words within your voice to get this figure. Subtract the total amount of words from the number of minutes it took us to deliver your presentation. The trajectory of the least monitoring time necessary for each load percentage is displayed against by the proportion of overall burden after the rhythmic record has been recreated.

4. Analysis of the application effect of artificial intelligence software based on semantic Web technology in English learning and teaching

This paper constructs an application system of artificial intelligence software based on semantic Web technology in English learning and teaching. On this basis, the performance of the system is tested and analyzed. The system constructed in this paper can be used for English autonomous learning and English intelligent teaching. Therefore, the practical effect test of the system in this paper is mainly carried out by investigating user satisfaction. This article separately surveys the students and teachers who use this system, and counts the satisfaction of learning effect and teaching effect. The obtained learning effect satisfaction is shown in Table 1 and Figure 10.

Table 1. Statistical table of learning effect satisfaction

NO	Student satisfaction	NO	Student satisfaction	NO	Student satisfaction
1	83.5	26	74.5	51	93.3
2	86.8	27	89.9	52	88.3
3	83.4	28	81.0	53	79.9
4	92.8	29	82.9	54	94.3
5	74.6	30	94.2	55	80.0
6	77.1	31	83.1	56	92.4
7	94.2	32	91.4	57	85.0
8	84.2	33	76.0	58	89.8
9	85.7	34	92.0	59	80.0
10	74.0	35	82.1	60	91.9
11	83.9	36	77.5	61	84.7
12	86.3	37	85.3	62	85.5
13	73.1	38	88.4	63	81.2
14	90.2	39	92.0	64	81.0
15	85.0	40	87.1	65	79.2
16	89.9	41	76.9	66	76.8
17	80.1	42	86.2	67	80.0
18	80.8	43	72.5	68	86.4
19	91.1	44	94.8	69	78.7
20	77.7	45	76.4	70	90.8
21	94.9	46	88.4	71	90.6
22	72.5	47	83.3	72	80.1
23	84.5	48	94.7	73	86.0
24	72.5	49	94.1	74	93.2
25	74.7	50	90.4	75	86.2

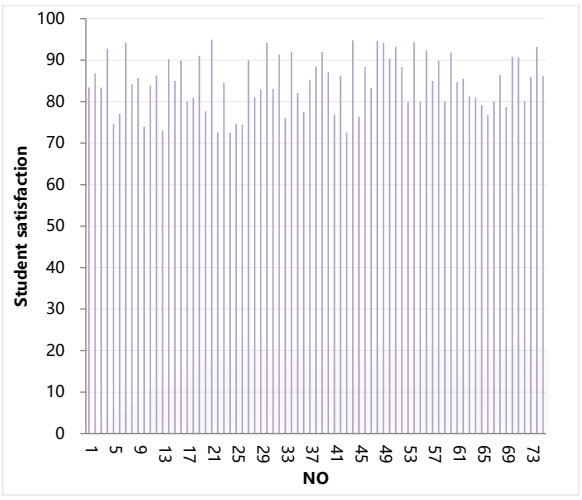


Figure 10. Statistical diagram of learning effect satisfaction

From the above results, it can be seen that students can start their own English learning more conveniently and quickly through this system. The teacher's teaching satisfaction obtained from the test is shown in Table 2 and Figure 11 below.

Table 2. Statistical table of teaching effect satisfaction

NO	Teacher satisfaction	NO	Teacher satisfaction	NO	Teacher satisfaction
1	94.9	26	85.9	51	95.8
2	86.1	27	88.7	52	83.6
3	95.6	28	91.5	53	85.8
4	96.9	29	79.6	54	95.6
5	96.3	30	90.4	55	84.9
6	89.9	31	83.8	56	81.3
7	90.1	32	87.7	57	91.4
8	82.9	33	87.6	58	90.8
9	78.8	34	84.9	59	84.8
10	90.6	35	94.1	60	90.9
11	87.0	36	96.6	61	84.9
12	81.1	37	83.2	62	95.8
13	82.0	38	91.6	63	95.0
14	86.4	39	91.3	64	85.7
15	86.5	40	83.8	65	88.5
16	84.2	41	78.6	66	82.2
17	96.0	42	93.9	67	87.8
18	80.3	43	90.6	68	87.1
19	96.2	44	83.3	69	82.1
20	94.4	45	79.6	70	85.0
21	92.6	46	86.5	71	91.1
22	87.5	47	89.9	72	96.0
23	88.7	48	81.2	73	84.0
24	89.5	49	85.3	74	82.6
25	82.4	50	96.0	75	86.2

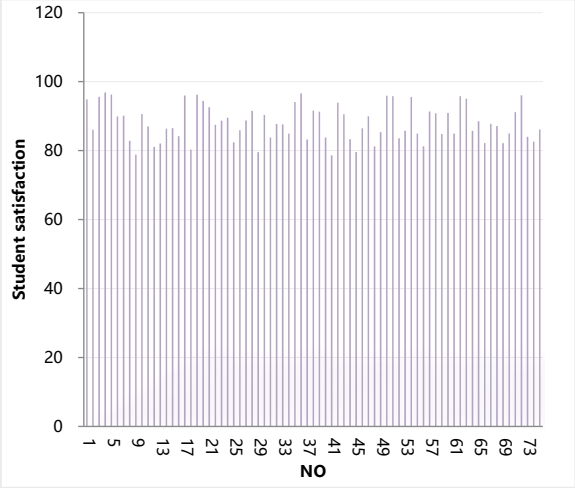


Figure 11. Statistical diagram of teaching effect satisfaction

From the above analysis results, the method proposed in this paper can play an important role in English autonomous learning and teaching.

5. Conclusion

In computer-assisted speech learning, speech recognition technology and speech evaluation technology are the core. Among them, speech recognition technology is particularly critical and plays a vital role. The reason is that speech recognition is an important basis and prerequisite for speech evaluation, and only high-accuracy speech recognition can further obtain good speech evaluation results. This paper combines semantic Web technology and artificial intelligence technology to construct an English teaching system that can be used for students' autonomous learning and teachers' intelligent teaching. Moreover, this paper takes college students' spoken English as the research object, and comprehensively analyzes the relationship between the various indicators based on the evaluation of multiple pronunciation quality indicators such as intonation, speaking speed, rhythm and intonation. In addition, this paper focuses on considering the weight of each indicator in the overall pronunciation quality evaluation, and establishes a multi-parameter English pronunciation quality evaluation model and method for college students. Finally, this paper verifies the performance of the system through experimental research. From the experimental research results, it can be seen that the system constructed in this paper meets the needs of English teaching and autonomous learning. The future advice is to focus on continual evaluation; elevated assessment may become obsolete, and wider data may be employed to evaluate abilities and competencies.

Acknowledgement

Project funded by Education Department of Anhui Province: Translation Teaching and Material Selection from the Perspective of Ideological and Political Theory Teaching (Approval No.: 2020KCSZYJXM174); Key Teaching Research Project funded by Chaohu University: OBE-Oriented Test and Teaching Assessment of Foreign Language Courses (Approval No: 2020JYXM1251)

References

[1] A. Rajkomar, J. Dean, I. Kohane, Machine learning in medicine, *New England Journal of Medicine*, Vol. 380, No. 14, pp. 1347-1358, April, 2019.

[2] Y. Xin, L. Kong, Z. Liu, Y. Chen, Y. Li, H. Zhu, M. Gao, H. Hou, C. Wang, Machine learning and deep learning methods for cybersecurity, *IEEE Access*, Vol. 6, pp. 35365-35381, May, 2018.

[3] L. Ward, A. Agrawal, A. Choudhary, C. Wolverton, A general-purpose machine learning framework for predicting properties of inorganic materials, *npj Computational Materials*, Vol. 2, No. 1, pp. 1-7, August, 2016.

[4] P. Feng, B. Wang, D. L. Liu, C. Waters, Q. Yu, Incorporating machine learning with biophysical model can improve the evaluation of climate extremes impacts on wheat yield in south-eastern Australia, *Agricultural and Forest Meteorology*, Vol. 275, pp. 100-113, September, 2019.

[5] K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis, D. I. Fotiadis, Machine learning applications in cancer prognosis and prediction, *Computational and Structural Biotechnology Journal*, Vol. 13, pp. 8-17, 2015.

[6] S. Amershi, M. Cakmak, W. B. Knox, T. Kulesza, Power to the people: The role of humans in interactive machine learning, *Ai Magazine*, Vol. 35, No. 4, pp. 105-120, December, 2014.

[7] V. Rodriguez-Galiano, M. Sanchez-Castillo, M. Chica-Olmo, M. Chica-Rivas, Machine learning predictive models for mineral prospectivity: An

evaluation of neural networks, random forest, regression trees and support vector machines, *Ore Geology Reviews*, Vol. 71, pp. 804-818, December, 2015.

[8] W. Coley, R. Barzilay, T. S. Jaakkola, W. H. Green, K. F. Jensen, Prediction of organic reaction outcomes using machine learning, *ACS Central Science*, Vol. 3, No. 5, pp. 434-443, May, 2017.

[9] A. Chowdhury, E. Kautz, B. Yener, D. Lewis, Image driven machine learning methods for microstructure recognition, *Computational Materials Science*, Vol. 123, pp. 176-187, October, 2016.

[10] S. Basith, B. Manavalan, T. H. Shin, G. Lee, SDM6A: A web-based integrative machine-learning framework for predicting 6mA sites in the rice genome, *Molecular Therapy-Nucleic Acids*, Vol. 18, pp. 131-141, December, 2019.

[11] C. Voyant, G. Notton, S. Kalogirou, M.-L. Nivet, C. Paoli, F. Motte, A. Fouilloy, Machine learning methods for solar radiation forecasting: A review, *Renewable Energy*, Vol. 105, pp. 569-582, May, 2017.

[12] C. Folberth, A. Baklanov, J. Balkovič, R. Skalský, N. Khabarov, M. Obersteiner, Spatio-temporal downscaling of gridded crop model yield estimates based on machine learning, *Agricultural and Forest Meteorology*, Vol. 264, pp. 1-15, January, 2019.

[13] J. Sieg, F. Flachsenberg, M. Rarey, In need of bias control: evaluating chemical data for machine learning in structure-based virtual screening, *Journal of Chemical Information and Modeling*, Vol. 59, No. 3, pp. 947-961, March, 2019.

[14] F. Thabtah, D. Peebles, A new machine learning model based on induction of rules for autism detection, *Health Informatics Journal*, Vol. 26, No. 1, pp. 264-286, March, 2020.

[15] F. A. Narudin, A. Feizollah, N. B. Anuar, A. Gani, Evaluation of machine learning classifiers for mobile malware detection, *Soft Computing*, Vol. 20, No. 1, pp. 343-357, January, 2016.

[16] D.-P. Deng, G.-S. Mai, C.-H. Hsu, C.-L. Chang, T.-R. Chuang, K.-T. Shao, Linking Open Data Resources for Semantic Enhancement of User–Generated Content, in: H. Takeda, Y. Qu, R. Mizoguchi, Y. Kitamura (Eds.), *Semantic Technology. Lecture Notes in Computer Science*, Springer, 2013, pp. 362-367.

[17] Q. Yao, H. Yang, R. Zhu, A. Yu, W. Bai, Y. Tan, J. Zhang, H. Xiao, Core, mode, and spectrum assignment based on machine learning in space division multiplexing elastic optical networks, *IEEE Access*, Vol. 6, pp. 15898-15907, March, 2018.

[18] X. Xu, D. Li, M. Sun, S. Yang, S. Yu, G. Manogaran, G. Mastorakis, C. X. Mavromoustakis, Research on Key Technologies of Smart Campus Teaching Platform Based on 5G Network, *IEEE Access*, Vol. 7, pp. 20664-20675, January, 2019.

[19] W. Zheng, B. Muthu, S. N. Kadry, Research on the Design of Analytical Communication and Information Model for Teaching Resources with Cloud-Sharing Platform, *Computer Applications in Engineering Education*, Vol. 29, No. 2, pp. 359-369, March, 2021.

[20] D. Bzdok, A. Meyer-Lindenberg, Machine learning for precision psychiatry: opportunities and challenges, *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, Vol. 3, No. 3, pp. 223-230, March, 2018.

[21] M. Gunasekaran, N. Chilamkurti, C.-H. Hsu, Editorial Note: Machine Learning for Visual Analysis of Multimedia Data, *Multimedia Tools and Applications*, Vol. 79, pp. 7-8, 5003-5003, February, 2020.

[22] M. Chen, Y. Hao, K. Hwang, L. Wang, L. Wang, Disease prediction by machine learning over big data from healthcare communities, *IEEE Access*, Vol. 5, pp. 8869-8879, April, 2017.

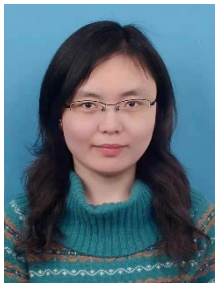
[23] L. Itu, S. Rapaka, T. Passerini, B. Georgescu, C. Schwemmer, M. Schoebinger, T. Flohr, P. Sharma, D. Comaniciu, A machine-learning approach for computation of fractional flow reserve from coronary computed tomography, *Journal of Applied Physiology*, Vol. 121, No. 1, pp. 42-52, July, 2016.

[24] U. Jayasinghe, G. M. Lee, T.-W. Um, Q. Shi, Machine learning based trust computational model for IoT services, *IEEE Transactions on Sustainable Computing*, Vol. 4, No. 1, pp. 39-52, January-March, 2019.

[25] B. Thoms, N. Garett, T. Ryan, The Design and Evaluation of a Peer Ratings System for Online Learning Communities, *43rd Hawaii International Conference on System Sciences*, Honolulu, HI, USA, 2010, pp. 1-10.

[26] S. Khan, A. Al-Dmour, V. Bali, M. R. Rabbani, K. Thirunavukkarasu, Cloud computing based futuristic educational model for virtual learning, *Journal of Statistics and Management Systems*, Vol. 24, No. 2, pp. 357-385, March, 2021.

Biography



Yan Dong, born in Dec. 1977, female, Han nationality, MA degree, associate professor at Chaohu University, research fields are English teaching, translation.