Hindawi
Mathematical Problems in Engineering
Volume 2022, Article ID 3879266, 11 pages
https://doi.org/10.1155/2022/3879266

Research Article

Decomposition and Recognition of English Speech Features Based on Neutrosophic Set Fuzzy Control and Random Matrix Theory

Na Su[1] **and Rohani Othman**[2]

[1] *Pinghu Normal School, Jiaxing University, Pinghu 314200, Zhejiang, China*
[2] *Language Academy, Faculty of Social Sciences and Humanities, Universiti Teknologi Malaysia, Johor Bahru 81310, Johor, Malaysia*

Correspondence should be addressed to Na Su; sunajxxy@zjxu.edu.cn

Received 29 April 2022; Revised 31 May 2022; Accepted 6 June 2022; Published 29 June 2022

Academic Editor: Ning Cao

In order to improve the effect of special decomposition and recognition of English speech, based on the idea of neutrosophic set fuzzy control, this paper uses Bayesian method as the basic algorithm of speech recognition to improve the algorithm in combination with English waveform characteristics. Moreover, this paper uses a semi-supervised learning method to process English speech waveform data, collects relevant data through the English speech input system, and then labels the data and obtains a new English speech data set through training and learning. In addition, this paper uses multiple iterations of labeling to obtain the ideal output data, uses neutrosophic set fuzzy control algorithms and machine learning algorithms to perform English speech feature decomposition and recognition, and uses feature parameter extraction methods to perform signal feature extraction. Finally, this article combines the needs of English speech recognition to build a system model and uses simulation tests to perform performance analysis of English speech feature decomposition and recognition model. The results of the research show that the improved algorithm and system model proposed in this paper have relatively good effects.

1. Introduction

English feature decomposition and recognition is the basic technology of intelligent translation system, and it is also a necessary technical means under the background of world economic integration. Moreover, through English feature decomposition and recognition, the accuracy of English translation can be effectively improved. Therefore, it is necessary to effectively develop speech recognition technology to improve technical reliability.

Feature extraction refers to the processing of the feature parameters of the sound signal by the recognizer, and the feature parameters should have the following characteristics. First of all, the feature parameters are better distinguished, and the sound unit can be accurately modeled. Second, the feature parameters must be robust and can minimize the influence of speakers, channels, and other components to prevent noise interference. Finally, it is necessary to include enough effective information, and the feature dimension is as low as possible to reduce the data size and improve the efficiency of the system. Dimension reduction can be used to remove redundant features to solve multi-space problems. For example, we have two variables: “time spent on the treadmill over a period of time” and “calorie consumption,” which are highly correlated.

At present, the general acoustic characteristics of speech recognition are Mel Frequency Cepstral Coefficient (MFCC) and Perceptual Linear Prediction Coefficient (PLP) [1]. MFCC features are filtered using a special set of “Mel filter groups” in the extraction process and are matched according to the auditory characteristics of human ears. The characteristics of PLP are based on the parameters of the auditory perception model and adapt to the noise environment. Taking into account the dynamic characteristics of the sound signal, it is usually based on its characteristics and its first-order and second-order differences to obtain better performance. In a low-resource environment, traditional shallow features such as MFCC and PLP lack stability and cannot meet the requirements of system modeling. Therefore, in order to obtain more robust feature parameters, traditional shallow features are usually nonlinearly transformed to extract deep features. Since this deep feature has less uncertainty in different environments and different speaker conditions, less training data can be used to build a more robust acoustic model. Nowadays, it is widely used to extract deep features using multi-layer sensors (MLP) or deep neural networks (DN). Currently, the commonly used deep features are tandem features and bottleneck (BN) features [2], which are extracted from the output layer and implicit layer of MLP or DN, respectively. Moreover, tandem features use principal component analysis (PCA) or Karhunen–Loeve transform (KLT) technology to correlate and protect the linear output value or logarithmic detection probability of the MLP or DN output layer, and connect with the original shallow features. The characteristics of BN are extracted through a neural network with a special structure. The network has many hidden layers with a relatively small number of nodes, and the output of the hidden layer is the BN feature. In addition, the tandem feature uses the computing power of the neural network to enhance the difference between the original features, and the BN feature is a powerful nonlinear degradation. Compared with the traditional feature parameters, the experiment proves that the feature parameters of neural network are more discriminative and robust, and show a certain degree of language dependence. The sound signal is a complex time-varying signal with complex and diverse correlations in different time ranges. GMM, SGMM, and DN models are limited to the length of a fixed window and can only model limited time data in the window. However, the recurrent neural network (RNN) has a feedback network structure, the output of the past time is used as a part of the current input, and the current network output result is obtained together with the current input. This mechanism allows the RNN model to use past time information to balance the data correlation of different time ranges. Therefore, compared with the conventional GMM, SGMM, and feed-forward DN, the RNN model can better describe the speech signal. In particular, a new RNN with a long-term storage (long short-term memory, LSTM) structure [3] can effectively overcome problems such as the disappearance of gradients in conventional RNNs and can effectively model long-term sound information.

Based on neutrosophic set fuzzy control technology, this article explores the intelligent technology that can be used for English speech feature decomposition and constructs a corresponding intelligent model.

2. Related Work

Traditional speech recognition acoustic modeling usually adopts an HMM model that estimates the probability distribution of observations in GMM, that is, the GMM-HMM model. However, the GMM-HMM model is independent between each state, and because the parameters are not shared with each other, many training samples are needed to obtain large-scale, accurate, and robust model parameter estimates. Therefore, in a low-resource environment, the GMM-HMM model has a problem of data sparseness, and the estimation of model parameters becomes inaccurate. On the other hand, assisting modeling training with other language data is one of the common methods in low-resource speech recognition, but the GMM model is not enough to distinguish multi-language modeling. Therefore, the GMM model has been widely used and good results in the field of speech recognition, but it is difficult to achieve ideal recognition performance in a low-resource environment [4]. Three levels of natural language understanding are as follows: 1 at the grammatical level, the structure of sentences and phrases is analyzed to find out the relationship between words and phrases and their roles in sentences. 2. Semantic level: find out word meaning, structural meaning, and their combined meaning through analysis, so as to determine the real (actual) meaning or concept expressed by the language. 3. Pragmatic level studies the impact of the external environment where the language is located on language use. It describes the environmental knowledge of language and the relationship between language and language users in a given language environment.

The literature [5] improved the GMM model with a parameter sharing strategy, proposed a subspace Gaussian mixture model, and constructed an SGMM-HMM acoustic model. Unlike traditional acoustic models, the state model parameters in SGMM are located in a parameter subspace, and each state is represented by one or several low-dimensional vectors. In low-resource speech recognition, global parameters can be trained in combination with other language data, and limited target language training data are used to estimate state-related parameters to obtain a more robust acoustic model. When the traditional GMM is directly regularized, the parameter amount of the model can be reduced and a more compact acoustic model can be obtained. For example, the sparse precision matrix modeling method directly imposes sparse restrictions on the inverse matrix of the Gaussian covariance matrix in the model and obtains a bend between the number of parameters and the modeling accuracy, thereby obtaining a better acoustic model in a low-resource environment [6]. The DN model has strong segmentation and robustness. In [7], the use of DN instead of GMM to estimate the posterior probability of observations is proposed. The results show that the recognition rate of the DN-HMM acoustic model system constructed for acoustic modeling is significantly higher than that of the traditional GMM-HMM acoustic model system. The literature [8] used the clustered state (the coupled triphone state) to replace the monophone state as the output unit of the neural network and called the improved model the context-dependent DN-HMM. In addition to DN, other neural network structures are also used for acoustic modeling. The literature [9] used CNN instead of DN for acoustic modeling, which is successfully applied to low-resource speech recognition. The literature [10] used the recurrent neural network structure to construct the RNN-HMM acoustic model. However, its system recognition rate is lower than that of the DN-HMM system.

For low-resource speech recognition features, HMM also has research and improvement. The literature [11] proposed an HMM model based on Kullback–Leibler distance. This model uses KL distance timing to dynamically match the posterior probability of each state and the HMM model unit. The experimental results show that the KL-HMM model can be applied to low-resource speech recognition tasks. On this basis, the literature [12] also combined the DN model with KL-HMM to perform multilingual sound data training, which achieved good recognition performance even when the data resources of the target language were extremely low.

The literature [13] proposed to perform voice conversion based on Gaussian mixture model. This method uses Gaussian mixture model to fit the speaker's spectral envelope vector and uses continuous parameter function to represent the parameters of the training data optimized by the least square method. However, the estimation of the GMM component is not based on the coupled feature vector, but based on the source feature vector. The literature [14] proposed a method called coupled density Gaussian mixture model, which can consider the source space and target space at the same time during training. In addition, the parameters of the transfer function can be directly estimated by the combined GMM to avoid the calculation of large matrix inferences. However, the method based on GMM has the shortcomings that the spectrum is too smooth and the details are not good. The reason is that the parameter relationship between the source speaker and the target conversation is inconsistent. Therefore, the sound quality and similarity of the converted voice are not enough. The literature [15] suggested to use frequency bending to achieve voice conversion, and the frequency bending can obtain a high sound similar to the sound quality of the sound. Then, the literature [16] suggested to use a combination of GMM and least squares to realize speech conversion, which can overcome the over-fitting problem of GMM. In the literature [17], the combined technology of GMM and dynamic frequency folding realizes voice conversion and overcomes the over-fitting problem of traditional GMM. Moreover, it used an adaptive weighted spectral interpolation speech analysis/synthesis model to extract the vowel frequency and spectral envelope of the sound, and it suggested to use weighted residual compensation to improve the personality similarity of the converted speech. The literature [18] also proposed a Gaussian mixture model method based on maximum likelihood estimation. This method uses static feature and dynamic feature statistics as the spectrum conversion sequence to introduce global dispersion characteristics, which reduces the over-smoothing effect and significantly improves the conversion performance.

3. Uncertainty Calculation in Bayesian Modal Parameter Identification Method

Text is composed of every word. When it comes to word vectors, one hot is the simplest word vector, but there are problems such as dimension disaster and semantic gap. By constructing co-occurrence matrix and using SVD to solve construction word vector, the computational complexity is high. Nnlm's so-called distributed hypothesis can be expressed in one sentence. Words in the same context have similar meanings. This leads to word2vec and fastText. Although their essence is still the language model, their goal is not the language model itself, but the word vector. A series of optimizations are made to get word vectors faster and better. The Bayesian method is to quantify the uncertainty of the recognition result through the post-covariance matrix based on the model parameters, and to estimate the inverse of the Hessian matrix of NLFF. When the number of degrees of freedom is n and the number of modes is m, the Hessian matrix is a $n^p = (m+1)^2 + mn$-dimensional square matrix, and the elements in the matrix consist of the second derivative of NLFF with respect to each modal parameter, including [19]

$$
\begin{aligned}
&\{f_i, \xi_i \colon i = 1, \cdots, m\},\\
&\{S_{ii} \colon i = 1, \cdots, m\},\\
&\left\{U_{ij}, V_{ij} \colon i = 1, \cdots, m; j < i\right\}, S_e,\\
&\left\{\Phi_{ij} \colon i = 1, \cdots, n; j = 1, \cdots, m\right\}.
\end{aligned}
\tag{1}
$$

Among them, U_{ij} and V_{ij} represent the real and imaginary parts of $S_{ij}\,(j < i)$, respectively.

$\widehat{H}_L \in R^{n_p \times n_p}$ is the Hessian matrix of the modal parameters at MPV, $\left\{\kappa_1, \kappa_2, \cdots, \kappa_{n_p}\right\}$ is the eigenvalues of $\widehat{H}_L$ in ascending order, and the corresponding eigenvector is $\left\{b_1, b_2, \cdots, b_{n_p}\right\}$. Since NLFF has achieved the minimum value at MPV, the eigenvalues of $\widehat{H}_L$ are all non-negative numbers. In addition, since the mode shapes are all processed as unit vectors in the recognition process, that is, NLFF has nothing to do with the size of the mode shape, if the curvature of $\widehat{H}_L$ in the mode shape direction is zero, the first m eigenvalues of the matrix must be zero, that is, $\kappa_i = 0; i = 1, 2, \cdots, m$. The eigenvectors corresponding to these eigenvalues are

$$
\begin{aligned}
b_1 &= \left[0^{(m+1)^2}; \Phi(1); 0_n; \cdots; 0_n\right],\\
b_2 &= \left[0^{(m+1)^2}; 0_n; \Phi(2); 0_n; \cdots; 0_n\right],\\
&\vdots\\
b_m &= \left[0^{(m+1)^2}; 0_n; \cdots; 0_n; \Phi(m)\right].
\end{aligned}
\tag{2}
$$

Among them, $0_n \in R^n$ represents a zero vector of length n. Then, $\widehat{H}_L$ and its inverse matrix can be expressed in the following form:

$$
\begin{aligned}
\widehat{H}_L &= \sum_{i=1}^{m} 0 \times b_i b_i^T + \sum_{j=m+1}^{n_p} \kappa_j b_j b_j^T,\\
\widehat{H}_L^{-1} &= \sum_{i=1}^{m} 0^{-1} \times b_i b_i^T + \sum_{j=m+1}^{n_p} \kappa_j^{-1} b_j b_j^T.
\end{aligned}
\tag{3}
$$

Among them, the posterior covariance matrix of $[\{f_i\}; \{\xi_i\}; \{S_{ii}\}; \{U_{ij}, V_{ij}\}; S_e]$ corresponds to the part of

$(m+1)^2 \times (m+1)^2$ in the upper left corner of $\widehat{H}_L^{-1}$. The first m singular terms of the above formula are meaningless for this part of the calculation, because [20]

$$\sum_{i=1}^{m} b_i b_i^T = \begin{bmatrix} 0_{(m+1)^2\times(m+1)^2} & 0_{(m+1)^2\times(m+1)^2} \\ \Phi(1)\Phi(1)^T & \\ \ddots & \\ 0_{mn\times(m+1)^2} & \Phi(m)\Phi(m)^T \end{bmatrix}. \quad (4)$$

The posterior covariance matrix $\widetilde{C}$ of the modal parameters can be expressed as

$$\widetilde{C} = \sum_{j=m+1}^{n_p} \kappa_j^{-1} b_j b_j^T. \quad (5)$$

It can be seen that the first m singular items with no computational significance have been eliminated. When the English speech features are framed, the result is shown in Figure 1

$\{\kappa_{ij} \geq 0; j = 1, 2, \cdots, n\}$ represents the eigenvalue of the posterior covariance matrix corresponding to the ith mode shape. Since the mode shape subspace has been excluded by the above formula, $\kappa_{t1} = 0$. The global posterior covariance of the mode shape can be evaluated with a scalar similar to the modal assurance criterion (MAC) [21]:

$$\rho_i = \left(1 + \sum_{j=2}^{n} \kappa_{ij}\right)^{-1/2}. \quad (6)$$

Obviously, ρ_i is a decreasing function of κ_{ij}, $0 \leq \rho_i \leq 1$, and if and only if all κ_{ij} is equal to 0, $\rho_i = 1$. This property makes ρ_i an effective measure to measure the quality of mode shape recognition. The larger ρ_i is, the smaller the uncertainty of the mode shape.

The standardized form EASI program of adaptive blind source separation is

$$B_{t+1} = B_t - \lambda_t \left(\frac{y_t y_t^T - I}{1 + \lambda_t y_t^T y_t} + \frac{g(y_t)y_t^T - (y_t)g(y_t)^T}{1 + \lambda_t \left| y_t^T g(y_t) \right|} \right) B_t. \quad (7)$$

It can be seen from the update formula of EASI that the choice of step size has a great impact on the separation performance and stability of the algorithm, so it needs to be selected carefully. Generally, in the case of incomplete signal separation, the steps should be increased to speed up the update speed of the separation matrix, so that the mixing matrix of the signal can be adapted more quickly. If the separation signal points to the original signal, the step size should be reduced to prevent a few time abnormal signals from having a greater impact on the separation matrix, so as to improve the stability of the algorithm. Therefore, the determination of the signal separation status is to adjust the key of the EASI update step. For the time-varying blind source separation of hybrid systems, the online algorithm is generally used for tracking. The time-varying ability of the tracking system is improved by accelerating the convergence speed of the algorithm. In order to speed up the convergence of the algorithm, the idea of variable step size is proposed. The cost function is used to update the step size adaptively, and a variable step size blind source separation algorithm is proposed. The stability of natural gradient algorithm can be improved by introducing momentum into the learning process of natural gradient. At the same time, an estimation function is used to adjust the step size and momentum factor, which greatly improves the convergence speed of the algorithm.

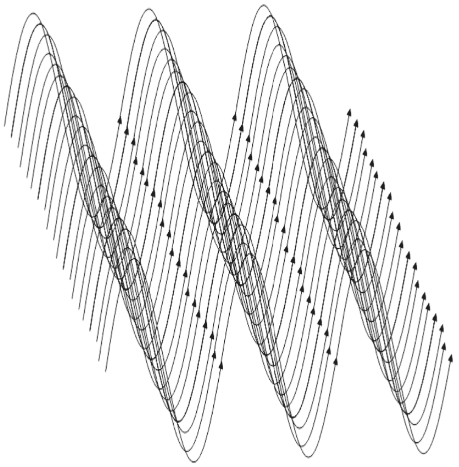

Figure 1: English speech frame processing.

When any serial update algorithm reaches a stable point, the separation matrix should make the mathematical expectation of the objective function equal to zero. Corresponding to the EASI algorithm, this condition can decompose two parts of symmetry and oblique symmetry, namely [22],

$$E\{yy^T\} = I, \quad (8)$$

$$E\{f'(y_t)y_t^T - y_t f'(y_t)^T\} = 0. \quad (9)$$

The matrix is defined as

$$S = \sum_{t=1}^{w} \frac{\left(f'(y_t)y_t^T - y_t f'(y_t)^T\right)}{w}. \quad (10)$$

Among them, S is an approximation to equation (10), and w is the selected data length, which is mainly used to estimate mathematical expectations. After calculating S, the modulus of each element of the matrix is calculated separately, and the largest modulus is selected as the separation index of this segment of signal, namely,

$$SI = \max(\text{abs}(S)). \quad (11)$$

Among them, SI is the separation index. In an ideal situation, S should be a zero matrix, and the greater the difference from the zero matrix, the more incomplete the separation. Therefore, the maximum value of each element pattern of the matrix can reflect the signal separation status and serve as a separation index. In the initial stage of program operation, the degree of separation of mixed signals is still very low, so the separation index is also very large. However, as the action of the program continues, the separated signal approaches the source signal. That is, the higher the degree of signal separation, the smaller the value of the separation index. Therefore, as a whole, as the separation process progresses, the separation index gradually decreases and shows a stable trend. In the above iterative update algorithm, the update step size has a great impact on the results. If the step size is too large, it is easy to produce oscillation or even divergence; if the step size is too small, the convergence will be too slow. It requires too many samples and requires a large amount of calculation. The simulated annealing strategy is very simple, but it can effectively combine the advantages of the above two steps, rapidly decline in the initial stage, and accurately converge in the later stage. The results show that it is simple and effective to use simulated annealing strategy to adjust the learning rate in this experiment.

The step size selection strategy of the EASI program is an important parameter that controls the size of each update of the separation matrix. If the step is too large, the convergence speed of the program will of course become faster, but the stability error will also become larger, which will also cause the program to diverge. On the contrary, if the step size is small, the update of the separation matrix is too slow, and the program will lose the role of real-time monitoring of input data. Therefore, in order to better balance the convergence speed and the stability error, the step size should be adjusted adaptively according to the separation index. In this article, we use the nonlinear mapping relationship to adjust the step size according to the separation index [23]:

$$\lambda(t) = \beta \tanh\{\alpha[SI(t) - \delta]\} + \gamma. \tag{12}$$

Among them, tanh represents the hyperbolic tangent function, α is the shape parameter, β is the scale parameter, and δ represents the information of SI and should be equal to half of the maximum value of SI under ideal circumstances. γ is a supplementary parameter to ensure that when $SI = 0$, $\lambda(t) = 0$. Therefore, the expression of γ can be obtained as

$$\gamma = -\beta \tanh(-\alpha \cdot \delta). \tag{13}$$

Through the above algorithm, the waveform signal processing of English speech is performed. The result before processing is shown in Figure 2 and the result after processing is shown in Figure 3.

The significance of using the mapping relationship based on the hyperbolic tangent function is firstly a monotonically increasing nonlinear function, and the larger the separation index, the more suitable the steps are. Second, the adjustment of the function curve shape is very convenient, and it only needs to change the relevant parameters. Third, the hyperbolic tangent function has boundaries. Sometimes, even if the separation index is too large, the steps are always limited to a certain range to prevent calculation divergence. The final result can be obtained through the membership degree of different language attributes (which cannot be directly divided by numerical values). For example, there are different methods for different people to control language pronunciation, which cannot be indicated by certain data, but can only be expressed by investigating and determining a function.

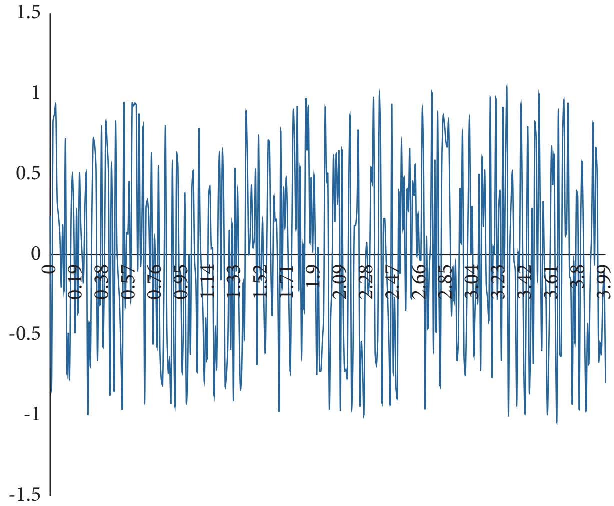

Figure 2: English speech waveform before processing.

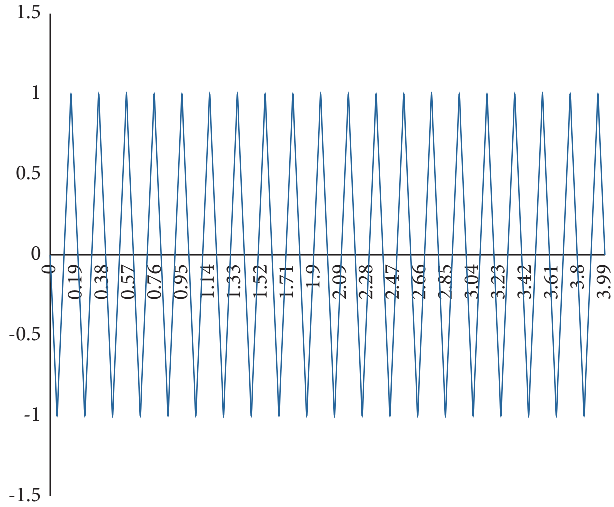

Figure 3: English speech waveform after processing.

With the action of the EASI program, new output signals continue to occur, and the separation matrix is constantly updated. If the input signal changes from a certain point in time, after a certain period of update, the change is reflected in the separation matrix. This is the theoretical basis for using EASI to detect structural changes.

EASI has a global matrix C_t, which is equal to the multiplication of the mixing matrix A and the separating matrix B_t:

$$C_t = B_t A. \tag{14}$$

In an ideal situation, the global matrix should converge to the identity matrix. In the structural modal recognition area based on blind source separation, the hybrid matrix A is the modal vibration type of the structure. Therefore, only by

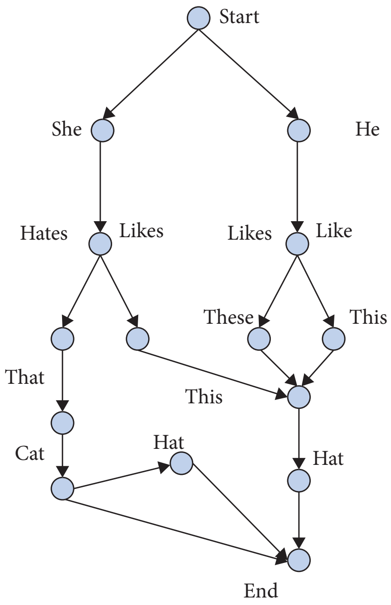

Figure 4: Decomposition diagram of English speech features.

multiplying the real-time updated separation matrix with the structural oscillator matrix, the difference between the unit matrix and the unit matrix can be re-examined to obtain performance indicators for quickly judging the real-time structural condition. However, because the true vibration shape of the structure is not known, it is assumed here that the vibration type obtained by the fast Bans FFT pattern recognition method is the true vibration shape of the structure, namely,

$$C_t = B_t \Phi. \tag{15}$$

At this time, the performance index is defined as

$$PI = \sum_i \left(\sum_j \frac{|C_{ij}|}{\max_k |C_{ik}|} - 1 \right) + \sum_j \left(\sum_i \frac{|C_{ij}|}{\max_k |C_{kj}|} - 1 \right). \tag{16}$$

Among them, C_{ij} is the (i, j)th element of the global matrix, and $\max_k |C_{ik}|$ represents the maximum absolute value of the ith row element in C. Similarly, $\max_k |C_{kj}|$ represents the maximum absolute value of the element in the jth column of C. Obviously, the closer the PI is to 0, the closer the global matrix C is to the identity matrix, that is, the better the separation effect.

Figure 4 shows an example of the word graph structure generated by the speech “he likes this hat.”

After running for a certain period of time, the EASI program converges, the PI is close to 0, and the value changes are stable. When the structural parameters change suddenly, the original convergence point cannot meet the stability conditions of EASI, and the global matrix is not an approximation of the identity matrix, but a continuous abnormal rise in the value of PI. Therefore, the change status of PI can reflect the structural status, and when the PI value exceeds the threshold, a more comprehensive system identification method (i.e., FFT pattern recognition method) needs to be used for in-depth analysis of the structural response data.

The main idea of the multi-scale online mode analysis method based on blind source separation and Bayes is to use EASI to monitor the structural response in real time. If the PI value is not abnormal, no additional operations are required. When the PI value becomes abnormally high and exceeds

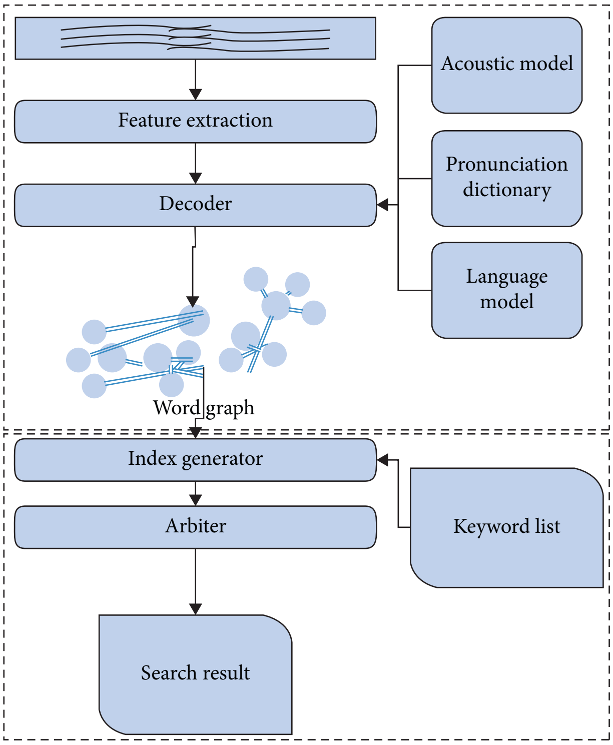

Figure 5: English speech keyword retrieval system.

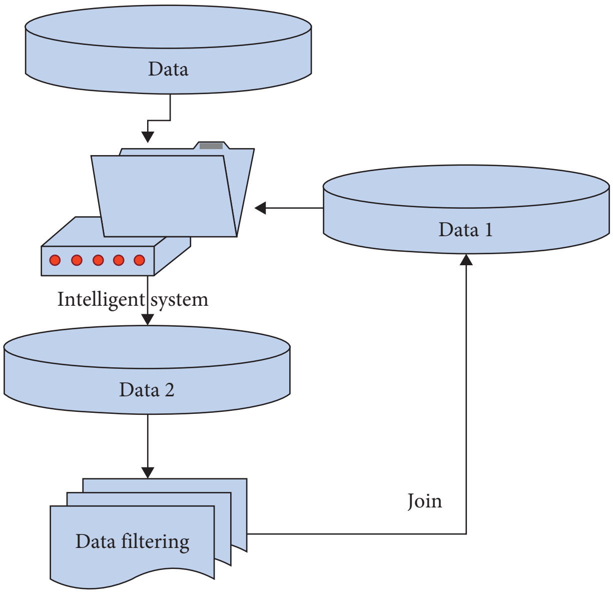

Figure 6: Application of semi-supervised learning in English speech recognition system.

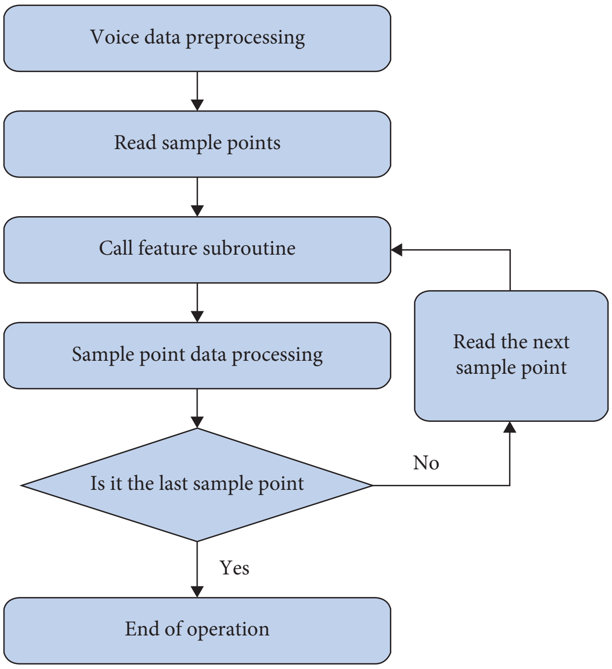

Figure 7: Flowchart of feature decomposition and recognition of English speech.

the limit value, the high-speed Bans FFT pattern recognition method is used to identify the structure frequency, attenuation, vibration type, and corresponding uncertainty. This method generally only performs conventional monitoring and obtains a small amount of information such as structural vibration type or mode response. When monitoring possible changes in the structure, it is called a multi-scale online mode analysis method due to the in-depth evaluation of structural mode parameter information. The implementation steps of this method are as follows.

(1) The structural response in the environmental stimulus is measured.

(2) The selected frequency band is executed by Bayes' fast FFT pattern recognition algorithm, and the pattern parameters accompanying the uncertainty estimation are obtained.

(3) The EASI program is executed, the structure response data are processed in real time, and the separation matrix is updated.

(4) The mode calculated in step (2) is set as the true vibration mode of the system, and it is multiplied by the separation matrix updated in real time in step (3) to obtain a global matrix to calculate PI.

(5) When the PI value exceeds the set threshold, the algorithm returns to step (2). Otherwise, the algorithm will return to step (3).

4. English Speech Feature Recognition Model Based on Neutrosophic Set Fuzzy Control

On the basis of the above analysis, the English speech feature recognition model is carried out, and the decomposition algorithm mentioned above is used as the system fitting algorithm. The basic structure of the English speech keyword retrieval system constructed is shown in Figure 5.

In this paper, the learning and training of English speech recognition is realized through semi-supervised learning, and semi-supervised training is also carried out through iterative methods. After collecting relevant data through the English speech input system, the data are labeled, and a new English voice data set is obtained through training and learning. After multiple iterations of labeling, the ideal output data are finally obtained. The process is shown in Figure 6.

When using neutrosophic set fuzzy control algorithm and machine learning algorithm for English speech feature decomposition and recognition, signal feature extraction is mainly performed by feature parameter extraction. After extracting the features of the speech signal, the corresponding iterative processing is performed to obtain a digitized signal that can be recognized by the system. Then, through conversion, the recognized signal becomes the output result that people can intuitively observe and hear. The English speech feature decomposition and extraction process is shown in Figure 7.

Through the above analysis, an English speech feature decomposition and recognition system based on neutrosophic set fuzzy control is constructed. On this basis, system performance verification analysis can be carried out. In the process of recognizing English speech features, the system first needs to digitize the speech signal and decompose the speech signal to make the signal a recognizable result of the system. Therefore, in the experimental research, the effect of system speech signal decomposition is studied through simulation research. A total of 78 sets of

Table 1: Statistical table of the accuracy rate of English speech digitization decomposition.

Num	Speech decomposition(%)	Num	Speech decomposition(%)	Num	Speech decomposition(%)
1	93.83	27	94.15	53	91.81
2	93.77	28	95.97	54	92.64
3	91.35	29	95.49	55	94.49
4	95.13	30	94.40	56	93.68
5	91.69	31	95.48	57	91.79
6	93.10	32	93.28	58	92.08
7	93.29	33	93.62	59	93.75
8	94.68	34	93.00	60	92.60
9	95.90	35	92.36	61	92.01
10	93.17	36	91.09	62	95.11
11	93.35	37	94.55	63	93.98
12	92.80	38	92.87	64	95.81
13	94.91	39	93.06	65	91.87
14	94.85	40	94.07	66	93.53
15	91.48	41	94.54	67	94.72
16	94.00	42	92.09	68	93.35
17	91.82	43	95.02	69	93.59
18	94.33	44	94.05	70	94.23
19	95.77	45	94.13	71	92.18
20	92.03	46	91.31	72	93.11
21	91.66	47	93.73	73	95.72
22	94.81	48	93.91	74	92.05
23	94.96	49	93.30	75	93.73
24	93.41	50	93.27	76	94.82
25	95.01	51	94.41	77	92.38
26	94.48	52	92.64	78	94.88

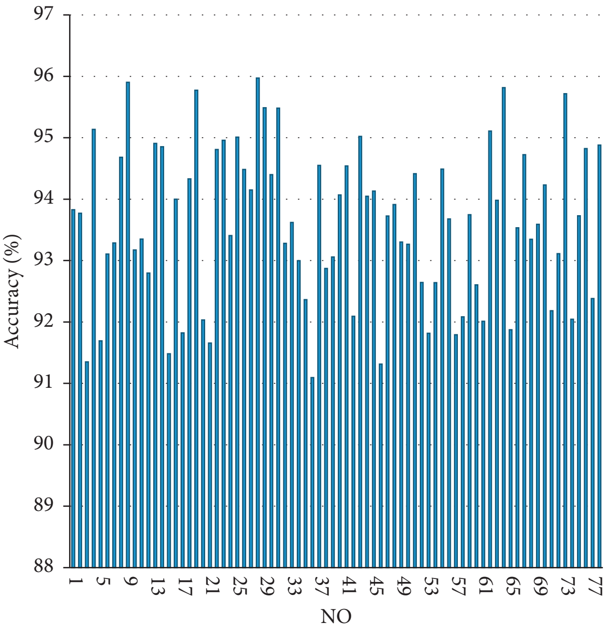

Figure 8: Statistical diagram of the accuracy rate of English speech digitization decomposition.

experiments are carried out to calculate the accuracy of speech digitization decomposition. The results are shown in Table 1 and Figure 8.

From the above experimental statistical results, the English speech feature recognition system based on neutrosophic set fuzzy control constructed in this paper has a good effect of English speech digitization. The system digitizes and decomposes English speech and then recognizes its features, which is also the core technology of intelligent translation. This paper uses simulation research to analyze the system's speech feature recognition effect, and the results are shown in Table 2 and Figure 9.

From the above statistical analysis, it can be seen that the English speech feature decomposition and recognition

Table 2: Statistical table of the recognition effect of English speech features.

Num	Speech recognition(&)	Num	Speech recognition(&)	Num	Speech recognition(&)
1	85.98	27	88.66	53	91.83
2	85.92	28	84.23	54	90.10
3	88.62	29	91.07	55	89.33
4	84.83	30	89.95	56	86.08
5	89.51	31	88.97	57	88.55
6	89.00	32	92.76	58	84.30
7	90.04	33	86.54	59	89.94
8	88.26	34	90.19	60	87.15
9	90.29	35	85.90	61	85.16
10	90.06	36	91.28	62	89.00
11	92.58	37	86.04	63	85.02
12	88.82	38	91.06	64	90.08
13	86.03	39	86.13	65	91.21
14	86.33	40	86.43	66	89.72
15	84.55	41	88.75	67	91.39
16	88.13	42	92.51	68	84.31
17	90.17	43	91.69	69	88.58
18	84.37	44	92.99	70	87.87
19	88.87	45	87.11	71	87.79
20	91.65	46	89.30	72	91.01
21	90.41	47	89.93	73	91.38
22	90.03	48	84.34	74	89.48
23	86.63	49	91.94	75	85.34
24	90.43	50	85.77	76	90.57
25	89.36	51	87.12	77	85.83
26	89.86	52	90.28	78	84.61

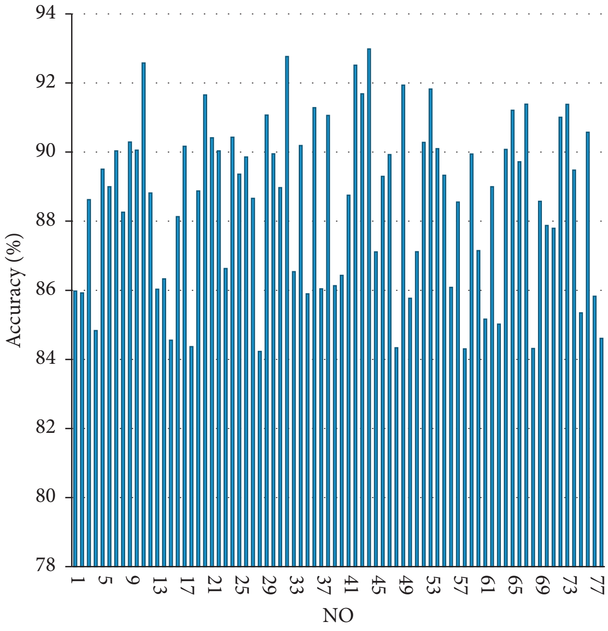

Figure 9: Statistical diagram of the recognition effect of English speech features.

system based on neutrosophic set fuzzy control constructed in this paper has a certain effect.

5. Conclusion

Based on neutrosophic set fuzzy control technology, this article explores the intelligent technology that can be used for English speech feature decomposition and constructs a corresponding intelligent model. The learning and training of English speech recognition is realized through semi-supervised learning, and semi-supervised training is also carried out through iterative methods. Moreover, after collecting relevant data through the English voice input system, the data are labeled, and a new English voice data set is obtained through training and learning. After multiple iterations of labeling, the ideal output data are finally obtained. In the process of English speech feature recognition, the system constructed in this paper first needs to digitize the speech signal and decompose the speech signal to make the signal a recognizable result of the system. Compared with the traditional feature parameters, the feature parameters of the neural network proposed in this paper have better recognition and robustness, and show certain language dependence. Sound signal is a complex time-varying signal, which has complex and diverse correlations in different time ranges. GMM, sgmm, and DN models are limited to fixed window lengths and can only model limited time data in windows. Therefore, the new RNN with long-term storage (long short-term memory, LSTM) structure can effectively overcome the problems of gradient disappearance in traditional RNN and can effectively model long-term sound information. The performance of the system is verified and analyzed through experimental research. From the results of experimental research, it can be seen that the English speech feature recognition system based on neutrosophic set fuzzy control constructed in this paper has a good effect of English speech digitization decomposition and speech feature recognition.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by school-level classroom teaching reform project of Jiaxing University, classroom reform and practice of Blended Teaching in Business English major courses under the background of "Internet +," Project no. 85152026.

References

[1] J. Cai, J. Luo, S. Wang, and S. Yang, "Feature selection in machine learning: a new perspective," *Neurocomputing*, vol. 300, no. 2, pp. 70–79, 2018.

[2] J. N. Goetz, A. Brenning, H. Petschko, and P. Leopold, "Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling," *Computers & Geosciences*, vol. 81, no. 1, pp. 1–11, 2015.

[3] H. Darabi, B. Choubin, O. Rahmati, A. Torabi Haghighi, B. Pradhan, and B. Kløve, "Urban flood risk mapping using the GARP and QUEST models: a comparative study of machine learning techniques," *Journal of Hydrology*, vol. 569, no. 5, pp. 142–154, 2019.

[4] A. Rajkomar, J. Dean, and I. Kohane, "Machine learning in medicine," *New England Journal of Medicine*, vol. 380, no. 14, pp. 1347–1358, 2019.

[5] Y. Xin, L. Kong, Z. Liu et al., "Machine learning and deep learning methods for cybersecurity," *IEEE Access*, vol. 6, no. 1, pp. 35365–35381, 2018.

[6] L. Ward, A. Agrawal, A. Choudhary, and C. Wolverton, "A general-purpose machine learning framework for predicting properties of inorganic materials," *Npj Computational Materials*, vol. 2, no. 1, p. 16028, 2016.

[7] P. Feng, B. Wang, D. L. Liu, C. Waters, and 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, no. 3, pp. 100–113, 2019.

[8] K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis, and D. I. Fotiadis, "Machine learning applications in cancer prognosis and prediction," *Computational and Structural Biotechnology Journal*, vol. 13, no. 5, pp. 8–17, 2015.

[9] S. Amershi, M. Cakmak, W. B. Knox, and T. Kulesza, "Power to the people: the role of humans in interactive machine learning," *AI Magazine*, vol. 35, no. 4, pp. 105–120, 2014.

[10] V. Rodriguez-Galiano, M. Sanchez-Castillo, M. Chica-Olmo, and 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, no. 3, pp. 804–818, 2015.

[11] C. W. Coley, R. Barzilay, T. S. Jaakkola, W. H. Green, and K. F. Jensen, "Prediction of organic reaction outcomes using machine learning," *ACS Central Science*, vol. 3, no. 5, pp. 434–443, 2017.

[12] A. Chowdhury, E. Kautz, B. Yener, and D. Lewis, "Image driven machine learning methods for microstructure recognition," *Computational Materials Science*, vol. 123, no. 8, pp. 176–187, 2016.

[13] S. Basith, B. Manavalan, T. H. Shin, and G. Lee, "SDM6A: a web-based integrative machine-learning framework for predicting 6mA sites in the rice genome," *Molecular Therapy - Nucleic Acids*, vol. 18, no. 6, pp. 131–141, 2019.

[14] C. Voyant, G. Notton, S. Kalogirou et al., "Machine learning methods for solar radiation forecasting: a review," *Renewable Energy*, vol. 105, no. 2, pp. 569–582, 2017.

[15] C. Folberth, A. Baklanov, J. Balkovič, R. Skalský, N. Khabarov, and M. Obersteiner, "Spatio-temporal downscaling of gridded crop model yield estimates based on machine learning," *Agricultural and Forest Meteorology*, vol. 264, no. 4, pp. 1–15, 2019.

[16] J. Sieg, F. Flachsenberg, and 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, 2019.

[17] F. Thabtah and 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, 2020.

[18] F. A. Narudin, A. Feizollah, N. B. Anuar, and A. Gani, "Evaluation of machine learning classifiers for mobile malware detection," *Soft Computing*, vol. 20, no. 1, pp. 343–357, 2016.

[19] Q. Yao, H. Yang, R. Zhu et al., "Core, mode, and spectrum assignment based on machine learning in space division multiplexing elastic optical networks," *IEEE Access*, vol. 6, no. 6, pp. 15898–15907, 2018.

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

[21] M. Chen, Y. Hao, K. Hwang, L. Wang, and L. Wang, "Disease prediction by machine learning over big data from healthcare communities," *IEEE Access*, vol. 5, no. 1, pp. 8869–8879, 2017.

[22] L. Itu, S. Rapaka, T. Passerini et al., "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, 2016.

[23] U. Jayasinghe, G. M. Lee, T.-W. Um, and Q. Shi, "Machine learning based trust computational model for IoT services," *IEEE Transactions on Sustainable Computing*, vol. 4, no. 1, pp. 39–52, 2019.