APPLICATIONS AND PROSPECTS OF ARTIFICIAL INTELLIGENCE IN LINGUISTIC RESEARCH

Shaohua Jiang

- School of Humanities, Fujian University of Technology, Fuzhou, Fujian, 350118, China
- Krirk University, Bangkok, 10220, Thailand
- sophia_FP@126.com

Zheng Chen*

- Concord University College, Fujian Normal University, Fuzhou, Fujian, 350000, China

Reception: 2 January 2024 | **Acceptance**: 22 January 2024 | **Publication**: 19 February 2024

Suggested citation:

Jiang, S. and Chen, Z. (2024). **Applications and Prospects of Artificial Intelligence in Linguistic Research.** *3C Tecnología. Glosas de innovación aplicada a la pyme 13(1)*, 57-76.

https://doi.org/10.17993/3ctecno.2024.v13n1e45.57-76

ABSTRACT

In modern linguistic research, the application of Artificial Intelligence has led the field and provided powerful tools and prospects for linguists. LSTM is used for extracting character features, joint vector representation and constructing text generation models and generating natural language text. LSTM is involved in the design of speech recognition network to process the input speech signals for generators and discriminators to improve the accuracy of speech recognition. By continuously optimizing the training objectives, the translation system will more accurately translate text from one language to another, thus facilitating cross-cultural communication. Through the application of artificial intelligence, the F1 value has been improved by 3.9% compared with the previous value, and the cumulative variance contribution rate of the five factors is more than 60%, with all subloadings reaching 0.4 or more. Artificial intelligence will promote the development of the field of linguistics, improve research efficiency and accuracy, and promote the innovation of language technology.

KEYWORDS

Artificial intelligence; LSTM; joint vector; speech recognition; F1 value

https://doi.org/10.17993/3ctecno.2024.v13n1e45.57-76

INDEX

ABSTRACT **2**
KEYWORDS **2**
1. INTRODUCTION **4**
2. LITERATURE REVIEW **4**
3. APPLICATION OF LSTM IN LINGUISTICS **5**
3.1. Application of LSTM in text analysis 6
3.1.1. Extracting character features 6
3.1.2. Joint vector representation 7
3.1.3. LSTM cell structure 7
3.2. Role of LSTM in processing speech signals 8
3.2.1. Speech Recognition Network Design 8
3.2.2. nput Speech Processing 10
3.2.3. Generator 10
3.2.4. Discriminators 11
3.2.5. Training objectives 12
3.3. Creating text generation models using LSTMs 12
4. PROSPECTIVE ANALYSIS OF ARTIFICIAL INTELLIGENCE IN LINGUISTIC RESEARCH **14**
4.1. Quality of translation in different languages 14
4.2. Identification accuracy 16
4.3. Validation of creative writing skills 17
5. CONCLUSION **17**
ACKNOWLEDGMENTS **18**
REFERENCES **18**
ABOUT THE AUTHOR **19**

https://doi.org/10.17993/3ctecno.2024.v13n1e45.57-76

1. INTRODUCTION

Perhaps because of the wider range of application scenarios, or because of the input from established technology companies such as Google, Baidu, and Tech Data, speech recognition technology is often the concept of artificial intelligence technology that comes to mind [1]. Speech recognition technology is indeed used in a large number of specific scenarios in the language learning process [2]. However, if the role and value of speech recognition technology is not understood accurately enough, it would be biased to even expect that relying on speech recognition technology can solve the challenges of language learning intelligence [3]. The key to speech recognition is recognition. No matter how high the recognition degree and accuracy, the ultimate goal is to recognize what the learner has said and display the specific text. This function and process, however, is not strictly pedagogical [4]. That is to say, the result of the recognition is simply a textual result, and it does not yet address the really important matter of how to improve the quality of what is being said. For language learners, the required results and value are much greater than for conventional translation tools [5]. This means that even if 100% recognition accuracy can be achieved, at best it will enable fast and accurate translation or presentation, and will not provide learners with methods and suggestions for learning and improvement [6].

In this paper, LSTM network is used to extract character level features from text data to capture important information and patterns in the text. LSTM is used to create joint vector representations and the structure and functionality of LSTM units are described. LSTM network is used to design speech recognition system to recognize and understand the speech content in the speech signal. Generators and discriminators are used in speech signal processing to improve the recognition accuracy and STM network is used to achieve the training objectives to improve the performance and effectiveness of speech signal processing. The present generative model is created to be used for tasks such as natural language generation. In addition, the innovation of this paper is the use of LSTM networks to create a text generation model, which is potentially valuable for natural language generation tasks. This model can be used to generate natural language text such as articles, comments, or conversations, which is expected to have a wide range of applications in the field of automated writing and chatbots.

2. LITERATURE REVIEW

Rasulova, Z emphasizes the importance of studying the processes and mechanisms of translation, referring to the methodological and psychologist's view that the issue of translation skills and their formation has an important place in translation theory and practice. It is shown that when studying translation, it is important to focus not only on the outcome of the translation, but also to delve into the skills and strategies of the translator and how these skills are formed [7]. Braithwaite, B suggests that there is a rapidly growing scholarly interest in sign languages of the

https://doi.org/10.17993/3ctecno.2024.v13n1e45.57-76

Global South, especially those emerging in small sign language communities. Neutral theoretical constructs about these communities and sign languages may be too abstract and may lead to a tendency to exoticize and objectify research by ignoring the actual needs and concerns of community members [8]. Bafoevna, N. D et al. point out that theological linguistics emerged partly due to the fact that religions have an important place in the social consciousness and are an integral part of any culture. Therefore, if the religious factor is ignored, the study of language will appear incomplete and may even become unfeasible in some cases [9]. Mizumoto, A et al. point out that in the field of corpus linguistics, the application of RS/MA has been very limited and confined to very few subfields. Given that corpus linguistics covers a wide range of issues, meta-analysis is considered to have great potential as a method for systematically synthesizing research results in the field [10]. Su, H et al. proposed a local grammar approach to the study of non-synchronous discourse behavior in academic texts, aiming to provide a new avenue for the study of non-synchronous academic discourse. The local grammar approach captures the realization patterns of discourse acts at both the lexico-grammatical and discourse semantic levels, which helps to understand how the realization of a particular discourse act varies across time and contexts [11]. Awad Al-Dawoody et al. selected a corpus of 60 randomly selected research articles and used them according to Hyland's classification of metadiscourse markers, using the AntConc.3.2.4 for qualitative and quantitative analysis. It was found that there is a gap between Egyptian and Saudi researchers in the use of different metadiscourse markers [12]. Chen, L et al. analyzed by binary logistic regression based on a corpus that recently published articles were more likely to express surprises triggered by a priori knowledge as compared to earlier published articles. These results can be explained by the fact that surprises are heuristic in nature and also by the pressure of academics in strategically promoting their research directions [13]. Umarova, N. R discusses conceptual terminology which is the most active and controversial terminology in modern linguistics, with a focus on the importance of concepts and their linguisticization in the way that language perceives the world, and expresses the national and cultural characteristics of the language. Cognitive approach is one of the methods of recognizing and explaining natural phenomena related to language through language. Cognitive linguistics is a discipline that studies human cognitive activity. Its main aim is to determine the involvement and share of the language system in the process of recognizing the world [14]. Hamzah, M. H et al. objective was to conduct a linguistic literature review of the aboriginal languages of Malaysia, using a systematic evaluation approach and focusing on the three main aboriginal groups of Peninsular Malaysia. The study covered linguistic subfields such as phonology, morphology, sociolinguistics, syntax, semantics, vocabulary and grammar. Further linguistic research is clearly necessary to protect and preserve these languages [15].

3. APPLICATION OF LSTM IN LINGUISTICS

Artificial Intelligence, and in particular LSTMs, are crucial for understanding and processing natural language. LSTMs are a special type of recurrent neural network

https://doi.org/10.17993/3ctecno.2024.v13n1e45.57-76

especially suited for processing and predicting sequential data. In linguistics, this means being able to efficiently process sequences of words, understand sentence structure, and even entire texts. Language contains complex long-term dependencies, for example the subject of a sentence may influence the verb form at the end of the sentence. LSTM is important because it can capture these long-term dependencies better than traditional RNNs [16]. This is crucial for understanding the meaning of text, for language generation and translation. Another advantage of LSTM is its ability to store and process large amounts of historical information, different languages have different grammatical structures and expression conventions, the flexibility of LSTM makes it a powerful tool for understanding and processing multiple languages.

3.1. APPLICATION OF LSTM IN TEXT ANALYSIS

3.1.1. EXTRACTING CHARACTER FEATURES

In natural language processing, CNNs are often used to extract text features, and some researchers have found that using CNNs to extract character-level features can represent the morphological features of words well [17]. Figure 1 shows the network structure for extracting character features in the model of this paper, for example, suyimen is the Latin Viennese word for I like. In this paper, the character vector dimension is set to 30 and is randomly initialized. The maximum character length of each word is 50, if the maximum length is exceeded, the first 50 letters are intercepted, and if the length is less than 50, Padding is used to make up. The character feature representation vectors of the words are extracted through the convolutional and maximum pooling layers. The size of the convolution kernel is 30 and the length of the convolution kernel is 3.

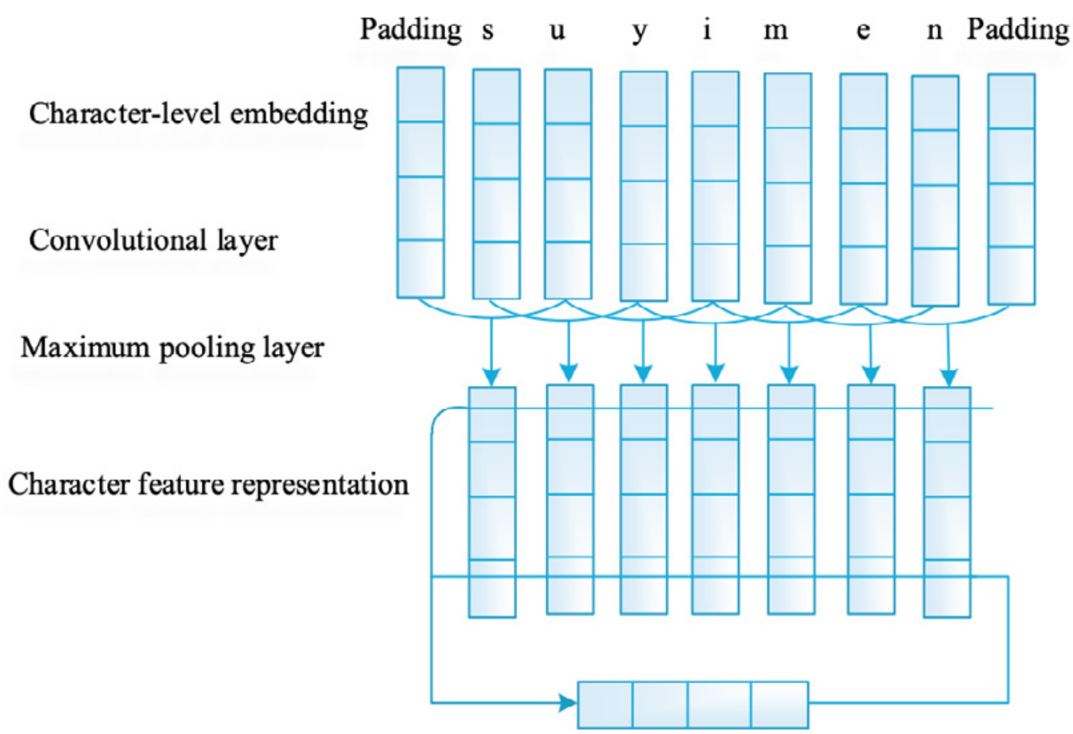

Figure 1 Character feature extraction

https://doi.org/10.17993/3ctecno.2024.v13n1e45.57-76

3.1.2. JOINT VECTOR REPRESENTATION

The cascade of word vectors, character feature vectors, and linguistic feature vectors is used as the input vector representation of the neural network. Assuming that V_{word} denotes the word vector, V_{char} denotes the character feature vector, and V_{fi} denotes the i th linguistic feature vector, the overall input vector can be represented as $V=\left[V_{\text{word}}:V_{\text{char}}:V_{f1}:\cdots:V_{f10}\right]$. The joint feature result is shown in Fig. 2.

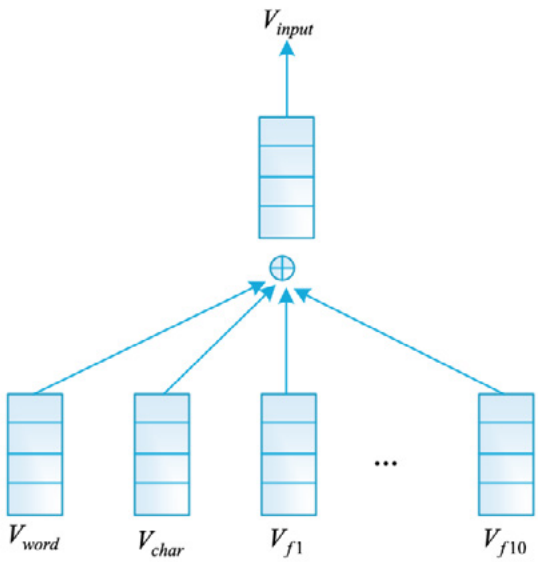

Figure 2 Joint feature representation

3.1.3. LSTM CELL STRUCTURE

Figure 3 shows the basic structure of an LSTM cell, which controls the input and output information through three special gate structures [18]

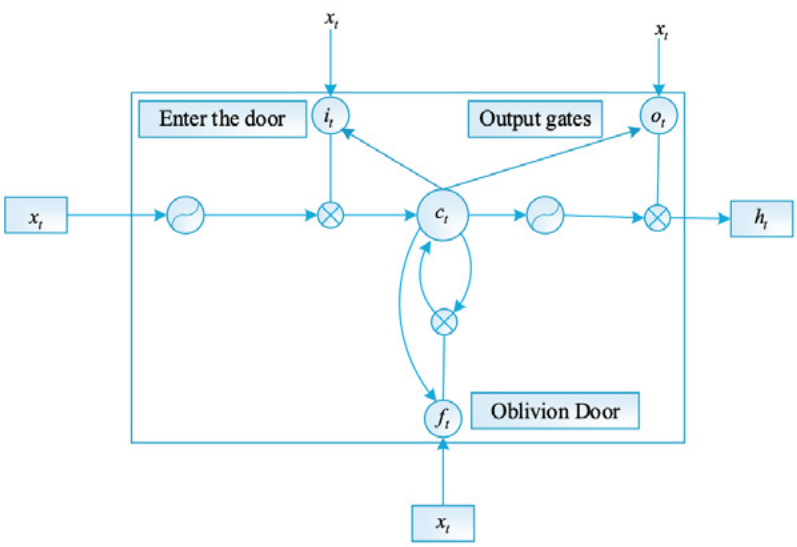

Figure 3 LSTM cell structure

https://doi.org/10.17993/3ctecno.2024.v13n1e45.57-76

$$i_t = \sigma\left(W_{xx}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i\right) \quad (1)$$

$$f_t = \sigma\left(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f\right) \quad (2)$$

$$c_t = f_t c_{t-1} + i_t \tanh\left(W_{xc}x_t + W_{hc}h_{t-1} + b_c\right) \quad (3)$$

$$o_t = \sigma\left(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_{t-1} + b_o\right) \quad (4)$$

$$h_t = o_t \tanh\left(c_t\right) \quad (5)$$

where σ is the Sigmoid activation function, *i* is the input gate, f is the forgetting gate, *c* is the memory cell, *o* is the output gate, *h* is the hidden layer, *tanh* denotes the hyperbolic tangent activation function, W is the weight matrix, e.g., W_{xi} is the weight matrix between the inputs *x* and the input gate, W_{hi} is the weight matrix from the hidden layer to the input gate, and *b* is the bias vector.

3.2. ROLE OF LSTM IN PROCESSING SPEECH SIGNALS

3.2.1. SPEECH RECOGNITION NETWORK DESIGN

Under the assumption that speech and noise are independent of each other, the speech signal and the noise signal are superimposed to form a mixed speech signal Z_t, and then the mixed speech signal is transformed into a two-dimensional time-frequency signal $Y \in {}^{\sim M \times N}$ by a short-time Fourier transform, and then the spectral coefficients of the speech are deduced, where M denotes the time frame corresponding to the speech and N denotes the frequency. The spectrum $\hat{Y}^{j} \in \sim M \times N$ of the speech signal is obtained by the following equation:

$$\hat{Y}^{j} = Y \otimes M^{j} \quad (6)$$

where $\otimes$ denotes the inner product of matrix elements and $M^{j} \in {}^{\sim M \times N}$ is called the time-frequency mask, the time-frequency mask value characterizes the interrelationships between different sources in a mixed signal, such as the target and interfering speakers in speech separation, and the time-frequency mask M^j is estimated by using Wiener filtering of the $\alpha-$ power amplitude spectrum, with the following equation:

$$M^{j} = \frac{\left|\hat{Y}^{j}\right|^{\alpha}}{\sum_{j}\left|\hat{Y}^{j}\right|^{\alpha}} \quad (7)$$

where $|*|$ denotes the absolute value of the matrix and α is an index chosen based on the probability distribution of the hypothesized speech, which is taken as 0.5 in this paper.

https://doi.org/10.17993/3ctecno.2024.v13n1e45.57-76

The generative adversarial network structure consists of two parts, the generator (G) and the discriminator (D). In this paper, we propose a learnable time-frequency mask generator that introduces a recursive derivation algorithm with a neural network structure and another sparse coding layer for generating the time-frequency mask M^j. In particular, the generator consists of a multilayer recurrent neural network (RNN) and a sparse coding layer, the RNN outputs to the sparse coding layer, and the output of the sparse coding layer is the corresponding time-frequency mask M. The method eliminates the need for subsequent processing such as signal filtering, and there is no need for manually defining the number of layers of the neural network.

The generative adversarial network is shown in Fig. 4, where the generator acts as an encoder ($RNN_{dec)}$ through a layer of bi-directional recurrent neural network, a layer of recurrent neural network as decoding, and a layer of feed-forward neural network as a sparse coding layer. The output of the sparse coding layer is time-frequency masked M^j, which is then multiplied by the matrix elements with the mixed signal to obtain the target speech signal. The discriminator consists of a layer of feed-forward neural network encoder and a layer of feed-forward neural network decoder FNN_{dec} and outputs as values in the interval [0,1]. The generator and the discriminator are iteratively optimized to obtain the optimal time-frequency mask M^j, which is used to estimate the amplitude spectrum of the target speech signal, and then combined with the phase spectrum of the mixed signal to reconstruct the time-domain signal with a short-time Fourier inverse transform.

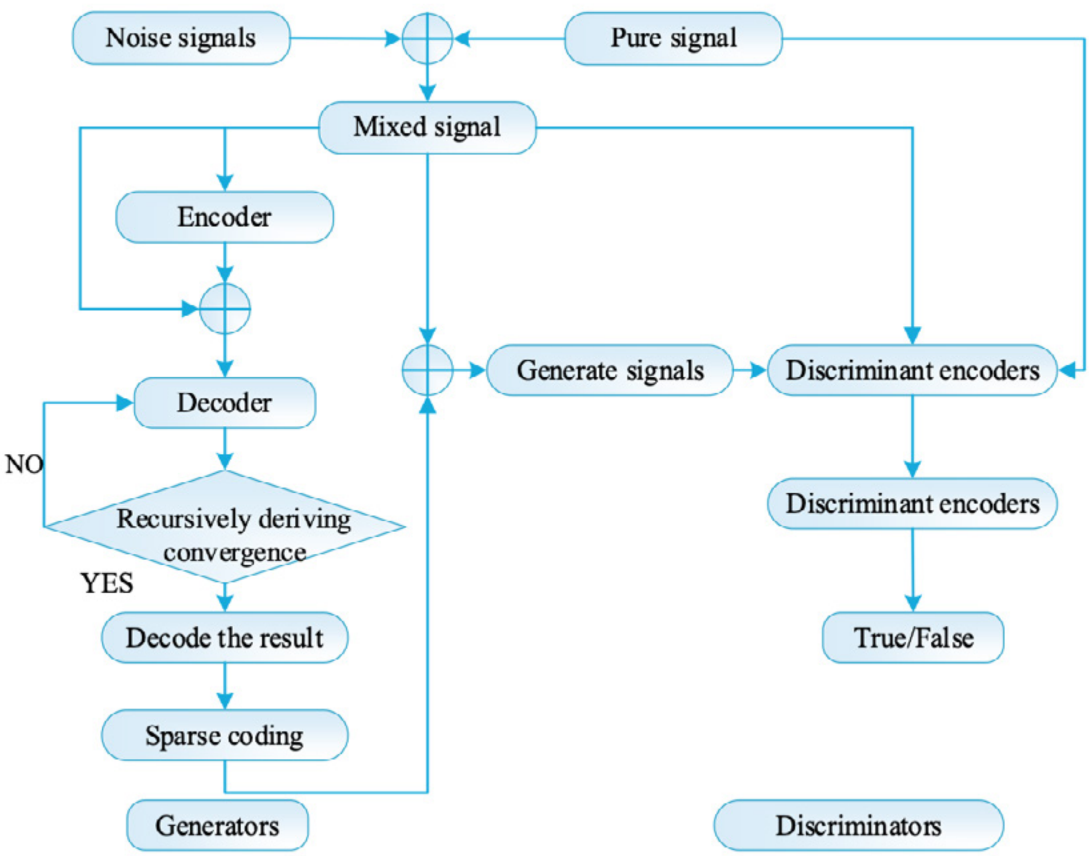

Figure 4 Generates adversarial network structure

https://doi.org/10.17993/3ctecno.2024.v13n1e45.57-76

3.2.2. NPUT SPEECH PROCESSING

Let Z_t be the speech time-domain signal sampled at 44.1kHz and mixed with 0dB signal-to-noise ratio, and Z_t be converted into a two-dimensional time-frequency signal YeRMxN by the short-time Fourier transform (STFT), which is a frame-adding window in accordance with the method of overlapping segmentation, and the window function adopts the Hamming window, with the length of the frame being set to 23ms, and the frame shift being set to 6ms, i.e., each frame contains N=1024 sample points, and there is an overlap of 256 sample points between neighboring time frames. After conversion, the time-frequency signal Y is partitioned into sub-band clusters B with batch data (Batch size) = M/T in a time period T. The remaining frames are padded with values of 0 so that the time dimension expands to T. In order to maintain correlation at the articulation of speech segments, the sub-bands of the latter frame overlap with the former by a time period .L x 2 The amplitude spectrum $\left|Y_{\text{in}}\right| \in^{\sim T\times N}$ of each subband b in Y is used as input to the generator, but considering that the high-frequency portion of the sound is small in energy and relatively insensitive to human hearing, the high-frequency portion of the sound larger than the frequency F is ignored during the training phase, and $\left|Y_{\text{filter}}\right| \in \sim T' \times F$ is used as the input to minimize the number of training parameters and to preserve the most important information of the speech.

3.2.3. GENERATOR

After the input speech is processed to $||Y_{\text{filter}}|$ as the input to the encoder *RNNenc*, *RNNenc* using a bi-directional *RNN (Bi-GRU)*, the output of each time frame h_t updated with the iteration of time frames *t* and a residual network is superimposed as:

$$\begin{cases} llh_{enc_t} = h_t + \left|y_{filter_t}\right| \\ \left|Y_{filter}\right| = \left[\left|y_{filter_T}\right|, \ldots, \left|y_{filter_t}\right|, \ldots, \left|y_{filter_1}\right|\right], \left|y_{filter_t}\right| \in^{\sim F} \end{cases} \quad (8)$$

where h_{enc_t} is denoted as the amplitude spectral vector of the output h_t superimposed on $\left|y_{\text{filter }t}\right|$ at each time t. The residual network facilitates faster training.

The h_{enc_1} of each time frame $\text{t} \in T$ in the merged time period T is denoted as $H_{\text{enc}} \in^{\sim T\times(2\times F)}$, and the overlapping time period *L x 2* is subtracted to obtain the loss $H_{\text{enc}} \in^{\sim T\times(2\times F)}$, where $T' = T-(L\times 2)$, specifically:

$$\widetilde{H}_{enc} = \left[h_{enc_{1+L}}, h_{enc_{2+L}}, , , h_{enc_{T-L}}\right] \quad (9)$$

where L is denoted as the time period in which the sub-bands overlap and is merged according to the above equation to obtain $\widetilde{H}_{enc}$.

https://doi.org/10.17993/3ctecno.2024.v13n1e45.57-76

The introduced recursive derivation algorithm generates temporary variables H^{j}_{dec} continuously and recursively through the encoder RNN_{dec} until the convergence criterion is satisfied, which is the mean-square error L_{MSE} between neighboring valuations of the temporary variables H^{j}_{dec} and the threshold is τ_{term} Let the maximum number of iterations be iter, and $\text{func}^{j}_{\text{dec}}$ denotes the training function of the decoder RNN.

H^{j}_{dec} After the decoding converges, it is passed to the sparse coding layer that generates the time-frequency mask M^j and shares the sparse coding layer weight parameter for each time period T:

$$\tilde{M}^{j} = \text{Re}LU\left(H^{j}{}_{dec}W_{\text{mask}} + b_{\text{mask}}\right) \tag{10}$$

The modified linear unitary function is defined as follows:

$$\text{Re}LU(x) = \begin{cases} x & \text{if} \quad x > 0 \\ 0 & \text{if} \quad x < 0 \end{cases} \tag{11}$$

where ReLU is a segmented linear function that sets all negative values to 0 while positive values remain constant, a setting known as unilateral inhibition, which gives the neurons sparse activation, and the sparsification process is done to improve interference suppression while restoring the frequency dimensions to the target speech signal frequency dimension N. $W_{\text{mask}}\,\epsilon^{\sim(2\times F)\times N}$ is the weight coefficients matrix for the feed-forward neural network, and $b_{\text{mask}} \in \sim N$ is the corresponding deviation.

The amplitude spectrum $\left|\hat{Y}^{j}_{\text{filter}}\right| \in \sim T \times N$ of the target speech signal is obtained by the encoder and decoder defined earlier with the following equation:

$$\begin{cases} \left|\hat{Y}^{j}_{\text{filter}}\right| = \left|Y_{\text{filter}}\right| \otimes \tilde{M}^{j} \\ \left|Y_{\text{filter}}\right| = \left[\left|y_{in_L}\right|, \ldots, \left|y_{in_{T-L}}\right|\right] \end{cases} \tag{12}$$

where $\left|Y_{\text{filter}}\right|$ is the real input to the generator.

3.2.4. DISCRIMINATORS

The time-frequency mask generated by the generator contains perturbations from the noise signal, and the discriminator plays a role in noise reduction by determining the true and false speech signals, so that the generated signal Й constantly approximates the target speech signal [19-20]. The discriminator consists of the codecs of feedforward neural networks FFN_{enc} and FFN_{dec}. The inputs are divided into

https://doi.org/10.17993/3ctecno.2024.v13n1e45.57-76

two types, one is the speech signal $\hat{Y}^{j}_{\text{filter}}$ and the mixed signal Y_{in} generated by the generator, and the other is the real speech signal Y^j and the mixed signal Y_{in}, and the inputs are merged into Y^{j}_{concat}. FFN_{enc} and FNN_{dec} share the weight parameters through the time period T. The output of the discriminator is:

$$\text{Re al / Fake} = \text{Re}LU\left(\text{Re}LU\left(Y^{j}_{\text{conat}} W_{\text{enc}} + b_{\text{enc}}\right) W_{\text{dec}} + b_{\text{dec}}\right) \tag{13}$$

where $W_{\text{enc}} \in \sim 2N \times (N/2)$ and $W_{\text{dec}} \epsilon^{\sim (N/2)\times 1}$ denote the weight coefficient matrices of feedforward neural networks FFN_{enc} and FFN_{dec}, respectively, with corresponding deviations of $b_{enc} \in^{\sim N/2}$ and $b_{dec} \in^{\sim 1}$.

3.2.5. TRAINING OBJECTIVES

Based on the input of the generator as well as the input of the discriminator, the objective function is adjusted to:

$$\min_{G} \max_{D} V_{CGAN}(G, D) = E\left[\log D\left(Y^{j}, Y_{\text{in}}\right)\right] + E\left[\log\left(1 - D\left(G\left(Y_{\text{in}}\right), Y_{\text{in}}\right)\right)\right] \tag{14}$$

Where Y^j is the real speech signal, Y_{in} is the input mixed signal, and $G(z)$ is the generated speech signal. The input to the discriminator is not only the original speech signal r_i and the corresponding signal generated by the generator Y^j, but also an additional mixed signal Z_t obtained by short-time Fourier transformation of the time-frequency signal Y_{in}, Y_{in} which constrains the generation direction of the generator. The GAN network enables the generated speech signal not only to approximate the probability distribution of the target speech signal, but also learns the spectral structure of the audio signals in this environment.

3.3. CREATING TEXT GENERATION MODELS USING LSTMS

The traditional machine translation model only associates the learned expression of the last word with the current word to be predicted for translation, whereas the addition of the attention mechanism associates the learned expression of each word at the source language end with the current word to be predicted for translation. Compared with the traditional machine translation, the effect of the model after adding the attention mechanism is significantly improved, two LSTM classification models, one is to use the output of the last moment of the LSTM as a higher level of representation, and the other is to average all the moments of the LSTM output as a higher level of representation. Both of these representations have certain defects, the first one is missing the previous output information, and the other averaging does not reflect the different importance of the output information at each moment. In order to solve this problem, the Attention mechanism is introduced, and the LSTM model is improved in this paper, and the LSTM-Attention model is shown in Figure 5.

https://doi.org/10.17993/3ctecno.2024.v13n1e45.57-76

The input sequence in the figure is the vector representation of each word of a text segmentation $x_0, x_1, x_2, \cdots, x_t$, and each input is passed into the LSTM unit to get the output of the corresponding hidden layer $h_0, h_1, h_2, \cdots, h_t$. Here, Attention is introduced in the hidden layer, and the probability distribution value of the attention assigned to each input is calculated $\alpha_0, \alpha_1, \alpha_2, \cdots, \alpha_t$, and the idea is to compute the proportion of the matching score of the output of the hidden layer and the whole text representation vector to the overall score at that moment, $\alpha_i, j \in [0, t]$ the formula is as follows:

$$\alpha_i = \frac{\exp\left(\mathrm{score}\left(\bar{h}, h_i\right)\right)}{\sum_j \exp\left(\mathrm{score}\left(\bar{h}, h_j\right)\right)} \tag{15}$$

where h_t is the output state of the hidden layer at the i nd moment, and $\bar{h}$ can be regarded as a text representation vector one level higher than the word. As mentioned above, both text representation methods have defects, so here $\bar{h}$ is randomly initialized as a parameter to be gradually updated during the training process. $\mathrm{score}\left(\bar{h}, h_i\right)$represents the score of the i th hidden layer output h_i in the text representation vector $\bar{h}$, the larger the score, the greater the attention of the input word in the text at this moment, the formula is as follows:

$$\text{score } \left(\bar{h}, h_i\right) = w^{\mathrm{T}} \tanh\left(W\bar{h} + Uh_i + b\right) \tag{16}$$

Where w, W, U is the weight matrix, b is the bias, and $tahn$ is the nonlinear activation function. After obtaining the value of the probability distribution of attention at each moment, the feature vector v containing the text information is calculated as follows:

$$v = \sum_{i=0}^{t} \alpha_i h_i \tag{17}$$

Finally, the softmax function is utilized to obtain the prediction category as, which is calculated as follows:

$$y = \mathrm{softmax}\left(W_v v + b_v\right) \tag{18}$$

In this paper, we use the gradient descent method to train the model, and gradually update the parameters of the model by calculating the gradient of the loss function, and finally reach the convergence. In order to make the objective function converge more smoothly, and also to improve the efficiency of the algorithm, only a small number of samples are taken for training each time. The model uses the cross-entropy loss function, and the calculation formula is as follows:

$$H_{y'}(y) = -\sum_i y_i' \log y_i \tag{19}$$

where y_i' is the actual category label value and y_i is the predicted category label value calculated using the softmax function

https://doi.org/10.17993/3ctecno.2024.v13n1e45.57-76

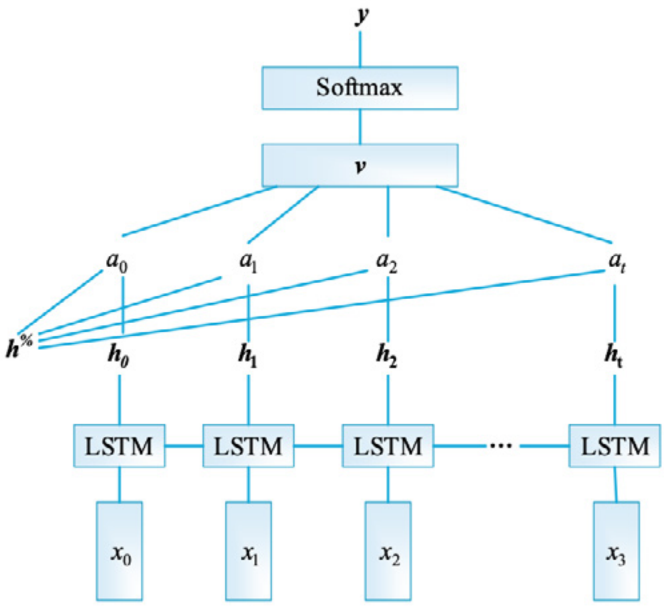

Figure 5 LSTM-Attention model

4. PROSPECTIVE ANALYSIS OF ARTIFICIAL INTELLIGENCE IN LINGUISTIC RESEARCH

4.1. QUALITY OF TRANSLATION IN DIFFERENT LANGUAGES

In order to verify the diagnostic ability of LSTM for language translation system, in the experiment, the LSTM-based artificial intelligence system is applied to different language translations to examine the ability of the diagnostic system in revealing the translation quality, strengths, weaknesses and characteristics of the translation system. The language translation systems that participated in the experiment included three statistical language translation systems, a rule-based language translation system that included diagnostic scores for each linguistic category at the lexical and phrase levels, a lexical category group that included all lexical categories, and a phrase category group that included all phrase-level category scores, system-level scores, and system-level scores computed using BLEU. Here, the small size of the test corpus resulted in a small number of sentence-level detection points with low reliability, so they were not considered for the time being. The first column in the table is the name of the diagnostic category or group of categories. The second and third columns are the diagnostic scores from System A and System B, respectively. The fourth column is the Paired t-statistic significance test score from the scores of the two systems. This score was obtained by repeating the experiment on a random subset of the test set 134). In this experiment, a Paired t-statistic value greater than 2.17 would indicate that the difference between the two scores is significant (>95%). The fifth column is the standard deviation of the diagnostic scores for Systems A and B. The sixth column is the 95% confidence interval for the diagnostic scores of System A and

https://doi.org/10.17993/3ctecno.2024.v13n1e45.57-76

B, accurate only to 0.01 due to space space. As can be seen by the BLEU scores. System B is 0.005 points higher than System A.

The translation system diagnostic results are shown in Table 1, where the difference in diagnostic scores between the two systems on the lexical category groups is not significant. On the diagnostic scores for each linguistic category at the lexical level, the two also have their own distinctions, and there is no obvious advantage for either one. However, on the phrase category group, the score advantage of System B, or LSTM, was more pronounced, and on the diagnostic scores for each linguistic category at the phrase level, System LSTM was higher than System A across the board, especially on the discontinuous distant phrase category. This result shows the advantage of System B in dealing with complex phrases and distant relations, an advantage that comes from recurrent neural network-based processing. Paired t-statistic statistics also show that the differences between the two systems are significant for all diagnostic scores. This comparison shows that the diagnostic system accurately captures the microscopic differences and commonalities between two systems with very similar macroscopic performance.

Table 1 Diagnostic results of translation system

	System A	System B	T Score	Score variance (A/B)	95% confidence interval (A/B)
	Lexical level				
Ambiguous word	0.59	0.59	2.88	0.00/0.00	0.58-0.61/0.58-0.61
Neologism	0.18	0.19	5.56	0.03/0.03	
Idiom	0.19	0.23	13.38	0.04/0.04	
Noun	0.59	0.59	2.68	0.00/0.00	
Verb	0.51	0.51	9.41	0.00/0.00	
Adjective	0.58	0.55	17.43	0.01/0.02	
Pronoun	0.75	0.73	13.49	0.02/0.02	
Adverb	0.53	0.54	7.11	0.01/0.01	
Preposition	0.65	0.64	6.21	0.01/0.01	
Quantifier	0.58	0.57	4.68	0.02/0.02	
Reduplicated word	0.33	0.39	9.86	0.10/0.08	
Match	0.66	0.65	8.07	0.01/0.01	
	Phrase level				
Subject-predicate collocation	0.51	0.51	7.36	0.01/0.01	
Predicate-object collocation	0.41	0.41	15.52	0.01/0.01	
Interobject collocation	0.44	0.51	9.51	0.01/0.01	
Quantifier collocation	0.51	0.51	3.56	0.01/0.01	
Azimuth collocation	0.52	0.53	2.83	0.03/0.04	
	Category group				

 https://doi.org/10.17993/3ctecno.2024.v13n1e45.57-76

Vocabulary	0.48	0.48	8.03	0.01/0.01	
Phrase	0.47	0.49	13.97	0.01/0.01	
	System level				
Department of linguistics Class score	0.42	0.43	16.51	0.00/0.00	
BLEU series Class score	0.35	0.36	7.91	0.00/0.00	

4.2. IDENTIFICATION ACCURACY

In this paper, we use the BIO annotation specification, and the named entity category includes three categories, person name, organization name and place name. In order to determine whether this linguistic feature is useful for Uyghur named entity recognition, the four features Pos1-Pos4 are added to the LSTM intelligent model at the same time, which is used to compare whether the addition of the Pos4 feature, is helpful for the overall named entity recognition task. The affixed lexical features are shown in Table 2.It can be seen that, in terms of the F1 value, the addition of all of them improves the lexical features to some extent. There is an improvement of 0.5.

Table 2 affix characteristics /%

Trait	P	R	F1
Just_token	75.8	74.7	75.3
Pos1	76.7	74.7	75.6
Pos2	76.4	75.0	75.7
Pos3	74.4	75.8	75.4
Pos4	75.6	73.0	74.3
Pos1-Pos4	76.2	75.5	75.9(↑0.5)

After Table 2, it is found that linguistic features can improve the language named entity recognition accuracy, therefore, all the linguistic features will be added, and the comparison experiments with Pos1-Pos4 features and Suffix1-Suffix4 features will be conducted, and the comparison of linguistic features is shown in Table 3. The final F1 value is improved by 3.9%, which fully indicates that for complex morphological languages, adding linguistic features can improve named entity recognition accuracy.

https://doi.org/10.17993/3ctecno.2024.v13n1e45.57-76

Table 3 affix characteristics /%

Trait	P	R	R
Just_token	75.8	74.7	74.7
Pos1-Pos4	76.2	75.5	75.5
Suffix1-Suffix4	78.6	75.0	75.0
All_feature	77.5	81.1	81.1

4.3. VALIDATION OF CREATIVE WRITING SKILLS

In order to facilitate statistical analysis, before conducting exploratory factor analysis, the suitability of factor analysis of questionnaire data N=727 was tested by KMO and Bartlett's test of sphericity, and the results showed that KMO=0.98 (>0.9), good level. LSTM was used to extract the common factors from the questionnaire data and the final factor loading matrix was obtained by the maximum variance method with orthogonal rotation, Table 4 shows the results of total variance interpretation of writing strategies. Five factors were extracted using the writing strategy, and the eigenvalues of each factor reached an acceptable value greater than 1. The cumulative variance contribution of the five factors was 66.5%, which is a desirable level of more than 60%. The common degree of each item, except R40, is greater than 0.5, and the factor loadings have reached 0.4 or more, indicating that the five factors extracted by the AI are all valid and can explain writing strategy ability better.

Table 4 Interprets the total variance of writing strategies

Inicial eigenvalue				Sum of squares of factor loads		
Divisor	Total	Variance %	Accumulate to %	Total	Variance %	Accumulate to %
1	25.8	52.7	52.7	8.1	16.6	16.6
2	2.1	4.4	57.1	8.1	16.5	33.1
3	1.9	4.1	61.2	6.5	13.4	46.5
4	1.4	2.9	64.1	5.9	12.2	58.7
5	1.1	2.4	66.5	3.7	7.7	66.5

5. CONCLUSION

In this paper, LSTM was used as the main tool to explore several aspects in the field of linguistics, including text analysis, speech signal processing and text generation. The suitability test (KMO=0.98) indicated that the data were at a good

https://doi.org/10.17993/3ctecno.2024.v13n1e45.57-76

level and suitable for factor analysis. Five common factors were extracted from the questionnaire data using the LSTM method, and the results showed that these five factors had high eigenvalues with a cumulative variance contribution rate of more than 60% of the desirable level, which indicated that these factors were able to explain the writing strategy ability better. In addition, the common degree of each item is greater than 0.5, and the factor loadings are all above 0.4, which further verifies the validity of these five factors extracted by AI. In addition, the article uses the BIO annotation specification for named entity recognition, which classifies named entities into three categories: personal names, institutional names, and place names. By adding the affix lexical features to the LSTM intelligent model, the results show some improvement in the F1 value, indicating that these features are helpful for the Uyghur named entity recognition task, which provides a strong support and innovation for the application of artificial intelligence in linguistic research.

ACKNOWLEDGMENTS

1. This research was supported by the funding of the following research project: Exploration on the Reform of College English Grammar Teaching by Educational Informationization (No.JZ180077).

2. This research was supported by the funding of the following research project: Corpus-assisted English Grammar Teaching Innovation (No.2018CG02644).

3. This research was supported by the funding of the following research project: An Innovative Model of Blended English Teaching by SPOC (No. FJJKCGZ18-793).

REFERENCES

(1) Yang, L., Fan, Z., & Zhou, J. (2022). Borderless Fusion Financial Management Innovation Based on Speech Recognition Technology. Scientific Programming.

(2) Dokuz, Y. , & Tufekci, Z. . (2020). Mini-batch sample selection strategies for deep learning based speech recognition. Applied Acoustics,171.

(3) Ho, N. H., Yang, H. J., Kim, S. H., & Lee, G. (2020). Multimodal approach of speech emotion recognition using multi-level multi-head fusion attention-based recurrent neural network. IEEE Access, 8, 61672-61686.

(4) Tsunemoto, A., Trofimovich, P., & Kennedy, S. (2023). Pre-service teachers' beliefs about second language pronunciation teaching, their experience, and speech assessments. Language Teaching Research, 7(1), 115-136.

(5) Hyland Bruno, J., Jarvis, E. D., Liberman, M., & Tchernichovski, O. (2021). Birdsong learning and culture: analogies with human spoken language. Annual review of linguistics, 7, 449-472.

(6) Bernardo, M. L. P. (2022). Localizing theory in a Spanish-language translation program. Teaching Literature in Translation: Pedagogical Contexts and Reading Practices, 262.

https://doi.org/10.17993/3ctecno.2024.v13n1e45.57-76

(7) Rasulova, Z. (2022). TRANSLATION CONCEPTS IN THE CONTEXT OF MODERN LINGUISTIC RESEARCH. International Bulletin of Applied Science and Technology, 2(11), 161-165.

(8) Braithwaite, B. (2020). Ideologies of linguistic research on small sign languages in the global South: A Caribbean perspective. Language & Communication, 74, 182-194.

(9) Bafoevna, N. D., & Ikromdjonovna, K. N. (2023). The Main Directions of Theo linguistic Research In Modern Linguistics. Journal of Survey in Fisheries Sciences, 10(2S), 2127-2136.

(10) Mizumoto, A., Plonsky, L., & Egbert, J. (2021). Meta-analyzing corpus linguistic research. In A practical handbook of corpus linguistics (pp. 663-688). Cham: Springer International Publishing.

(11) Su, H., Zhang, Y., & Lu, X. (2021). Applying local grammars to the diachronic investigation of discourse acts in academic writing: The case of exemplification in Linguistics research articles. English for Specific Purposes, 63, 120-133.

(12) Awad Al-Dawoody Abdulaal, M. (2020). A cross-linguistic analysis of formulaic language and meta-discourse in linguistics research articles by natives and Arabs: Modeling Saudis and Egyptians. Arab World English Journal (AWEJ) Volume, 11.

(13) Chen, L., & Hu, G. (2020). Surprise markers in applied linguistics research articles: A diachronic perspective. Lingua, 248, 102992.

(14) Umarova, N. R. (2021). A linguistic approach to conceptual research. ASIAN JOURNAL OF MULTIDIMENSIONAL RESEARCH, 10(4), 62-66.

(15) Hamzah, M. H., Halim, H. A., Bakri, M. H. U. A. B., & Pillai, S. (2022). Linguistic Research on the Orang Asli Languages in Peninsular Malaysia. Journal of Language and Linguistic Studies, 18, 1270-1288.

(16) Oh, Y. R., Park, K., Jeon, H. B., & Park, J. G. (2020). Automatic proficiency assessment of Korean speech read aloud by non-natives using bidirectional LSTM-based speech recognition. Etri Journal, 42(5), 761-772.

(17) Hou, W., Wang, J., Tan, X., Qin, T., & Shinozaki, T. (2021). Cross-domain speech recognition with unsupervised character-level distribution matching. arXiv preprint arXiv:2104.07491.

(18) Santoso, J., Setiawan, E. I., Purwanto, C. N., Yuniarno, E. M., Hariadi, M., & Purnomo, M. H. (2021). Named entity recognition for extracting concept in ontology building on Indonesian language using end-to-end bidirectional long short term memory. Expert Systems with Applications, 176, 114856.

(19) Peng, L., Fang, S., Fan, Y., Wang, M., & Ma, Z. (2023). A Method of Noise Reduction for Radio Communication Signal Based on RaGAN. Sensors, 23(1), 475.

(20) Budinsky, R. , Ozmeral, E. J. , & Eddins, D. . (2023). The impact of hearing aid user's own voice on device signal processing. The Journal of the Acoustical Society of America.

ABOUT THE AUTHOR

Shaohua Jiang is working as a lecturer of School of Humanities, Fujian University of Technology. His research is focused within the fields of English Language

 https://doi.org/10.17993/3ctecno.2024.v13n1e45.57-76

Education, Translation Education under the heading Smart Education and Artificial Intelligence.

Zheng Chen is an Associate Professor at the Department of Foreign Languages, Concord University College, Fujian Normal University. Her research is focused within the fields of English Language Education and American Literature Studies under the heading Artificial Intelligence.