EAI Endorsed Transactions
on Scalable Information Systems

Research Article EAI.EU

An automatic scoring method for Chinese-English spoken translation based on attention LSTM

Xiaobin Guo[1,*]

[1]Zhengzhou Railway Vocational and Technical College, Zhengzhou 450000, China

Abstract

In this paper, we propose an automatic scoring method for Chinese-English spoken translation based on attention LSTM. We select semantic keywords, sentence drift and spoken fluency as the main parameters of scoring. In order to improve the accuracy of keyword scoring, this paper uses synonym discrimination method to identify the synonyms in the examinees' answer keywords. At the sentence level, attention LSTM model is used to analyze examinees' translation of sentence general idea. Finally, spoken fluency is scored based on tempo/rate and speech distribution. The final translation quality score is obtained by combining the weighted scores of the three parameters. The experimental results show that the proposed method is in good agreement with the result of manual grading, and achieves the expected design goal compared with other methods.

Keywords: Chinese-English spoken translation, attention LSTM, sentence level.

Received on 06 January 2022, accepted on 12 January 2022, published on 13 January 2022

doi: 10.4108/eai.13-1-2022.172818

*Corresponding author. Email: 9106750[illegible]@qq.com

1. Introduction

Chinese-English spoken translation quality assessment is one of the hot topics in the field of automatic assessment of Chinese-English translation quality in recent years [1]. In the automatic scoring of oral English, some studies are mostly aimed at the oral English evaluation of pronunciation quality [2], such as reading questions and repeat questions. Zheng et al. [3] scored English reading questions through maximum likelihood linear regression and maximum posterior probability algorithm, and achieved certain results. However, there is still a lack of effective evaluation strategies for question types related to text content (keywords, sentence drift, etc.), such as interpretation questions and retelling questions. Although some scholars have carried out the corresponding research, but really applied to large-scale speaking test scoring results are very limited. Zhang et al. [4], for example, used speed, text coverage, keyword coverage and other indicators to score oral retelling questions, but this method lacked the overall analysis of sentence general idea. Suyoun Yoon et al. [5] used Siamese convolutional neural network to extract key information features of examinees' spoken sentences for scoring, but also did not conduct further research on the general idea of sentences. Therefore, there are still many challenges in establishing an effective automatic scoring model for Chinese-English spoken translation questions.

In recent years, many methods based on deep neural network (DNN) have been widely used in natural language processing [6-10]. Automatic feature extraction using DNN greatly alleviates the feature dependence problem of traditional methods. At the same time, word-distributed embedding is used as the input of DNN, which makes the extracted features contain rich semantic

information. These DNN methods combine with distributed word vector have achieved success in many tasks, with better accuracy and efficiency than traditional methods. For part-of-speech tagging tasks, as much as possible in order to avoid according to the characteristics of the specific tasks, reference [11] used many hidden layers to automatically extract features. For the part of speech tagging, entity recognition and chunking analysis tasks, it designs a unified architecture, which greatly eased the feature dependent problem in traditional method, significantly improved the labeling results on each task. For Chinese word segmentation and part-of-speech tagging, a more concise and efficient deep neural network model was designed in reference [12], which achieved a better effect with less use of artificially designed features. However, compared with traditional tagging models, although the neural network model mentioned above reduces the workload of artificial design features, the actual effect is limited by the size of word window, and the context information referenced by pos-tagging is very limited. However, the present research shows that the word categories are closely related to the contextual information around them. Reference [13] proposed that hierarchical long short-term memory (LSTM) was used to obtain a wider range of context information, and part-of-speech tagging and word segmentation tasks were combined to provide supplementary information to each other, thus improving the accuracy of part-of-speech tagging. Reference [14] proposed to add CRF layer in the output layer of LSTM network, and use CRF layer to realize tag inference at sentence level. The result was better than the traditional CRF model and the model using LSTM network alone. However, this reference only focused on pos-tagging, chunking analysis and entity recognition in English, and did not discuss the experiments on pos tagging and relevant corpus in Chinese.

Recently, attention mechanism has been introduced into the field of natural language processing, and has achieved good application effects in machine translation [15], syntactic analysis and automatic summarization [16]. The attention mechanism is used to assign different probability weights to the hidden layer units of neural network, so that the hidden layer can pay attention to the feature information which is more favorable to the classification task, and reduce the attention to some redundant information. Thus, in the same context sequence, adding the hidden layer of attention mechanism can further optimize the quality of extracted features. This is well proved by the application of attention mechanism in syntactic analysis [17], which enables the syntactic analysis model to learn long-distance syntactic dependency information. For part-of-speech tagging, as a syntactic functional category of words, the accuracy of part-of-speech tagging is obviously affected by the contextual information in the sentence. Especially for some long distance and specific syntactic structure information, it can solve the problem of tagging concurrent words well [18]. Adding attention mechanism to neural network annotation model can obtain these specific context information well and improve the accuracy of annotation model.

This paper introduces an automatic grading method for the quality of Chinese-English spoken translation. We select semantic keywords, sentence semantic similarity and oral fluency as evaluation indexes to evaluate translation quality. In sentence-level Chinese-English translation, the translation of key words must convey the meaning, and the general meaning of Chinese-English sentences should also be accurate. As for oral translation, fluency parameters are also very important, and fluency also reflects the overall level of the translator's oral English [19]. In the scoring of sentence-oriented Chinese-English oral translation questions, researchers generally pay attention to the evaluation of the accuracy of Chinese-English oral translation and the general idea of the whole sentence. This is the main reason why we choose the above three evaluation parameters. In many Spoken English proficiency tests in China, Chinese-English spoken translation is the main question type. Therefore, automatic scoring of Chinese-English spoken translation questions has practical significance.

In the oral test, the reference answer standard of the sentence-level Chinese-English oral translation question manual revision clearly states that 60% of the key information should be translated, and 40% of the overall comprehension and expression of the sentence should be accounted for. Therefore, this scoring criterion should also be considered when constructing automatic scoring model [20].

In terms of key word scoring, the scoring model should not only consider the score of the key words given in the answer, but also the score of the synonyms related to keywords [21]. Therefore, we should establish thesaurus related to answer keywords, and then give the score of keywords by the degree of thesaurus related to keywords.

In terms of sentence comprehension, scholars usually score by calculating the similarity between the standard answer sheet and the answers. With the development of deep learning technology [22-24], some specific neural network models based on deep learning can mine deeper semantic information in sentences. Because of this discovery, some researchers have applied neural network models to automatic scoring tasks. For example, Qian Hussein et al. [25] built an adaptive deep learning speaking scoring system. Bshary et al. [26] used the LSTM to evaluate oral pronunciation. This paper attempts to apply deep learning to sentence general idea scoring in automatic scoring of Chinese-English spoken translation.

2. A quality scoring model for Chinese-English spoken translation

2.1. Keyword translation calculation method based on synonym discrimination

As can be seen from the scoring criteria of Chinese-English spoken translation (Table 2), the scoring of key words is very important. In order to accurately evaluate the information of key word translation in the answer sheet, we need to consider the following two situations: first, the number of key words in the reference answer; Second, refer to the use of synonyms for keywords in the answer. Therefore, it is necessary to judge whether the answer of the examinee contains the key information required by the question, that is, to inspect the examinee's grasp of keywords and their synonyms. In order to judge candidates' mastery of keywords and their synonyms, this paper adopts the synonym discrimination method combined with Word2Vec and semantic tree [27] to carry out semantic analysis and grading of candidates' oral keyword information at the lexical level.

From the semantic level, a sentence is usually composed of "key words" and "general words". "Key words" can affect the meaning of the sentence, which is the key information required by the oral answer. However, "general words" do not have a decisive influence on understanding the meaning of the whole sentence [28]. Therefore, according to the requirements of manual scoring, this paper mainly scores the key words in the sentence, and the influence of "general words" on the sentence is also considered when scoring the sentence drift. Because Word2Vec can define the semantics of keywords and their synonyms and represent them as vectors, it can be used to calculate semantic similarity. In order to score the key information of the answer, we need to build a corpus for the key information of the answer, which should contain key information words and synonyms with high frequency of use. In order to avoid too many synonyms, candidates' knowledge should be considered at the same time [29,30]. Before the experiment, we recorded all the keywords and synonyms with high usage to build a language library. At the same time, manual annotation was carried out to form a text corpus for the use of the two experimental schemes.

The following is a Chinese-English spoken translation question and standard answer. Titled is “学习决定成绩,成绩又促进学习进步”. The standard answer is “Learning determines grade, and achievement promotes progress in learning.” Among them, "Learning", “grade”, "achievement", and "progress" are the key words of the standard answer. Keywords and synonym structures are shown in Figure 1.

Figure 1. Semantic tree

In Figure 1, the node words in each right single-branched tree are synonyms of each other, and the similarity decreases gradually to the right.

At the lexical level, this paper uses the method of synonym discrimination to score key information, and the specific steps are shown in Figure 2. First, it identifies the position of the candidate's key word in the semantic tree. If the candidate's key word is not found in the semantic tree, the word is not included in the thesaurus representing the standard answer, and the key information point is lost. If there are nodes in the semantic tree, the keywords that examinee answers are converted into semantic feature vectors through the previously trained Word2Vec model, and then the semantic similarity between two words is calculated by cosine similarity. " ⊕ " in Figure 2 represents the calculation of cosine similarity. The last key information score is the mapping score of semantic similarity corresponding to the word.

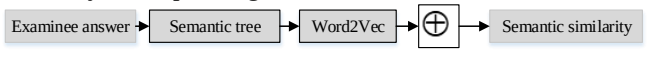

Figure 2. Flow chart of lexical semantic similarity calculation

2.2. Attention LSTM

LSTM model

LSTM is a Recurrent Neural network (RNN) [31]. It solves the problem of gradient disappearance existing in traditional RNN by introducing memory cell and gated mechanism, and performs better in representing the context information of elements in sequence data and extracting long-distance dependencies. Figure 3(a) shows a single LSTM unit, and Figure 3(b) shows its internal structure.

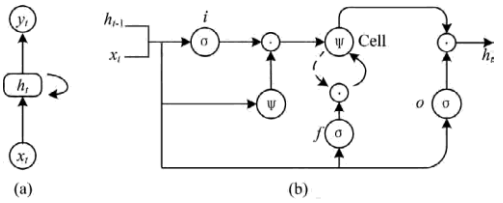

Figure 3. Internal structure of LSTM unit

There are three kinds of gates in LSTM unit: input gate i, forgetting gate f and output gate o. The input gate is used to control the update information of memory unit. The forgetting gate is used to control the amount of memory unit information used at the previous moment. Output gate is used to control the amount of information output to the next hidden state. At time t, given input vector x_t and the hidden state h_{t-1} at the previous moment, LSTM unit calculates the hidden state h_t at the current moment through internal loop and update:

$$i_t = \sigma(U^i x_t + W^i h_{t-1} + b^i) \quad (1)$$

$$f_t = \sigma(U^f x_t + W^f h_{t-1} + b^f) \quad (2)$$

$$i_t = \sigma(U^i x_t + W^i h_{t-1} + b^i) \quad (3)$$

$$c_t = f_t * c_{t-1} + i_t * \varphi(U^c x_t + W^c h_{t-1} + b^c) \quad (4)$$

$$h_t = o_t \varphi(c_t) \quad (5)$$

Where, c_t represents the state information of the memory unit. The parameter set $\{U^i, W^i, U^f, W^f, U^o, W^o, U^c, W^c\}$ corresponds to the weight matrix of different gates. $\{b^i, b^f, b^o, b^c\}$ represents the corresponding offset term, σ and φ are sigmoid and tanh activation functions respectively. $*$ sign means vectors wisepoint multiplication.

Generally, the information in LSTM network is one-way, and LSTM can only use the information of the past moment, not the information of the future moment. Obviously, for some tasks such as word segmentation and part-of-speech tagging, both forward and backward information of sequences are very important. Therefore, a reverse layer can be added to the LSTM network to constitute a BLSTM. The BLSTM consists of two LSTM layers with opposite directions, and its structure is shown in Figure 4.

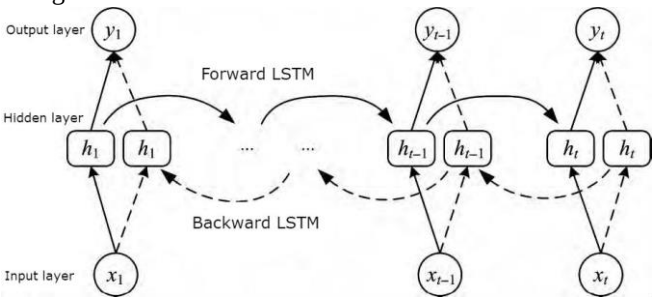

Figure 4. Unfolded BLSTM network

In Figure 4, the expanded BLSTM network structure is divided into three layers: input layer, hidden layer and output layer. x_t, h_t and y_t represent the input vector, hidden state vector and output vector at time t respectively. The hidden layer consists of forward LSTM and reverse LSTM, which are used to calculate forward and reverse hidden states respectively, and then projected to the common output layer. Compared with unidirectional LSTM, bidirectional LSTM has better performance in sequence feature acquisition and representation because the hidden layer information flows along two opposite directions and can obtain both forward and backward historical information. Therefore, it has been applied in many natural language sequence annotation tasks.

Attention mechanism

As a syntactic function category of words, the tagging process of part of speech is influenced by the information of sentence syntactic structure, and is more closely related to the words that have important syntactic dependence. However, other words have no obvious marking effect on the current words. Attention mechanism is a good probability weight allocation mechanism. By calculating the probability weight of attention at different moments, some nodes which are very related to the annotation of the target word get more attention and are assigned to larger probability weight. In this way, the quality of feature vector of hidden layer can be improved. The structure of the basic attention model is shown in Figure 5.

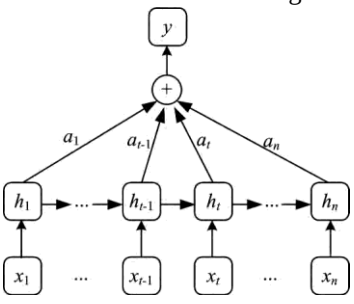

Figure 5. Structure of attention model

In the neural network model with added attention mechanism, the new hidden state vector s is jointly determined by the initial hidden state vector h_i at each moment, and the calculation formula is as follows:

$$s = \sum_{i=1}^{t} \alpha_i h_i \tag{6}$$

Where α_i represents the weight of the initial hidden state h_i relative to the new hidden layer, and its calculation formula is as follows:

$$\alpha_i = \frac{\exp(e_i)}{\sum_{j=1}^{n} \exp(e_i)} \tag{7}$$

$$e_i = v \tanh(wh_i + b) \tag{8}$$

Where e_i represents the energy value of the hidden state at moment i, which is mainly determined by the hidden state vector h_i at that moment. w and v are the weight matrices, and b is the corresponding offset value. The corresponding process of formulas (6)~(8) realizes the transformation from the initial hidden layer to the new attention layer, and the weight coefficient α_i corresponding to the hidden layer at each moment reflects its influence on the current output.

Chinese-English Spoken Translation based on attention LSTM

The Chinese-English spoken translation model proposed in this paper adds attention mechanism on the basis of BLSTM, and the specific model structure is shown in Figure 6. It mainly includes three parts: input layer, hidden layer and output layer. Among them, the hidden layer is composed of one-way LSTM layer, two-way LSTM layer and attention layer, which are respectively introduced below.

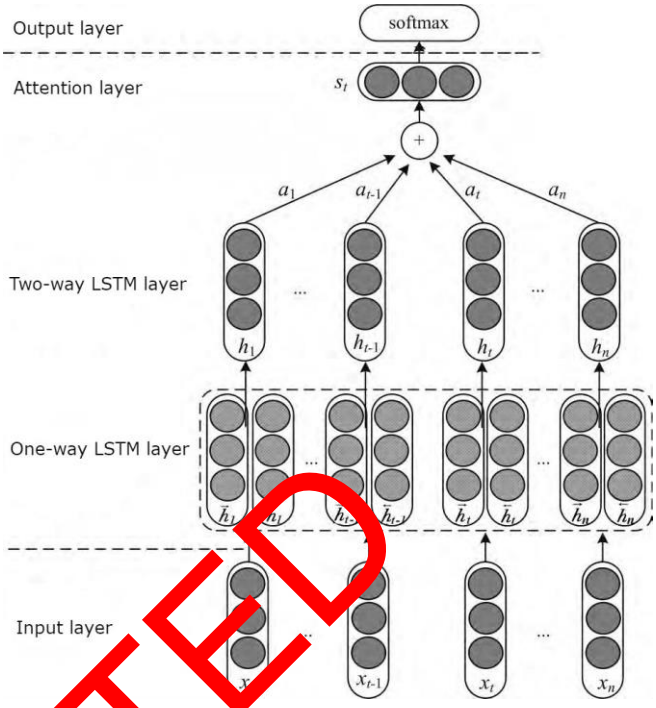

Figure 6. Structure of attention-based LSTM model for Chinese-English spoken translation

1) Input layer. Onehot representation is traditionally used to represent word vectors. This form of expression only symbolizes the words and does not contain any semantic information. And because of the large dimension, the element value is mostly 0. Therefore, there is a serious data sparsity problem. Distributed word vector uses low-dimensional and dense real vector to vectorize words, which contains rich semantic information and is widely used in many natural language processing tasks. This paper adopts the distributed word vector representation method, using Google word2vec tool. The word vector matrix M is formed through pre-training and indexed in the word vector matrix [32,33], and each word is transformed into its corresponding word vector form x_t, which is used as the input of BLSTM network.

2) Hidden layer. The calculation is mainly divided into three steps.

Step 1. The forward LSTM and reverse LSTM hidden states are calculated according to the LSTM model. One-way LSTM contains only one hidden layer in one direction. According to the input at the current moment, vector x_t, and the hidden state vector h_{t-1} at the previous moment, the hidden state h_t at the current moment is calculated. However, the bidirectional LSTM contains the forward layer and the reverse layer, and the hidden state

vector $\vec{h}_t \in R^{m\times 1}$ and the reverse hidden state vector $\overleftarrow{h}_t \in R^{m\times 1}$ of the forward layer at the current moment need to be calculated respectively.

$$\vec{h}_t = LSTM(x_t, \vec{h}_{t-1}) \quad (9)$$

$$\overleftarrow{h}_t = LSTM(x_t, \overleftarrow{h}_{t-1}) \quad (10)$$

Where m is the implicit element dimension. LSTM() function represents nonlinear transformation of LSTM network, and its main function is to encode input word vector into corresponding implicit state vector.

Step 2. According to LSTM forward hidden state and reverse hidden state, the BLSTM hidden layer is calculated.

The forward hidden state vector $\vec{h}_t$ and reverse hidden state vector $\overleftarrow{h}_t$ are linearly combined by weighted summation method to obtain the hidden layer vector h_t of BLSTM.

$$h_t = W_1\vec{h}_t + V_1\overleftarrow{h}_t + b_1 \quad (11)$$

Where $W_1 \in R^{m\times m}$ and $V_1 \in R^{m\times m}$ are weight matrices. $b_1 \in R^{m\times 1}$ is the corresponding offset term. The hidden layer aggregates the forward and backward sequence information of the current element in the input sequence, which can provide richer contextual features for oral translation.

Step 3. According to the attention mechanism, probability weight is assigned to the BLSTM hidden layer, and the new attention hidden layer is calculated. Since BLSTM contains both forward and reverse layers, both forward $\vec{h}_t$ and reverse hidden states $\overleftarrow{h}_t$ need to be considered. In this paper, the aggregated hidden state vector h_t is used to calculate the hidden layer energy at this moment e_t:

$$e_t = V_2 \tanh(W_2 h_t + b_2) \quad (12)$$

Where, $W_2 \in R^{l\times m}$ and $V_2 \in R^{1\times l}$ are weight matrices, $b_2 \in R^{l\times 1}$ is the corresponding offset term, and l is the dimension of vector V_2. Then, according to the energy value of the hidden state vector at each moment, the corresponding attention probability weight of the hidden state at this moment is calculated:

$$\alpha_t = \frac{\exp(e_t)}{\sum_{t=1}^{n}\exp(e_t)} \quad (13)$$

Where, α_t is the probability weight of attention corresponding to the hidden state h_t. The hidden state energy value and probability weight obtained by equations (12) and (13) reflect the effect of the hidden state at each moment on classification, and this probability distribution is used to assign different importance to different context features. Finally, multiply and sum the hidden states and corresponding probability weights at each moment to obtain the new attention hidden layer vectors $s_t \in R^{m\times 1}$:

$$s_t = \sum_{t=1}^{n}\alpha_t h_t \quad (14)$$

The dimension of the new attention hidden layer obtained in Equation (14) is the same as that of the original hidden layer. Since the probability distribution of attention is different at each moment, the new hidden layer of attention can pay attention to the part of speech tagging that is different from the initial hidden layer and the input sequence [35, 36]. Therefore, at the beginning of each moment, the hidden layer plays a different role in Chinese-English spoken. Among them, the probability of attention of the hidden state which has a great influence on the current word labeling is correspondingly greater.

3) Output layer. Softmax function is used to calculate the probability distribution of tags on the annotation set at each time.

$$y_t = soft\max(W_3 s_t + b_3) \quad (15)$$

Where $W_3 \in R^{L\times m}$ represents the weight matrix between the hidden layer of attention and the output layer. $b_3 \in R^{L\times 1}$ is the corresponding offset term. $y_t \in R^{L\times 1}$ represents the probability distribution of words at the current moment on the annotation set, such as the k-th of y_t (k=1,2,..,L) dimension $y_k = p(y_t = k)$ represents the probability of the k-th part of speech in the current word allocation annotation set. L is the total number of elements in the tagging set.

2.3. The scoring method of oral fluency

In the aspect of pronunciation, this paper mainly analyzes the examinee's oral fluency. Oral fluency is an important indicator for teachers to directly evaluate candidates' oral pronunciation ability. However, oral fluency is mainly reflected in the speed of speakers, so this paper evaluates oral fluency based on tempo/rate and speech distribution. Among them, the characteristic of speed is the average pronunciation time of each word.

First, the candidate's pronounce time is derived from the number of words n and the length of the i-th word in

oral pronounce by a double threshold cut lexical method based on short-term energy and zero crossing rate. Then, formula (16) is used to calculate the characteristics of speech speed. If the speed of the candidate is greater than the set threshold, it is judged to be fluent. Then, fluency score is given by fractional mapping function. If the speed of the examinee is less than the set threshold, the examinee's speech distribution is judged to be uneven and does not meet the pronunciation requirements of the oral answer.

$$speed = \frac{\sum_{i=1}^{n} pronounce\_Time}{n} \quad (16)$$

2.4. Fractional fusion model

The total score of the automatic scoring model for Chinese-English spoken translation is obtained through the following steps:

(1) Introduce the examinee's oral answering voice, calculate the average length of the speech segment and the average pause time based on zero-energy integral-cut lexical model, and calculate the examinee's speaking speed, and then calculate the fluency score.

(2) Using the trained Word2Vec vector, the degree of fitting between the keywords in the answer audio and the keywords in the model library is determined. The position of the keyword in the semantic tree was matched and the score of the keyword was calculated.

(3) Using Word2Vec and short and long memory network model, the semantic features of all sentences in the corpus are transformed into vectors. Match the similarity between the examinee's pronunciation and the general idea of the sentence in the standard answer, and give the score of the general idea of the sentence.

(4) The final evaluation total score is the weighted result of the three scoring indexes, as shown in equation (17).

$$Y = \alpha_1 * X_{keywords} + \alpha_2 X_{sentence} + \alpha_3 X_{fluency} \quad (17)$$

Where, α_1, α_2 and α_3 are the weights of keyword score, sentence drift score and fluency score respectively. After analysis by linear regression prediction method, the weight values are set as 0.6, 0.3 and 0.1 in sequence. Finally, the total score is mapped to the A, B, C and D grade ranges.

3. Experiments and analysis

In order to verify the effectiveness of the proposed method, relevant experiments are carried out. The questions are selected from the interpreting and listening test of a university in June 2020. The first part is Chinese-English spoken translation. There are 5 questions in this section (Table 1), and each question is worth 2 points. We collected 3100 real audio data with accurate manual marking, and each question had 620 audio data within 20 seconds, which were recorded by 620 students in 6 different examination rooms. In order to reduce the influence of the grading teachers' subjectivity on the grading results, each phonetic answer paper was graded independently by two grading teachers, and the average scores of the two teachers were taken as artificial grading.

Table 1. Translation questions and reference answers

Question	Answer
近年来,**机器翻译**研究取得了长足的进步,**译文质量**不断提高。	In recent years, the research of **machine translation** has made great progress, and the **translation quality** has been constantly improved
在人**机交互**和高级用户接口应用领域中,我们希望未来的**机器**能像人一样与我们更加容易和便捷地交流,如**手势驱动控制、手语翻译**等。	In the field of **human-computer interaction** and advanced user interface applications, we hope that the future **machines** can communicate with us more easily and conveniently like human beings, such as **gesture driven control, sign language translation** and so on.

Note: the bold is keyword.

Before the key word score and sentence general idea score, 3100 samples of phonetic answer paper need to be manually translated into text. Meanwhile, in order to extract effective semantic features of examinees, stop words such as "the", "a", "to", "this" and "can" are all removed. In the experiment process, this paper will conduct modeling and experiment on 5 topics respectively. Finally, the experimental results of the overall scoring model were taken as the average of 5 experiments. For the division of data set of each question, the ratio of "training:test=7:3" was used for the experiment.

This paper establishes a scoring model based on three scoring indexes: keywords, sentence drift and oral fluency. The weights of the three indicators are respectively set as 0.6, 0.3 and 0.1. Referring to the suggestions of teachers, this paper sets four grading grades A, B, C and D for a single translation question with a total score of 2 points, combining the translation principle of "faithfulness, expressiveness and elegance" and teachers' grading standards for translation questions. The scoring standard

and corresponding score range of each grade are shown in Table 2.

Table 2. Grading criteria and ratings

score	level	description
1.5≤score≤2.0	A	The key information is accurate, the language expression is fluent, the vocabulary is used properly, the sentence general idea is correct, the overall grasp is excellent
1.0≤score<1.5	B	The key information is not accurate enough, the language expression is smooth, the vocabulary is used properly, and the general idea of the sentence is not well understood
0.5≤score<1.0	C	The key information is relatively accurate, but the language expression is not smooth, the sentence general deviation, general grasp
0≤score<0.5	D	Key information is inaccurate or irrelevant, the language expression is not smooth, the general meaning of the sentence has a large deviation, and the overall grasp is poor

The recording parameters of voice data are: mono, 22050Hz sampling rate, 16 bit coding. Keywords are converted into word vectors by Word2Vec, and the dimension of word vector adopts 100 dimensions. The learning rate is 0.001. The number of neurons in LSTM layer is set to 100, tanh function is used as the activation function, and Adam is used as the optimization method. The batch of data for each training is 3.

This paper uses consistency (accuracy) and Pearson correlation coefficient to evaluate the scoring ability of the automatic scoring model for Chinese-English spoken translation. The consistency rate and Pearson correlation coefficient (r value) are shown in equation (18) and equation (19).

$$A = \frac{\text{The number of samples with consistent system rating and teacher rating}}{\text{The total number of samples}} \tag{18}$$

$$r = \frac{\sum_i (x_i - \bar{x}) * (y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x}) \sum_i (y_i - \bar{y})}} \tag{19}$$

Where x_i represents the model score, and $\bar{x}$ is the mean value of the model score. y_i is the teacher's rating, and $\bar{y}$ is the mean of the teacher's rating. And, $0 < r < 0.2$ means very weak correlation. $0.2 < r < 0.4$ indicates weak correlation. $0.4 < r < 0.6$ means moderately relevant. $0.6 < r < 0.8$ indicates strong correlation. $0.8 < r < 1$ means very strong correlation.

In order to verify the effectiveness of the proposed method, the following comparative experimental scheme is adopted. 1) Oral fluency scoring method and keyword scoring method remain unchanged. LSTM model is not used in sentence drift scoring, and word vector generated by Word2Vec is directly averaged to obtain sentence semantic representation and scoring. 2) On the basis of experimental scheme 1, Word2Vec in keyword scoring and sentence general idea scoring is replaced by another Bert model with good pre-training effect at present. 3) On the basis of experimental scheme 2, Bert and LSTM models are used for sentence general idea scoring. The experimental results are shown in Table 3~Table 7.

Table 3. Comparison between proposed model scoring and teacher scoring (part)

Examinee number	Teacher scoring	Proposed
1	A	A
2	A	A
3	B	A
4	A	A
5	A	A
6	A	A
7	B	A
8	A	A
9	C	C
10	A	A

Table 4. The consistency between the new model score and the teacher's score

Number	Consistency

1	0.8853
2	0.8327
3	0.8643
4	0.8485
5	0.8537
Average	0.8569

Table 5. The consistency rate between Bert+LSTM model score and teacher score

Number	Consistency
1	0.8169
2	0.8493
3	0.8388
4	0.8702
5	0.8231
Average	0.8397

Table 6. Comparison of the average agreement rates of different experimental methods

Method	Consistency
Word2Vec	0.7422
Bert	0.7642
Bert+LSTM	0.8397
Word2Vec+LSTM	0.8569

Table 7. The correlation between the model score and the teacher score

Number	Consistency
1	0.8122
2	0.7822
3	0.8711
4	0.8702
5	0.9095
Average	0.8490

(1) Table 3 is the comparison table of some experimental results. It can be seen that there is a good similarity between the results of the model scoring in this paper and those of teachers. Table 4 shows that the average consistency rate of the automatic scoring model for Chinese-English spoken translation built in this paper is 0.8569 on five questions. Among them, the highest value can reach 0.8853, indicating that the accuracy of the model scoring is close to the real score of the teacher and has good effectiveness. In the actual grading process, the teacher will score at his discretion, while the model scoring is based on the established evaluation indicators. Therefore, the model scoring with unified scoring rules has higher objectivity and authenticity, and can explain the differences between the model scoring in this paper and the teacher scoring.

(2) According to Table 4, Table 5 and Table 6, the experimental effect of using Bert model alone is 0.022 higher than that of using word2Vec model alone, indicating that Bert does improve the acquisition of effective information to a certain extent through the multi-layer bidirectional decoding process. Although Word2Vec+LSTM has a lower agreement rate on questions 2 and 4 than Bert+LSTM, it achieves the best average agreement rate on the 5 questions. This indicates that the word2Vec and LSTM model proposed in this paper have a better combination effect, which improves the expression ability of keyword semantics and sentence semantics to a certain extent, thus improving the accuracy of automatic oral scoring.

(3) Table 7 shows that the correlation between the scoring model in this paper and the average score of teachers reaches 0.8490, indicating a strong correlation between the predicted score of the model and the real score of teachers. Among them, the highest value can reach 0.9095, indicating that the model score of question 5 has a very high correlation with the teacher's score. It is proved that the introduction of synonym discrimination, LSTM neural network model and the evaluation method of speed feature can enhance the scoring ability of the automatic scoring model of Chinese-English spoken translation.

4. Conclusion

This paper aims to analyze the scoring mechanism of Chinese-English spoken translation questions and establish an objective and effective automatic scoring model for Chinese-English spoken translation by taking Chinese-English spoken sentence translation in higher

education examination as the research object. This study shows that the method of synonym discrimination for key words, LSTM model for sentence translation analysis and fluency rating based on speed and pronunciation distribution have good results. The directions for further work are as follows: (1) The amount of corpus needs to be increased. At present, it is difficult to find public test data for oral translation questions, so the research work mainly builds relevant scoring standards and test databases for specific tests, resulting in a small database size, which will affect the extraction and training of semantic features by neural network model, and then affect the scoring accuracy of relevant semantics. (2) The addition of speech recognition module. In this paper, artificial translation is used to replace speech recognition. Only the combination of speech recognition and automatic scoring can establish a complete oral evaluation system. (3) Increase evaluation indicators. At present, this paper does not involve the study of grammar and intonation. Establishing the evaluation index of the scoring model will make the scoring mechanism more comprehensive and the scoring result more objective.

References

[1] Wang Y. On the Chinese-English Translation of Current Political Culture-loaded Words in News from the Perspective of Intercultural Communication[C]// 2021 5th International Seminar on Education, Management and Social Sciences (ISEMSS 2021). 2021.

[2] Shoulin Yin, Hang Li, Asif Ali Laghari et al. A Bagging Strategy-Based Kernel Extreme Learning Machine for Complex Network Intrusion Detection[J]. EAI Endorsed Transactions on Scalable Information Systems. 7(33), e8, 2021. http://dx.doi.org/10.4108/eai.6-10-2021.171247

[3] Zheng S. A Study of Chinese-English Translation of Culture-specific Items in Publicity Texts of Guangzhou's Intangible Cultural Heritage[J]. Theory and Practice in Language Studies, 2021, 11(6):749-755.

[4] Zhang X. A Study of Cultural Context in Chinese-English Translation[J]. Region - Educational Research and Reviews, 2021, 3(2):11-14.

[5] X Liu, H Lai, Wong D F, et al. Norm-Based Curriculum Learning for Neural Machine Translation[C]// Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020.

[6] Qingwu Shi, Shoulin Yin, Kun Wang, Lin Teng and Hang Li. Multichannel convolutional neural network-based fuzzy active contour model for medical image segmentation. Evolving Systems (2021). https://doi.org/10.1007/s12530-021-09392-3

[7] Liu, J., Zhang, J. & Yin, S. Hybrid chaotic system-oriented artificial fish swarm neural network for image encryption. Evolutionary Intelligence (2021). https://doi.org/10.1007/s12065-021-00643-5

[8] Jisi A and Shoulin Yin. A New Feature Fusion Network for Student Behavior Recognition in Education [J]. Journal of Applied Science and Engineering. vol. 24, no. 2, pp.133-140, 2021.

[9] Shoulin Yin, Hang Li, Desheng Liu and Shahid Karim. Active Contour Modal Based on Density-oriented BIRCH Clustering Method for Medical Image Segmentation [J]. Multimedia Tools and Applications. Vol. 79, pp. 31049-31068, 2020.

[10] S. Yin and H. Li. Hot Region Selection Based on Selective Search and Modified Fuzzy C-Means in Remote Sensing Images[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 13, pp. 5862-5871, 2020, doi: 10.1109/JSTARS.2020.3025582.

[11] Weizong W, Xin Z, Zhenyu, et al. The adaptation of sport assessment Catella Questionnaire into simplified Chinese version: cross-cultural adaptation, reliability and validity[J]. Health and quality of life outcomes, 18(1), 2020.

[12] H. Zhang, "Neural Network-Based Tree Translation for Knowledge Base Construction," in IEEE Access, vol. 9, pp. 38706-38717, 2021, doi: 10.1109/ACCESS.2021.3063234.

[13] Zhao L, Zhang A, Y Liu, et al. Encoding Multi-Granularity Structure Information for Joint Chinese Word Segmentation and POS Tagging[J]. Pattern Recognition Letters, 2020, 138(6).

[14] Ren Q, Yi L S, Wan W L. Research on the LSTM Mongolian and Chinese machine translation based on morpheme encoding[J]. Neural Computing and Applications, 2020, 32(1):41-49.

[15] Wang D, Fan J, Fu H, et al. Research on Optimization of Big Data Construction Engineering Quality Management Based on RNN-LSTM[J]. Complexity, 2018, 2018:1-16.

[16] Yin, S., Li, H. & Teng, L. Airport Detection Based on Improved Faster RCNN in Large Scale Remote Sensing Images [J]. Sensing and Imaging, vol. 21, 2020. https://doi.org/10.1007/s11220-020-00314-2

[17] Forster K I, Olbrei I. Semantic heuristics and syntactic analysis[J]. Cognition, 1973, 2(3):319-347.

[18] Dao P, Heinrichjosties E, Boroson T. Automated Algorithms to Identify Geostationary Satellites and Detect Mistagging using Concurrent Spatio-Temporal and Brightness Information[J]. Journal of Development Studies, 2005, 41(6):937-970.

[19] Huang P . On Chinglish in Chinese-English Translation and Its Countermeasures: Taking Translation of Modern Chinese Prose as an Example[J]. Open Access Library Journal, 2021, 08(6):1-8.

[20] Huang P. On Chinglish in Chinese-English Translation and Its Countermeasures: Taking Translation of Modern Chinese Prose as an Example[J]. Open Access Library Journal, 2021, 08(6):1-8.

[21] Wardell V, Esposito C L, Madan C R, et al. Semi-automated transcription and scoring of autobiographical memory narratives[J]. Behavior Research Methods, 2020, 53(2).

[22] Shoulin Yin, Hang Li, Lin Teng, Man Jiang & Shahid Karim. An optimised multi-scale fusion method for airport detection in large-scale optical remote sensing images [J]. International Journal of Image and Data Fusion, vol. 11, no. 2, pp. 201-214, 2020. DOI: 10.1080/19479832.2020.1727573

[23] Xiaowei Wang, Shoulin Yin, Desheng Liu, Hang Li & Shahid Karim. Accurate playground localisation based on multi-feature extraction and cascade classifier in optical remote sensing images [J]. International Journal of Image and Data Fusion, vol. 11, no. 3. pp. 233-250, 2020. DOI: 10.1080/19479832.2020.1716862

[24] Jing Yu, Hang Li, Shoulin Yin. Dynamic Gesture Recognition Based on Deep Learning in Human-to-Computer Interfaces [J]. Journal of Applied Science and Engineering, vol. 23, no. 1, pp.31-38, 2020.

[25] Hussein M, Hassan H A, Nassef M. A Trait-based Deep Learning Automated Essay Scoring System with Adaptive Feedback[J]. International Journal of Advanced Computer Science and Applications, 2020, 11(5).

[26] Bshary R, Oliveira R F. Cooperation in animals: toward a game theory within the framework of social competence[J]. Current Opinion in Behavioral Sciences, 2015, 3:31-37.

[27] Zhang D, Xu H, Su Z, et al. Chinese comments sentiment classification based on word2vec and SVMperf[J]. Expert Systems with Applications, 2015, 42(1):1857-1863.

[28] Wang X, Xu Y, Wang Y, et al. Representational similarity analysis reveals task-dependent semantic influence of the visual word form area[J]. Scientific Reports, 2018, 8(1):3047.

[29] Shoulin Yin, Jie Liu and Lin Teng. Improved Elliptic Curve Cryptography with Homomorphic Encryption for Medical Image Encryption[J]. International Journal of Network Security, Vol. 22, No. 3, pp. 419-424, 2020.

[30] Lin Teng, Hang Li, Shoulin Yin, Shahid Karim &Yang Sun. An active contour model based on hybrid energy and fisher criterion for image segmentation[J]. International Journal of Image and Data Fusion. Vol.11, No. 1, pp. 97-112. 2020.

[31] Rahman M H, Xie C, Sha Z. Predicting Sequential Design Decisions Using the Function-Behavior-Structure Design Process Model and Recurrent Neural Networks[J]. Journal of Mechanical Design, 2021, 143(8):1-46.

[32] Shoulin Yin and Jing Bi. Medical Image Annotation Based on Deep Transfer Learning[C]. 2018 IEEE International Congress on Cybermatics i-Things. Halifax, NS, Canada. DOI: 10.1109/Cybermatics_2018.2018.00042. 2019.6

[33] Yang Sun, Shoulin Yin, Hang Li, Lin Teng, Shahid Karim. GPOGC: Gaussian Pigeon-Oriented Graph Clustering Algorithm for Social Networks Cluster [J]. IEEE Access. Volume: 7, Page(s): 99254 - 99262, 03 July 2019.

[34] Shoulin Yin, Ye Zhang, Shahid Karim. Large Scale Remote Sensing Image Segmentation Based on Fuzzy Region Competition and Gaussian Mixture Model[J]. IEEE Access. volume 6, pp: 26069 - 26080, 2018.

[35] Shoulin Yin, Ye Zhang. Singular value decomposition-based anisotropic diffusion for fusion of infrared and visible images[J]. International Journal of Image and Data Fusion, 10(2), pp: 140-163, 2019.

[36] Yin Shoulin, Liu Jie, Teng Lin. A new krill herd algorithm based on SVM method for road feature extraction[J]. Journal of Information Hiding and Multimedia Signal Processing, vol.9, no.4, pp. 997-1005, July 2018.