WILEY | Hindawi

Research Article

Design and Application of Chinese English Machine Translation Model Based on Improved Bidirectional Neural Network Fusion Attention Mechanism

Qin Qin (ID)

Henan Polytechnic University, Jiaozuo 454000, China

Correspondence should be addressed to Qin Qin; qinqin@hpu.edu.cn

In order to improve the effect of Chinese-English machine translation, this paper combines attention mechanism and neural network algorithm and applies it to Chinese-English machine learning translation. Moreover, this paper uses Gaussian distribution instead of chi-square distribution to analyze the approximate error introduced by the Chinese and English speech energy detection method. In addition, this paper studies the overall and specific approximation errors by establishing the normalized mean square error function and the absolute error function, respectively. Finally, this paper proposes a new model for machine translation based on logarithmic position representation and self-attention mechanism. Through the experimental research, it can be seen that the Chinese-English machine translation model integrating attention mechanism and bidirectional neural network proposed in this paper has a good practical translation effect.

1. Introduction

In the process of processing Chinese and English languages, it is relatively easy to split English sentences [1]. The hardest part is reorganizing the simple sentences that have been split up, and it is easy to be too limited by the grammatical structure of English to make bold changes. At this time, nouns in the original text are translated into nouns in the target language, verbs are translated into verbs [2], and adverbs are translated into adverbs. Even, sometimes an attributive clause, we will copy the structure of the original text. In this way, the translated Chinese can only be obscure and incomprehensible, the words do not convey the meaning, and do not conform to the Chinese language habits. Moreover, the Chinese translated in this way has a strong translation cavity. Therefore, in fact, we should try our best to avoid falling into the trap of translation cavity [3]. One of the important translation methods is part-of-speech conversion, and almost any part-of-speech in the original text can be converted into other parts of speech when it is converted into the target language, such as verbs into nouns, nouns into verbs, prepositions or adverbs into verbs, and nouns into

adjectives. The conversion of parts of speech is a necessary means of doing a good job in translation. It can make the translation break away from the structure of the original text and not be bound by the original text. Therefore, the translated text will naturally avoid translation traces as much as possible.

This paper combines the attention mechanism and neural network algorithm and applies it to Chinese-English machine learning translation to improve the effect of Chinese-English machine translation and provide a reference for the development of subsequent artificial intelligence translation systems.

2. Related Work

The convolutional neural network proposed in [4] achieves the best performance on object classification in translation datasets. With the further development of artificial intelligence translation, most of the studies using convolutional neural networks to learn high-level semantic features in intelligent translation have been proposed to complete specific tasks such as classification [5]. Among these methods,

most of them use repeated several layers of convolutional networks, followed by a pooling layer after each convolutional network, and finally select more important features and output them through a fully connected layer [6]. Reference [7] proposed another classic convolutional neural network, which achieved good results in translation localization and classification tasks, and brought a new reference standard for the structural design of deep models. Reference [8] used a smaller convolution kernel and designed the model to be deeper, so that it could better learn the high-level semantics of intelligently translated content. Reference [9] deepens the depth of the model and improves the overall performance. Generally speaking, if the convolutional network wants to improve the expression learning ability, it mainly depends on increasing the number of output channels, which will consume a lot of computing resources and easily cause overfitting. Reference [10] uses a convolutional branch network to organize information across channels, which not only improves the performance of the model but also greatly reduces the amount of parameters that increase sharply due to the deepening of the depth and width of the model.

With the development of neural network research, some studies have begun to build neural topic models to represent local intelligent translation semantics, that is, object-level intelligent translation content topic models are derived from natural language processing and can discover or learn a class of statistical models for abstract topics in documents [11]. Reference [12] proposed an unsupervised neural topic model document neural autoregressive distribution estimator, which fuses a topic model and a neural network, assuming that the generation of each word is only related to the words generated before it and directly builds. The modulo document is the product of the conditional probabilities of all words, and each conditional probability is generated using a feedforward neural network to obtain object-level topic features with high-level semantic information. In the literature [13], the DocNADE model was introduced into the field of translation understanding, and it was proposed to do translation classification and annotation based on the DocNADE model, and it also achieved good performance. The translation understanding work based on deep learning has achieved fruitful results, shortening the translation semantic gap to a certain extent, but there is still some room for improvement. Reference [14] deepened the network depth to an unprecedented 152-layer M, directly reducing the translation classification error rate on ImageNet to less than 5%, approaching or even surpassing the human recognition accuracy rate to an unprecedented peak of development. With the help of deep learning, the task of object-level analysis has also achieved great breakthroughs. Object detection is the basic method of object-level analysis, and it has also entered a stage of rapid development. In the past ten years, the traditional machine intelligence translation field usually uses feature descriptors for object recognition tasks, and the progress is slow [15]. Due to the excellent performance of convolutional neural networks on ImageNet, the literature [16] general-ized the ability of AlexNet to recognize objects based on ImageNet training to the field of object detection, which is transfer learning. The proposal of R-CNN is a milestone in the application of convolutional neural networks to the field of target detection, and it is the pioneering work of using deep learning for target detection. After this, a series of far-reaching object detection models based on R-CNN, and regression methods based entirely on deep learning [17]. It is precisely with the help of the deep learning model that under the training of a large amount of ImageNet data, it has the ability to better capture the high-level semantics of intelligent translation content, so that the model can also achieve ideal results when migrating to other tasks or datasets [18]. Transfer learning lays the foundation for the improvement of most tasks in computer intelligent translation. Not only in the field of object detection but also in other computer intelligent translation tasks, transfer learning is also widely used, especially in the task of mining intelligent translation semantic information, including translation-based translation description and translation question answering, as well as video-based video, description, video understanding, and video question answering. As a result, the research progress of intelligent translation semantic understanding task has been greatly promoted [19].

3. Intelligent Chinese and English Audio Spectrum Perception

The principle of Chinese and English sound waveform energy detection is mainly based on the chi-square distribution theory. The Chinese and English sound waveform energy Y of the received signal of cognitive users obeys the chi-square distribution:

$$Y \sim \begin{cases} \chi^2_{2TW}, & H_0, \\ \chi^2_{2TW}(2\gamma), & H_1. \end{cases} \tag{1}$$

Among them, γ represents the signal-to-noise ratio of the received signal of the cognitive user, χ^2_{2TW} and $\chi^2_{2TW}(2\gamma)$, respectively, represent the central chi-square distribution and the noncentral chi-square distribution, both of which have 2TW degrees of freedom, and the parameter of $\chi^2_{2TW}(2\gamma)$ is 2γ. For the simplicity of the description, u is uniformly used in the following text to represent the time-bandwidth product TW in the above formula [20].

$$Y \sim \begin{cases} \chi^2_{2u}, & H_0, \\ \chi^2_{2u}(2\gamma), & H_1. \end{cases} \tag{2}$$

In the additive white Gaussian noise (AWGN) channel, the following important probability formulas can be obtained according to the above probability distribution

functions, and these formulas can also express the spectrum sensing performance of cognitive users:

$$\text{Detection probability}: P_d = P\left\{Y > \frac{\lambda}{H_1}\right\} = Q_u\left(\sqrt{2\gamma}, \sqrt{\lambda}\right), \tag{3}$$

The detection probability P_d represents the probability that a cognitive user can correctly detect the presence of an authorized user. The high detection probability indicates that cognitive users can accurately perceive the presence of authorized users, to maintain the sensing state to continue detection or release the spectrum in use and reduce interference to authorized users.

$$\text{False alarm probability}: P_f = P\left\{Y > \frac{\lambda}{H_0}\right\} = \frac{\Gamma(u, \lambda/2)}{\Gamma(u)}. \tag{4}$$

The false alarm probability P_f represents the false alarm (type 1 error) of the cognitive user when the authorized user does not exist, that is, the probability that the cognitive user thinks that the authorized user exists at this time. The low probability of false alarm indicates that cognitive users can more accurately perceive spectrum holes, so as to perform dynamic spectrum access.

$$\text{Missed alarm probability}: P_m = P\left\{Y \leq \frac{\lambda}{H_1}\right\} = 1 - P_d. \tag{5}$$

The missed alarm probability P_m represents the false alarm (type 2 error) of the cognitive user when the authorized user exists, that is, the probability that the cognitive user cannot detect the existence of the authorized user. For the same cognitive device, the probability of false alarm and probability of detection add up to a constant 1.

$\Gamma(a)$ and $\Gamma(a, b)$ in formula (4) represent the gamma function and incomplete gamma function, respectively, and $Q_u(a, b)$ in formula (3) represents the generalization Q function. The λ in the above two formulas is the threshold value of the Chinese and English sound waveform energy detector.

Many research methods on cooperative spectrum sensing are based on a centralized cognitive wireless network, that is, the network consists of a decision center and N cognitive users, and each cognitive user reports its own spectrum sensing results to the decision center. The decision center fuses the above information using information fusion criteria such as Chinese and English sound waveform energy fusion or decision fusion, to make a final decision to determine whether an authorized user exists. Decision fusion strategies can be mainly divided into the following categories:

"Majority" rule (majority-rule): if more than a certain number of cognitive users believe that authorized users exist, that is, the local decision is H_1, then the final decision of the decision center is H_1.

AND-rule: when all cognitive users in the network need to decide H_1, the decision center decides H_1.

OR-rule: as long as there is a cognitive user in the network to decide H_1, the decision center decides H_1.

This research will combine the existing Chinese and English phonetic waveform energy fusion and decision fusion criteria to propose a two-layer data fusion mechanism. The "or" criterion is used as the fusion criterion in the upper-layer decision fusion, and the method is also used as the comparison object in the simulation analysis. The cooperative spectrum sensing performance of cognitive networks using the OR criterion is listed below as [21].

$$Q_d = 1 - \prod_{i=1}^{N}(1 - P_{d,i}), \tag{6}$$

$$Q_m = \prod_{i=1}^{N} P_{m,i}, \tag{7}$$

$$Q_f = 1 - \prod_{i=1}^{N}\left(1 - P_{f,i}\right). \tag{8}$$

In the formula, Q_d, Q_f, Q_m represents the detection probability, false alarm probability, and missed alarm probability when N cognitive users in the cognitive network perform cooperative sensing, respectively. $P_{d,i}, P_{f,i}, P_{m,i}$ is the same as the above and represents the local detection probability, false alarm probability, and missed alarm probability of the ith cognitive user (SU_i).

In the traditional Chinese and English sound waveform energy detector, the cognitive user first passes the received signal through a band-pass filter, takes out the signal of a certain target frequency band, and then, performs a square operation on it. After that, it accumulates the energy value of Chinese and English sound waveforms in a period of time through the integrator and finally compares it with the preset threshold value (single threshold), to make a local decision on whether the authorized user exists. The single-threshold detection model is shown in Figure 1(a).

O_i in the figure represents the energy of the English sound waveform in the received signal of the cognitive user SU_i, and the decisions H_0 and H_1 of the existence of an authorized user are determined by comparing the Chinese and English sound waveform energy O_i with the threshold value λ_i. When $O_i \leq \lambda_i$, it is decided as H_1, when $O_i > \lambda_i$, it is decided as H_0.

In the double-threshold detection model, each cognitive user compares the energy of the Chinese and English sound waveforms in the received signal according to the preset two thresholds $\lambda_{1,i}$ and $\lambda_{2,i}$ in the Chinese and English sound waveform energy detector, so as to make a local decision, as shown in Figure 1(b). According to the above description, the local decision criterion of this method is as follows:

Decision: $H_1 : O_i > \lambda_{2,i}$
Decision: $H_0 : O_i \leq \lambda_{1,i}$
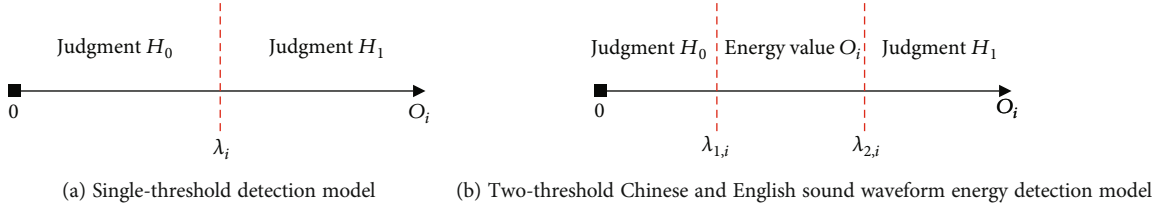Delayed decision: $\lambda_{1,i} < O_i \leq \lambda_{2,i}$

FIGURE 1: Single-threshold and dual-threshold detection models.

When the energy O_i of the English sound waveform in the received signal obtained by SU_i is located in the interval $\lambda_{1,i} < O_i \le \lambda_{2,i}$, SU_i thinks that the energy value of the Chinese and English sound waveform is relatively modular, and it cannot determine whether an authorized user exists or not according to its size. Therefore, the O_i value is reported to the decision center as it is. After obtaining the local decision results reported by all cognitive users in the cognitive network and the English sound waveform energy in the received signal, the decision center performs two-level information fusion to obtain the final decision result of the existence of authorized users. The specific spectrum sensing methods and information fusion rules are as follows:

$SU_i(i = 1,\cdots,N)$ first performs local spectrum detection within its own range and obtains the output O_i of the Chinese and English phonetic waveform energy accumulator. If O_i satisfies $O_i \le \lambda_{1,i}$ or $O_i > \lambda_{2,i}$, it is decided as H_0 or H_1, respectively. If O_i is located in the delayed decision interval, that is, $\lambda_{1,i} < O_i \le \lambda_{2,i}$, the cognitive user does not make a local decision result, but uploads O_i to the decision center and waits for its delayed decision. From this, it can be seen that the local perception result R_i reported by the cognitive user to the decision center consists of two kinds of information: the local decision and the energy value of the Chinese and English sound waveforms, namely,

$$R_i = \begin{cases} O_i, & \lambda_{1,i} < O_i \le \lambda_{2,i}, \\ L_i, & \text{other}, \end{cases} \tag{9}$$

$$L_i = \begin{cases} 0, & 0 \le O_i \le \lambda_{1,i}, \\ 1, & O_i > \lambda_{2,i}. \end{cases} \tag{10}$$

In order to facilitate the theoretical derivation of detection performance in the following, and without loss of generality, it is assumed here that the decision center receives N data packets reported by all cognitive users, including K L_i and $-KO_i$. Then, the decision center uses these $N-KO$ values and uses the Chinese and English sound waveform energy fusion method to obtain the delayed decision D:

$$D = \begin{cases} 0, & 0 \le \sum_{i=1}^{N-K} O_i \le \lambda, \\ 1, & \sum_{i=1}^{N-K} O_i > \lambda. \end{cases} \tag{11}$$

The λ in the expression of the delay decision D represents the threshold value obtained by the decision center according to the user's communication requirements and the required false alarm probability set by the spectrum sensing performance. From the cognitive user perception process described above, it can be seen that this method considers that $N-K$ users among all N cognitive users cannot make local decisions based on the energy of the English sound waveform in the received signal output by the English sound waveform energy accumulator. Therefore, these cognitive users report the Chinese and English sound waveform energy values of the received signal directly to the decision center without processing, and the decision center uses a large amount of historical perception data in the past to make a delayed decision based on the $N-K$ Chinese and English sound waveform energy values. It is equivalent to merging Chinese and English phonetic waveform energy on these values and replacing the unreliable local decision results of $N-K$ cognitive users. The Chinese and English sound waveform energy value $\sum_{i=1}^{N-K} O_i$ obtained by the decision center obeys the following chi-square distribution:

$$\sum_{i=1}^{N-K} O_i \sim \begin{cases} \chi^2_{2(N-K)u}, & H_0, \\ \chi^2_{2(N-K)u}(2\gamma_0), & H_1. \end{cases} \tag{12}$$

Among them, $\gamma_0 = \sum_{i=1}^{N-K} \gamma_i$ represents the summation of the signal-to-noise ratios of the N-K cognitive users who reported the energy values of the Chinese and English sound waveforms. Finally, according to the local decision results L_i of K cognitive users and the above-mentioned delayed decision D, the final decision obtained by the decision center using the "or" criterion for decision fusion is:

$$F = \begin{cases} 1, & D + \sum_{i=1}^{K} L_i > 1, \\ 0, & \text{other}. \end{cases} \tag{13}$$

Among them, $F = 1$ corresponds to H_1, indicating that the decision center finally determines that the authorized user exists. On the contrary, it corresponds to H_0 and determines that the authorized user does not exist.

Based on the dual-threshold Chinese and English sound waveform energy detection method discussed above, the

spectral sensing performance is analyzed theoretically. Similar to the previous article, $P_{d,i}, P_{f,i}, P_{m,i}$ is still used here to represent the local detection probability, false alarm probability, and missed alarm probability of SU_i itself. Here, two probability values $\Delta_{0,i}$ and $\Delta_{1,i}$ are used to denote the probability of SU_i that the energy value of the Chinese and English speech waveforms falls into the delay decision interval under the two assumptions of the existence of an authorized user or not, so as to facilitate the subsequent discussion. Its expression is:

$$\Delta_{0,i} = P\left\{\lambda_{1,i} < O_i \le \frac{\lambda_{2,i}}{H_0}\right\}, \qquad (14)$$

$$\Delta_{1,i} = P\left\{\lambda_{1,i} < O_i \le \frac{\lambda_{2,i}}{H_1}\right\}. \qquad (15)$$

Therefore, the aforementioned local spectrum sensing performance of cognitive users can be modified as:

$$P_{d,i} = P\left\{O_i > \frac{\lambda_{2,i}}{H_1}\right\} = Q_u\left(\sqrt{2\gamma_i}, \sqrt{\lambda_{2,i}}\right), \qquad (16)$$

$$P_{m,i} = P\left\{O_i \le \frac{\lambda_{1,i}}{H_1}\right\} = 1 - \Delta_{1,i} - P_{d,i}, \qquad (17)$$

$$P_{f,i} = P\left\{O_i > \frac{\lambda_{2,i}}{H_0}\right\} = \frac{\Gamma(u, \lambda_{2,i}/2)}{\Gamma(u)}. \qquad (18)$$

The same symbol Q_d, Q_f, Q_m is still used here to represent the detection probability, false alarm probability, and missed alarm probability when all cognitive users in the network cooperate:

$$Q_m = \sum_{K=0}^{N-1} \binom{N}{K} \prod_{i=1}^{K} P_{m,i} \prod_{i=K+1}^{N} \Delta_{1,i}\left[1 - Q_{(N-K)u}\left(\sqrt{2\gamma_0}, \sqrt{\lambda}\right)\right] + \prod_{i=1}^{N} P_{m,i}, \qquad (19)$$

$$Q_f = 1 - \prod_{i=1}^{N}\left(1 - \Delta_{0,i} - P_{f,i}\right) - \sum_{K=0}^{N-1} \binom{N}{K} \prod_{i=1}^{K}$$
$$\cdot \left(1 - \Delta_{0,i} - P_{f,i}\right) \prod_{i=K+1}^{N} \Delta_{0,i}\left[1 - \frac{\Gamma[(N-K)u, \lambda/2]}{\Gamma[(N-K)u]}\right], \qquad (20)$$

$$Q_d = 1 - Q_m. \qquad (21)$$

It can be seen from the above formula that the probability value of O_i between $\lambda_{1,i}$ and $\lambda_{2,i}$ under the two assumptions of the existence of authorized users has a very important influence on the spectrum sensing performance of the dual-threshold Chinese and English tone waveform energy detector. When $\Delta_{0,i}$ and $\Delta_{1,i}$ are both equal to zero, the method proposed in this paper degenerates into the traditional single-threshold Chinese and English sound waveform energy detection method. At this time, all cognitive

users in the cognitive network report their local decisions to the decision center. Moreover, the decision center also uses the "or" criterion to perform decision fusion on all local decision results based on only this kind of data, so as to obtain the final result.

According to the perception process and realization principle of the aforementioned dual-threshold Chinese and English tone waveform energy detector, this part will use the method of simulation analysis to analyze and discuss its detection performance in detail. The simulation is presented in the form of ROC (receiver-operating characteristic curve) and CROC (complementary receiver-operating characteristic curve). Two simulation scenarios are set here; the main difference is the difference of the Δ parameter, which are, respectively, set as: $\Delta_{0,i} = \Delta_{1,i} = 0.01$ and $\Delta_{0,i} = \Delta_{1,i} = 0.1$.

The comparison objects in the two scenarios are the traditional single-threshold detection methods when $\Delta_{0,i} = \Delta_{1,i} = 0.1$, and the rest of the parameters are set as follows:

$$N = 10,$$
$$\gamma_1 = \gamma_2 = \cdots = \gamma_N = 10\,\text{dB}, \qquad (22)$$
$$u = 5.$$

We first examine the curves in Figure 2. When Δ is small, the dual-threshold detection method improves the spectrum sensing performance less compared to the traditional method. When $\Delta_{0,i}$ and $\Delta_{1,i}$ are further increased, as shown in Figure 3, the improvement of spectrum sensing performance by this method becomes more significant. At this time, the probability that the energy value of the English sound waveform in the received signal falls into the delay decision interval is greater (0.1) than the first scenario. By carefully observing Figure 3, it can also be seen that when Q_f is equal to 10^{-4}, the detection probability of this method can obtain a gain of about 1.23 times compared with the traditional single-threshold detection method. Also, the gain in detection probability can be further increased as the false alarm probability decreases.

This paper studies the approximation error introduced by using Gaussian distribution instead of chi-square distribution in Chinese and English phonetic waveform energy detection methods. The overall and specific approximation errors are studied by establishing the normalized mean square error function and the absolute error function, respectively. Moreover, the reasonable lower limit of the degree of freedom of the chi-square distribution when the above approximation is performed is given by numerical simulation, and the corresponding chi-square distribution probability value when the approximation error is the smallest is listed.

Spectrum sensing technology based on Chinese and English speech waveform energy detection has the advantages of low cost and simple deployment, making it increasingly a research hotspot in this field. In these related studies, it is necessary to use the chi-square distribution, a
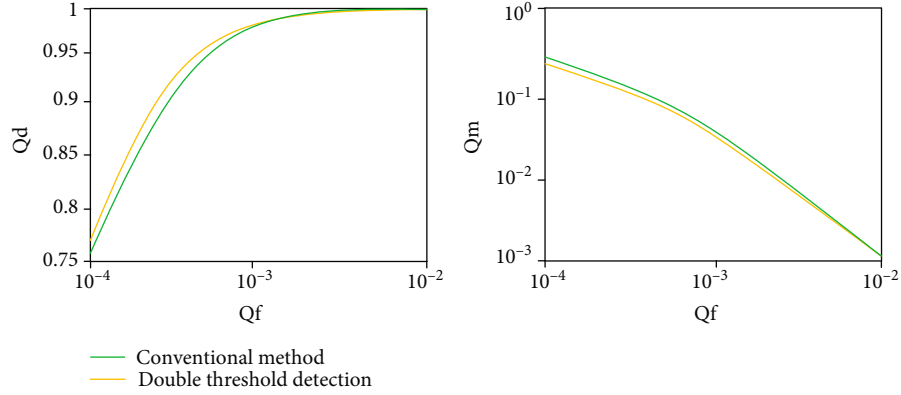
FIGURE 2: The working characteristic curve and complementary working characteristic curve of the receiver, $\Delta_{0,i} = \Delta_{1,i} = 0.01$.
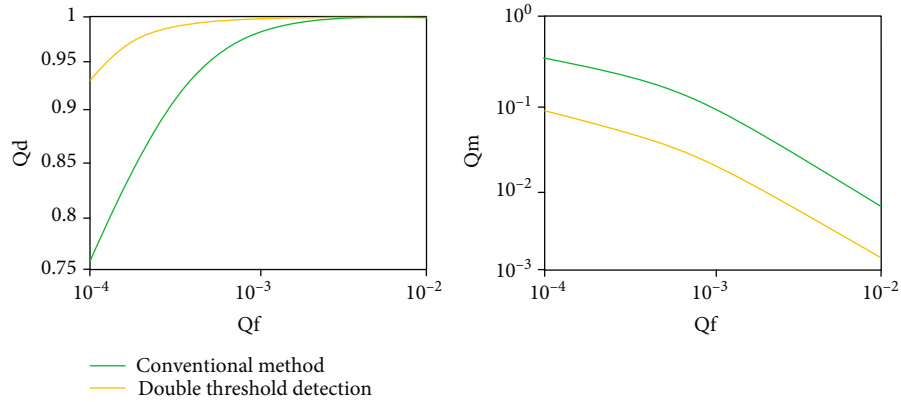

FIGURE 3: The working characteristic curve and complementary working characteristic curve of the receiver, $\Delta_{0,i} = \Delta_{1,i} = 0.1$.
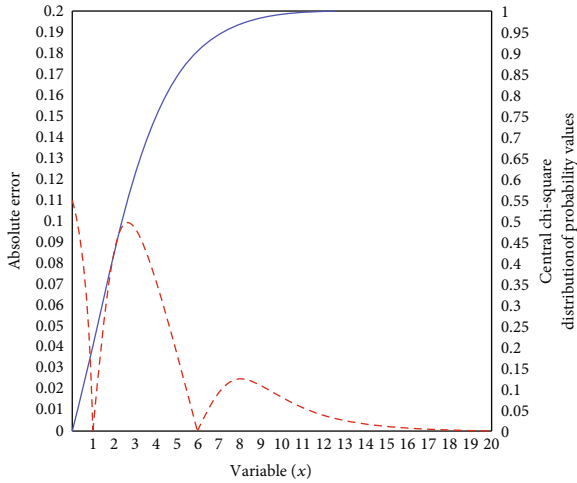

FIGURE 4: Absolute error curve and central chi-square distribution CDF curve when H_0 ($m = 3$).

mathematical tool. To analyze the performance of the Chinese and English sound waveform energy detectors: the central chi-square distribution corresponds to the hypothesis H_0, and the noncentral chi-square distribution corresponds to the hypothesis H_1.

The English sound waveform energy in the signal observed by the receiver should obey the following chi-square distribution.

$$O \sim \begin{cases} x_{2TW}^2, & H_0, \\ x_{2TW}^2(2\gamma), & H_1. \end{cases} \quad (23)$$

Among them, O represents the energy value of the English sound waveform in the signal received by the cognitive user, and χ_{2TW}^2 and $\chi_{2\pi W}^2(2\gamma)$ are the central and noncentral chi-square distributions with 2TW degrees of freedom, respectively. The noncentral parameter of the latter is 2γ, corresponding to the signal-to-noise ratio of the cognitive user receiver. For brevity of the description below, we use m for the delay-bandwidth product 2TW and mr for the degrees of freedom of the chi-square distribution, and we obtain:

$$O \sim \begin{cases} \chi_m^2, & H_0, \\ \chi_m^2(mr), & H_1. \end{cases} \quad (24)$$

If m is assumed to be large enough (this condition is easily satisfied when multiple cognitive users perform
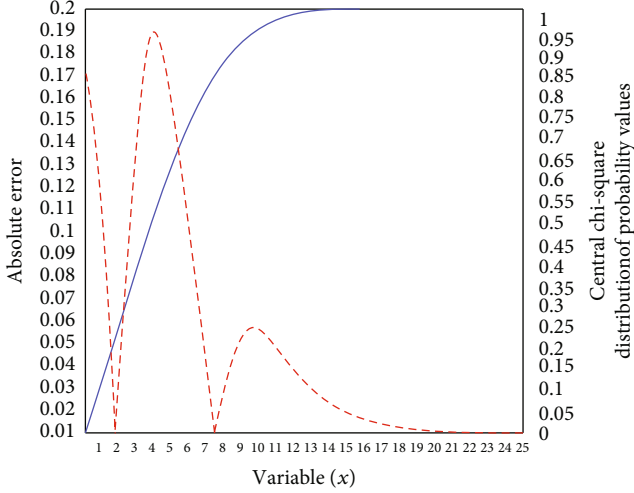
FIGURE 5: Absolute error curve and noncentral chi-square distribution CDF curve when H_1 ($m = 4$, $r = -20$ dB).

cooperative sensing), then according to the central limit theorem, O should tend to obey a Gaussian normal distribution as follows:

$$O \sim \begin{cases} N(m, 2m), & H_0, \\ N(m(1 + r), 2m(1 + 2r)), & H_1. \end{cases} \tag{25}$$

Among them, $O^\sim$ represents the approximate value of the English sound waveform energy O in the received signal, and $N(\mu, \sigma^2)$ represents the Gaussian normal distribution with mean μ and variance σ^2.

According to formulas (24) and (25), we can obtain the cumulative probability distribution functions (CDF) of O and $O^\sim$ under two assumptions:

$$\begin{cases} P_o(x \, ; m) = \int_0^x \dfrac{t^{m/2-1} e^{-t/2}}{2^{m/2} \Gamma(m/2)} \, dt, & H_0, \\ P_O(x \, ; m, r) = \sum_{j=0}^{\infty} e^{-\lambda/2} \dfrac{(\lambda/2)^j}{j!} P_o(x \, ; m + 2j), & H_1, \end{cases} \tag{26}$$

$$\begin{cases} P_{\bar{o}}(x \, ; m) = N(x \, ; m, 2m), \\ \quad = \dfrac{1}{2\sqrt{m\pi}} \int_{-\infty}^x e^{-(t-m)^2/4m} dt, \\ P_{\bar{o}}(x \, ; m, r) = N(x \, ; m(1 + r), 2m(1 + 2r)), \\ \quad = \dfrac{1}{2\sqrt{m\pi(1+2r)}} \int_{-\infty}^x e^{-[t-m(1+r)]^2/4m(1+2r)} dt H_0. \end{cases} \tag{27}$$

Among them, $\Gamma(x)$ represents the gamma function. From this, we can define the following absolute error function and normalized mean square error function:

$$\begin{cases} \text{AEF}(x \, ; m) = |P_o(x \, ; m) - P_{\bar{o}}(x \, ; m)|, & H_0, \\ \text{AEF}(x \, ; m, r) = |P_o(x \, ; m, r) - P_{\bar{o}}(x \, ; m, r)|, & H_1, \end{cases} \tag{28}$$

$$\begin{cases} \text{NEF}(m) = \dfrac{\sum_{i=1}^n [\text{AEF}(x_i \, ; m)]^2}{\sum_{i=1}^n [P_o(x_i \, ; m)]^2} \\ \quad = \dfrac{\sum_{i=1}^n [P_o(x_i \, ; m) - P_{\bar{o}}(x_i \, ; m)]^2}{\sum_{i=1}^n [P_o(x_i \, ; m)]^2}, & H_0, \\ \text{NEF}(m, r) = \dfrac{\sum_{i=1}^n [\text{AEF}(x_i \, ; m)]^2}{\sum_{i=1}^n [P_O(x_i \, ; m, r)]^2} \\ \quad = \dfrac{\sum_{i=1}^n [P_O(x_i \, ; m, r) - P_{\bar{o}}(x_i \, ; m, r)]^2}{\sum_{i=1}^n [P_O(x_i \, ; m, r)]^2}, & H_1. \end{cases} \tag{29}$$

In formulas (28) and (29), n represents the number of variables x, and i represents the subscript of each x. It can be seen from the expressions of AEF and NEF that NEF masks the influence of the variable x on the error, so NEF only depicts the approximate trend of the approximation error as a function of degrees of freedom. In contrast, AEF reveals precisely the exact effect of both x and m on the approximation error.

As mentioned earlier, NEF does not reveal the exact effect of x and m on the error when acting at the same time, but only shows the approximate variation law of the error for different variables x. Since there are two assumptions involved in spectrum sensing technology, the presence or absence of signals from authorized users (corresponding to H_1 and H_0), the errors in these two cases will be analyzed separately in the following sections.

The expression for NEF at H_0 is rewritten as follows:

$$\begin{aligned} \text{NEF}(m) &= \dfrac{\sum_{i=1}^n [\text{AEF}(x_i \, ; m)]^2}{\sum_{i=1}^n [P_o(x_i \, ; m)]^2} \\ &= \dfrac{\sum_{i=1}^n [P_o(x_i \, ; m) - P_{\bar{o}}(x_i \, ; m)]^2}{\sum_{i=1}^n [P_o(x_i \, ; m)]^2}. \end{aligned} \tag{30}$$

It can be seen that Equation (30) contains a fraction with two square sum functions, and each function contains an integral (see Equations (26) and (27)). In particular, $P_o(x_i \, ; m)$ contains the expression of the gamma function, so the theoretical analysis of the shape of the NEF function will be very complicated. In order to find another way, we use numerical calculation to discuss the NEF, and the parameters in the simulation are set as:

$$\begin{aligned} m &\in [1, 50], \\ x &\in [0, 200]. \end{aligned} \tag{31}$$

The expression for NEF at time H_1 is rewritten as follows:

$$\begin{aligned} \text{NEF}(m, r) &= \dfrac{\sum_{i=1}^n [\text{AEF}(x_i \, ; m)]^2}{\sum_{i=1}^n [P_o(x_i \, ; m, r)]^2} \\ &= \dfrac{\sum_{i=1}^n [P_o(x_i \, ; m, r) - P_{\bar{o}}(x_i \, ; m, r)]^2}{\sum_{i=1}^n [P_o(x_i \, ; m, r)]^2}. \end{aligned} \tag{32}$$
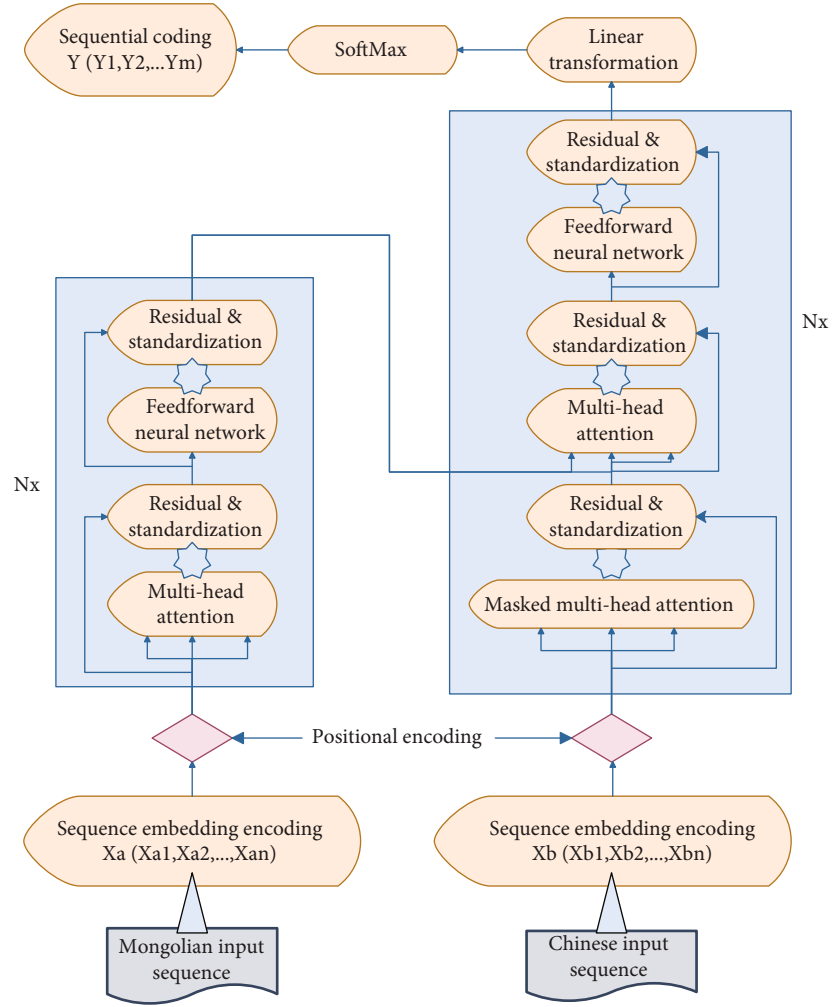
FIGURE 6: Attention model.

The difference from the previous assumption is that NEF at H_1 is a binary function of degrees of freedom m and a noncentral parameter r representing the real-time signal-to-noise ratio of a cognitive user. Therefore, in this subsection, we will discuss the impact of the above two variables on NEF.

Similar to the previous, we also investigate the morphology of NEF by numerical calculation. The parameters in formula (32) are set as:

$$m \in [1, 20],$$
$$x \in [0, 90]. \tag{33}$$

In addition, according to the recommendation of IEEE 802.22 WRAN, since the signal-to-noise ratio of the received signal of the cognitive user is very poor when the signal of the authorized user exists, the noncentral parameter r is set to -20 dB, -10 dB, and 0 dB.

First, we rewrite formula (28) at H_0 to simplify the following discussion, where m is fixed at 3 and is not explicitly expressed in the formula:

$$\mathrm{AEF}(x) = \left| P_O(x) - P_{\bar{O}}(x) \right. \tag{34}$$

Since both $P_o(x)$ and $P_{\bar{o}}(x)$ are cumulative probability distribution functions, they have limits of the form:

$$\lim_{x \longrightarrow \infty} P_o(x) = \lim_{x \longrightarrow \infty} P_{\bar{o}}(x) = 1. \tag{35}$$

Therefore, the limit of AEF is:

$$\lim_{x \longrightarrow \infty} \mathrm{AEF}(x) = 0. \tag{36}$$

Formula (36) shows that when x is quite large, the absolute error between the central chi-square distribution and the Gaussian distribution is negligible.

Figure 4 shows both the absolute error value and the probability value of the central chi-square distribution using a dual vertical axis plot, with the variable x on the horizontal
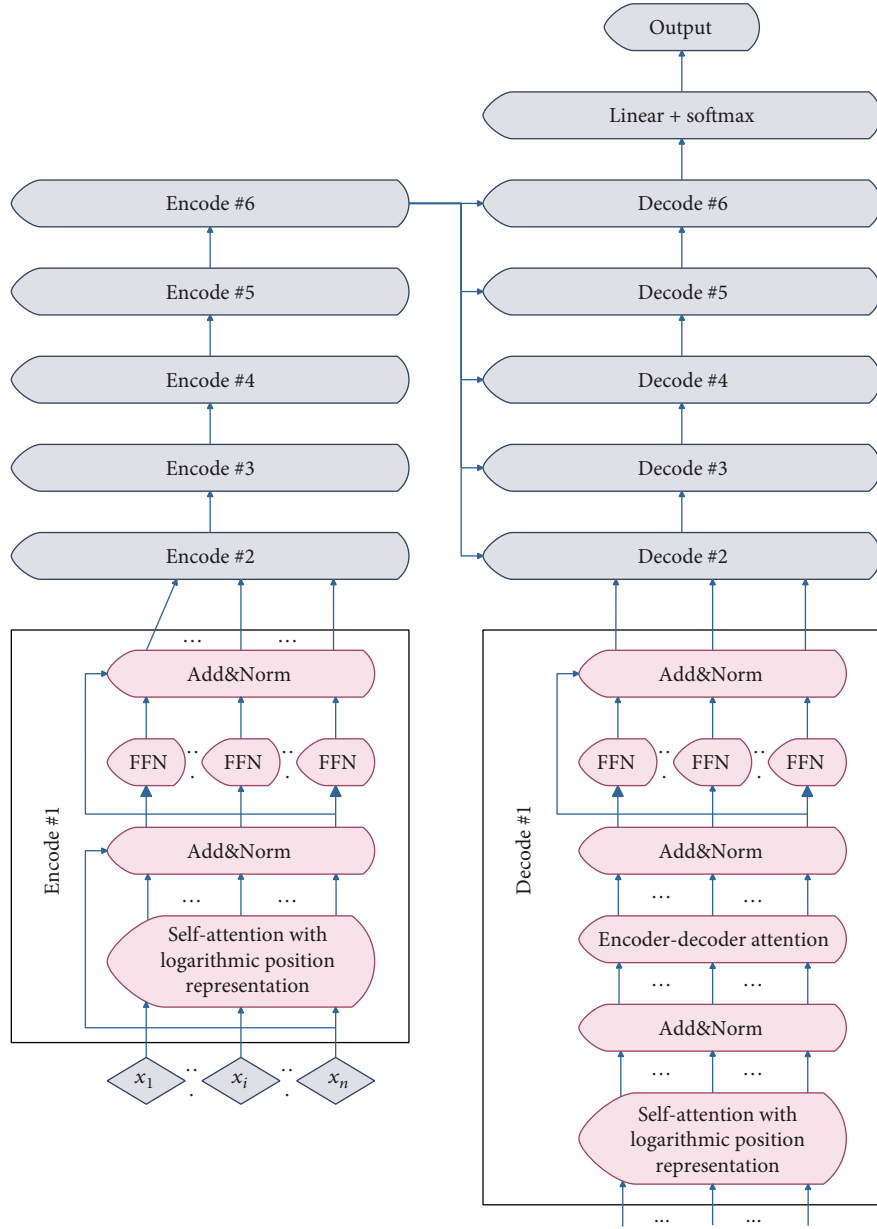
FIGURE 7: Chinese-English machine translation model integrating with attention mechanism and bidirectional neural network.

axis. The relationship between the above two can be clearly seen from the double vertical axis graph, and the absolute error will gradually become smaller than that shown in the above graph, especially in the area containing this special probability value. It can be seen from this that replacing the central chi-square distribution with the Gaussian distribution can obtain the best approximation effect at the value of the corresponding false alarm probability.

We rewrite formula (28) at time H_1 as follows to simplify the subsequent discussion. Among them, r is used as a parameter (set to -20 dB to represent poor channel conditions), and m is fixed at 4 (take a stricter lower limit than 3 in the previous section to show the difference).

Moreover, it is not explicitly expressed in the formula:

$$\text{AEF}(x;r) = |P_o(x;r) - P_{\bar{o}}(x;r)|. \tag{37}$$

Since both $P_o(x;r)$ and $P_{\bar{o}}(x;r)$ are cumulative probability distribution functions, they are both monotonically increasing functions of x and have limits of the form:

$$\lim_{x \to \infty} P_o(x;r) = \lim_{x \to \infty} P_{\bar{o}}(x;r) = 1. \tag{38}$$

TABLE 1: Evaluation of Chinese-English translation effect of Chinese-English machine translation model integrating with attention mechanism and bidirectional neural network.

Number	Translation effect	Number	Translation effect	Number	Translation effect
1	75.9	19	78.8	37	71.2
2	79.8	20	86.7	38	85.8
3	84.0	21	80.6	39	81.2
4	83.1	22	73.6	40	84.5
5	76.4	23	69.8	41	84.7
6	73.8	24	71.8	42	70.8
7	71.2	25	78.5	43	73.4
8	72.5	26	84.2	44	76.6
9	79.6	27	69.2	45	81.8
10	76.7	28	77.8	46	87.8
11	69.7	29	73.4	47	78.5
12	78.1	30	72.8	48	74.6
13	74.5	31	69.7	49	72.0
14	78.9	32	79.5	50	79.9
15	87.1	33	87.3	51	87.6
16	81.0	34	83.8	52	86.6
17	74.4	35	69.9	53	84.8
18	76.4	36	76.8	54	76.5

TABLE 2: Evaluation of English-Chinese translation effect of Chinese-English machine translation model integrating with attention mechanism and bidirectional neural network.

Number	Translation effect	Number	Translation effect	Number	Translation effect
1	85.0	19	73.4	37	73.6
2	89.7	20	76.7	38	79.0
3	89.4	21	73.6	39	74.2
4	76.4	22	83.7	40	81.1
5	87.6	23	83.5	41	73.3
6	81.9	24	88.8	42	85.1
7	80.4	25	75.1	43	81.9
8	90.0	26	81.5	44	89.7
9	88.9	27	89.9	45	83.2
10	90.8	28	78.0	46	89.4
11	73.4	29	84.4	47	85.1
12	78.2	30	84.0	48	82.6
13	76.9	31	82.0	49	77.2
14	89.8	32	83.8	50	80.8
15	84.4	33	86.7	51	89.3
16	80.6	34	83.0	52	79.5
17	88.0	35	82.3	53	77.6
18	78.3	36	80.5	54	82.6

Thus, the limit of ARF is obtained as:

$$\lim_{x \longrightarrow \infty} \text{AEF}(x;r) = 0. \tag{39}$$

Formula (39) shows that the absolute error between the noncentral chi-square distribution and the Gaussian distribution can be ignored when x is quite large.

Figure 5 shows both the absolute error value and the probability value of the noncentral chi-square distribution through a dual vertical axis graph, and the horizontal axis

is the variable x. The relationship between the above two can be clearly seen from the double vertical axis graph, and the absolute error will gradually become smaller than that shown in the above graph, especially in the area containing this special probability value. It can be seen that replacing the noncentral chi-square distribution with the Gaussian distribution can obtain the best approximation effect at the value of the corresponding detection probability.

4. The Chinese-English Machine Translation Model Integrating Attention Mechanism and Bidirectional Neural Network

The Chinese-English machine translation model constructed in this paper adopts the classic encoder-decoder architecture, and the internal structure is shown in Figure 6.

In this paper, a new model for machine translation based on logarithmic position representation and self-attention mechanism is proposed, as shown in Figure 7. In the encoder, there are self-attention combined with log position representation layer and fully connected FFN network layer. The decoder has self-attention combined with log position representation layer, encoder-decoder attention layer, and fully connected FFN network layer. The output layer contains a linear transformation layer and a Softmax fully connected layer.

On the basis of the above research, this paper verifies the effect of the Chinese-English machine translation model that combines the attention mechanism and the bidirectional neural network proposed in this paper. The translation effects of the Chinese-to-English and English-to-English translations of this model are evaluated through multiple sets of translation experiments, and the experimental results obtained are shown in Table 1 and Table 2, respectively.

From the above research, it can be seen that the Chinese-English machine translation model combining attention mechanism and bidirectional neural network proposed in this paper has a good practical translation effect.

5. Conclusion

There is a huge difference in language structure between Chinese and English. English focuses on form, while Chinese focuses on meaning. The grammatical structure of English is rigorous and in order, and the master-subordinate relationship is clear and clear. It is no exaggeration to say that sometimes, a long English sentence is mostly a compound sentence, and it is several lines long. Moreover, the structure is complex, including main clauses, adverbial clauses, attributive clauses, appositions, adjectives, and adverbs. However, the syntactic structure of Chinese is loose and flexible, and different meanings can be generated only by random collocation. This paper combines attention mechanism and neural network algorithm, applies it to Chinese-English machine learning translation, improves the effect of Chinese-English machine translation, and proposes a new model of machine translation based on logarithmic position representation and self-attention mechanism. Through the experimental research, it can be seen that the Chinese-

English machine translation model combining attention mechanism and bidirectional neural network proposed in this paper has a good practical translation effect.

Data Availability

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

Conflicts of Interest

The author declares no competing interests.

Acknowledgments

References

[1] E. Brynjolfsson, X. Hui, and M. Liu, "Does machine translation affect international trade? Evidence from a large digital platform," *Management Science*, vol. 65, no. 12, pp. 5449–5460, 2019.

[2] R. Schulte, "Expanding the horizon of translation studies," *Translation Review*, vol. 106, no. 1, pp. 1–3, 2020.

[3] K. Gurumurthy, D. Anand, and K. Karthik, "Democratizing translation using digital methods," *Language and Language Teaching*, vol. 10, no. 1, pp. 30–36, 2021.

[4] K. Karlgren, M. Lakkala, A. Toom, L. Ilomäki, P. Lahti-Nuuttila, and H. Muukkonen, "Assessing the learning of knowledge work competence in higher education–cross-cultural translation and adaptation of the collaborative knowledge practices questionnaire," *Research Papers in Education*, vol. 35, no. 1, pp. 8–22, 2020.

[5] C. Dinneen, "Students' use of digital translation and paraphrasing tools in written assignments on direct entry English programs," *English Australia Journal*, vol. 37, no. 1, pp. 40–51, 2021.

[6] Y. Sun and D. Li, "Digital humanities approaches to literary translation," *Comparative Literature Studies*, vol. 57, no. 4, pp. 640–654, 2020.

[7] M. M. J. F. Farooque and B. Jadad, "A comparative analysis of translation software using data science approach to Arabic statements," *International Journal of Advanced Computer Systems and Software Engineering*, vol. 2, no. 3, pp. 13–19, 2021.

[8] J. C. S. Hvidt, L. F. Christensen, C. Sibbersen, S. Helweg-Jørgensen, J. P. Hansen, and M. B. Lichtenstein, "Translation and validation of the system usability scale in a Danish mental health setting using digital technologies in treatment interventions," *International Journal of Human–Computer Interaction*, vol. 36, no. 8, pp. 709–716, 2020.

[9] Y. Tsai, "Collaborative translation in the digital age," *Research in Language*, vol. 18, no. 2, pp. 119–135, 2020.

[10] M. Y. Gamal, "Context, field and landscape of audiovisual translation in the Arab world," *ESSACHESS-Journal for Communication Studies*, vol. 13, no. 25, pp. 73–105, 2020.

[11] M. W. Heft, C. H. Fox, and R. P. Duncan, "Assessing the translation of research and innovation into dental practice," *JDR Clinical & Translational Research*, vol. 5, no. 3, pp. 262–270, 2020.

[12] C. Clivaz, "Lost in translation? The odyssey of 'digital human-ities' in French," *Studia Universitatis Babeş-Bolyai Digitalia*, vol. 62, no. 1, pp. 26–41, 2017.

[13] L. Shafirova, D. Cassany, and C. Bach, "Transcultural literacies in online collaboration: a case study of fanfiction translation from Russian into English," *Language and Intercultural Com-munication*, vol. 20, no. 6, pp. 531–545, 2020.

[14] S. Vandepitte and E. Lefever, "Translation as a multilingual activity in the digital era," *Revue francaise de linguistique appli-quee*, vol. XXIII, no. 2, pp. 59–72, 2018.

[15] B. Türkmen, "Utilising digital media as a second language (L2) support: a case study on Netflix with translation applications," *Interdisciplinary Description of Complex Systems: INDECS*, vol. 18, no. 4, pp. 459–470, 2020.

[16] H. Minors, "Desblache, Lucile (2019): music and translation: new mediations in the digital age. London: Macmillan, 407 p," *Meta: journal des traducteurs/Meta: Translators' Journal*, vol. 66, no. 1, pp. 230–232, 2021.

[17] J. Singh, "Lost in translation? The emergence of the digital Guru Granth Sahib," *Sikh Formations*, vol. 14, no. 3-4, pp. 339–351, 2018.

[18] Y. A. Rankin, J. O. Thomas, and N. M. Joseph, "Intersectional-ity in HCI," *Interactions*, vol. 27, no. 5, pp. 68–71, 2020.

[19] A. A. Fadele, A. Kamsin, K. Ahmad, and R. A. Rasheed, "A novel Hadith authentication mobile system in Arabic to Malay language translation for android and iOS phones," *Interna-tional Journal of Information Technology*, vol. 13, no. 4, pp. 1683–1692, 2021.

[20] P. Tracey, E. Dalpiaz, and N. Phillips, "Fish out of water: trans-lation, legitimation, and new venture creation," *Academy of Management Journal*, vol. 61, no. 5, pp. 1627–1666, 2018.

[21] R. N. Parsa, "Trends in e-tools and resources for translators and interpreters," *Translation & Interpreting*, vol. 13, no. 2, pp. 183–186, 2021.