Optimization of Unsupervised Neural Machine Translation Based on Syntactic Knowledge Improvement

Aiping Zhou
Department of Fundamental Courses, Southeast University Chengxian College, Nanjing, 210088, China

***Abstracts*—Unsupervised Neural Machine Translation is a crucial machine translation method that can translate in the absence of a parallel corpus and opens up new avenues for intercultural dialogue. Existing unsupervised neural machine translation models still struggle to deal with intricate grammatical relationships and linguistic structures, which leads to less-than-ideal translation quality. This study combines the Transformer structure and syntactic knowledge to create a new unsupervised neural machine translation model, which enhances the performance of the existing model. The study creates a neural machine translation model based on the Transformer structure first, and then introduces sentence syntactic structure and various syntactic fusion techniques, also known as the Transformer combines grammatical knowledge. The results show that the Transformer combines grammatical knowledge paired with Bi-Long Short-Term Memory proposed in this research has better performance. The accuracy and F1 value of the combined model in the training dataset are as high as 0.97. In addition, the time of the model in real sentence translation is controlled within 2s, and the translation accuracy is above 0.9. In conclusion, the unsupervised neural machine translation model proposed in this study has better performance, and its application to actual translation can achieve better translation results.**

***Keywords*—Unsupervised; Neural network; Machine translation; Grammatical knowledge; Transformer; LSTM**

I. INTRODUCTION

With the progress of globalization and the increasing frequency of information exchange, machine translation, is an important artificial intelligence technology, plays an important role in connecting different languages and cultures [1-2]. Machine Translation (MT) is the process of using computer software to convert text or speech from one natural language to another. MT, as a branch of computer-assisted translation, aims to achieve barrier-free language communication between people. Unsupervised Neural Machine Translation (UNMT) belongs to one kind of MT, and the use of UNMT model to carry out translation tasks can not only improve the translation speed and save the cost, but also be able to deal with multiple languages at the same time, which is an important value of language utilization. As a research direction that has been developing rapidly in recent years, the main purpose of UNMT is to provide the best solution to the problem in the field of unsupervised translation, and to provide the best solution to the problem. Its main purpose is to carry out automatic translation without parallel corpus, so as to improve the speed and accuracy of machine translation [3-4]. Currently, the traditional UNMT model is still facing a series of challenges, for example, in the environment without parallel corpus, UNMT is often difficult to obtain effective linguistic correspondences, thus affecting the quality and accuracy of translation [5]. Against this background, the emergence of the Transformer structure has revolutionized the field of machine translation, especially its demonstrated efficiency in sequence-to-sequence learning [6]. At the same time, how to better integrate grammatical knowledge into the UNMT model has become an urgent challenge. Based on the above problems and challenges, this study aims to deeply explore and propose a novel UNMT model, which not only incorporates the advantages of the Transformer structure, but also the theoretical features of grammatical knowledge. The newly constructed UNMT model aims to achieve higher translation quality and effect, and at the same time solve the technical and theoretical problems of the traditional UNMT model in machine translation, so as to provide certain technical reference value for the field of machine translation.

In order to facilitate readers to better understand the content of the article and the framework of the article, this research divides the article into a total of six sections, which are the introduction, literature review, method design, result analysis, discussion and conclusion chapters. Section I mainly introduces the background of the study, the current status of the study, the research methodology, and the significance of the study. The literature review chapter mainly analyzes and summarizes the related studies of others so as to prove the novelty of this research which is mentioned in Section II. The research methodology in Section III mainly explains how to build the optimized UNMT model and introduces some related machine translation techniques. The result analysis section mainly tests the performance and practical application effect of the UNMT model, so as to prove the effectiveness of the model which is mentioned in Section IV. The discussion in Section V is mainly to further analyze and summarize the reasons for the better performance of the model according to the experimental results. The conclusion in Section VI is a concise summary of the whole paper.

II. RELATED WORK

UNMT is an approach to MT whose main feature is to translate without a parallel corpus. Currently, it has been optimised by a number of experts in combination with deep learning. In order to address the drawback that remotely supervised relation extraction is seriously affected by mislabelling in practical applications, Xiao et al. proposed a Transformer module for remotely supervised relation extraction with multi-instance learning using a hybrid attention mechanism. On the remote supervised relation extraction task, experimental results demonstrated that the method developed by the research performed better than the state-of-the-art algorithms at the time [7]. The Transformer concept has additionally been used in the machining sector. The apriori knowledge of the target text could not be fully utilised by the conventional automatic speech mistake detection systems, according to Zhang et al. Therefore, Zhang et al. proposed to apply the Transformer model to it, and the results of this study showed that the method could obtain a relative improvement of 8.4% on the F-1 scoring metrics, which is advancing significance for the optimisation of automatic speech error detection methods [8]. Li et al. concluded that the existing anomaly detection methods in the power industry do not fully exploit the potential value of the data. Therefore, an anomaly detection model based on graph attention and Transformer was proposed. Li et al. designed experiments based on power data in a region of China [9]. Li et al. argued that current neural TTS models suffer from robustness problems thus leading to audio anomalies. In order to construct a Neural network (NN) model capable of synthesising both natural and stable audio, thus a Transformer based TTS model called RobuTrans was proposed in [10]. Through experiments, it was found that the model solves the robustness problem that exists in the TSS model. According to Xiao et al., the current entity and relation extraction suffers from noise labelling issues and is unable to recognise the relationship between relations and phrases. This led to the proposal of a hybrid depth NN model based on Transformer and other models. The outcomes of many experiments demonstrated that the model was superior at entity and relation extraction and could filter noisy words [11].

As a posteriori regularisation technique to direct the training efficiency of unsupervised MT models during repeated reverse translation, Ren et al. presented a phrase-based statistical MT model. This study jointly optimises the SMT and NMT models under a unified expectation maximisation framework and gradually improves the performance of both models during the iterative process. The results of the study show that the proposed scheme can achieve two advantages. Filtering the noise in the phrase table by SMT can promptly mitigate the negative impact of errors during iterative back translation. Meanwhile, NMT can make up for the inherent lack of fluency in SMT [12]. Li et al. utilised spatio-temporal maps obtained from videos and the spatial and temporal interactions of objects to facilitate potential spatial alignment and remove translation ambiguity in UNMT. The designed model employs multimodal backtranslation and feature pseudo-visual hubs, and learns a shared multilingual visual-semantic embedding space that fuses visual hub subtitles as additional weak supervision. The proposed model is validated on the VATEX Translation 2020 and HowToWorld datasets for translation in sentences and words with good generalisation performance [13]. Sun et al. empirically investigated the performance of four different languages (French, German, Chinese, and Japanese) on the English UNMT model. In addition, a simple general method is proposed for improving the translation performance of these four language pairs. To address the shortcoming that different language pairs have significant delayed convergence in the denoising process, Sun H et al. proposed a pseudo-data based UNMT [14].

In summary, a number of scholars have carried out a series of studies on Transformer structure and UNMT model. Among them, the research on Transformer structure mainly focuses on the detection of various abnormal signals and data, while the research on UNMT model focuses on the optimisation of model translation effect. Based on the above background, this research innovatively fuses the Transformer structure with GK and uses the fused model in the field of UNMT, aiming at better extraction of English grammatical error features and detection of its incorrect grammar.

III. UNMT Model Construction Based on GK Improvement

The emergence of neural MT models has led to the gradual replacement of end-to-end single MT, and the continuous optimisation of neural MT models has also made the translation effect of various small language translation models closer and closer to that of human translation. In this research, the traditional Long Short-Term Memory (LSTM) and the neural MT model under the Transformer structure are firstly introduced, and then it is optimised by combining with the GK (Grammatical Knowledge) structure, and a new UNMT model is proposed.

A. *Research on Neural MT Modelling Based on Transformer Structure*

With the continuous combination of deep learning and MT technology, the neural MT model has become the most mainstream intelligent translation model. The biggest advantage of Recurrent Neural Network (RNN) in English translation is its ability to remember the historical information of a sentence, to expand the individual words in a sentence in time steps, and to check its grammar [15]. RNN automatically corrects grammatical errors by transforming the input natural language text into the output of the RNN so that grammatical errors can be corrected automatically. The standard expression form of the RNN is shown in Eq. (1) and Eq. (2).

$$h_t = \varphi_h \left(W_h x_t + U_h h_{t-1} + b_h \right) \tag{1}$$

In Eq. (1), x_t denotes the input of the time step at moment t and $t = \left[1, 2, \cdots m\right]$. h_t denotes the implied state of the output of the time step at moment t. W_h, U_h, and b_h denote the relevant parameters of the output implied state, respectively. φ_h, on the other hand, denotes the nonlinear activation function of the output implied state [16].

$$y_t = \varphi_y \left(W_y h_t + b_y \right) \tag{2}$$

In Eq. (2), y_t denotes the output of the network at the moment t. W_y and b_y are the relevant parameters of the output state of the network, respectively. φ_y denotes the nonlinear activation function of the network output. Researchers optimised the RNN and designed new loop units by adjusting the nonlinear activation function in the network. The common LSTM gradually gets new applications in MT problems. The unit structure of LSTM is shown in Fig. 1.

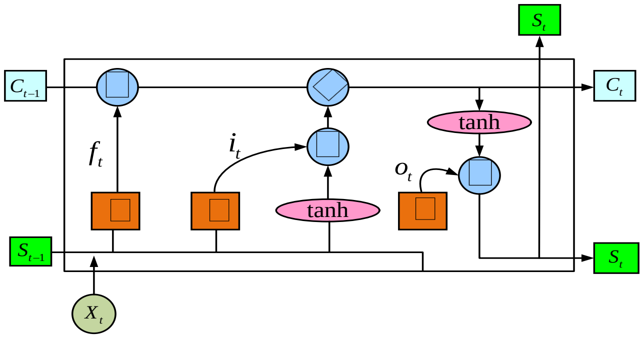


Fig. 1. LSTM cell structure diagram.

In Fig. 1, the LSTM consists of three gate structures, and selectively receives information through memory units. Where X_t denotes the input data of the input layer at the moment t. S_t denotes the neuron state of the hidden layer at the moment t. C_t denotes the memory unit at the moment t. The three σ in Fig. 1 indicate the three gate structures in LSTM from left to right. f_t, i_t, and o_t indicate the parameters of the three gates, respectively [17].

$$f_t = \sigma' W_x^f \cdot X_t + W_s^f \cdot S_{t-1} + b_f \tag{3}$$

In Eq. (3), W_x^f, W_s^f denotes the weight matrix, b_f denotes the input gate bias vector. σ' denotes the input gate. f_t denotes the input gate parameters.

$$i_t = \sigma'' W_x^i \cdot X_t + W_s^i \cdot S_{t-1} + b_i \tag{4}$$

In Eq. (4), W_x^i and W_s^i denote the weight matrix, b_i denotes the forgetting gate bias vector. σ'' denotes the forgetting gate. S_{t-1} denotes the neuron state at the moment of $t-1$. i_t denotes the forgetting gate parameter.

$$o_t = \sigma''' W_x^o \cdot X_t + W_s^o \cdot S_{t-1} + b_o \tag{5}$$

In Eq. (5), W_x^o, W_s^o denotes the weight matrix, b_o denotes the output gate bias vector. σ''' denotes the output gate. o_t is the output gate parameter.

$$\tilde{c}_t = \tanh\left(W_x^c \cdot X_t + W_s^c \cdot S_{t-1} + b_c\right) \tag{6}$$

In Eq. (6), $\tilde{c}_t$ is the output of the memory cell obtained by the forgetting gate parameter i_t after the calculation of the tanh function. W_x^c and W_s^c denote the weight matrix. b_c denotes the bias vector of the forgetting gate after the calculation of the tanh function.

$$c_t = i_t \odot \tilde{c}_t + f_t \odot c_{t-1} \tag{7}$$

In Eq. (7), $\odot$ denotes the Hardamard product.

$$S_t = o_t \odot \tanh\left(c_t\right) \tag{8}$$

Eq. (8) is the equation for the neuron state of the hidden layer at moment t.

Since both RNN and LSTM are prone to the problems of gradient vanishing and gradient explosion when performing parameter updates, the study builds a small language translation model by combining the Transformer structure. Transformer is an NN structure used for sequence-to-sequence learning, which is better able to deal with the problem of long text. The structure of Transformer is shown in Fig. 2.

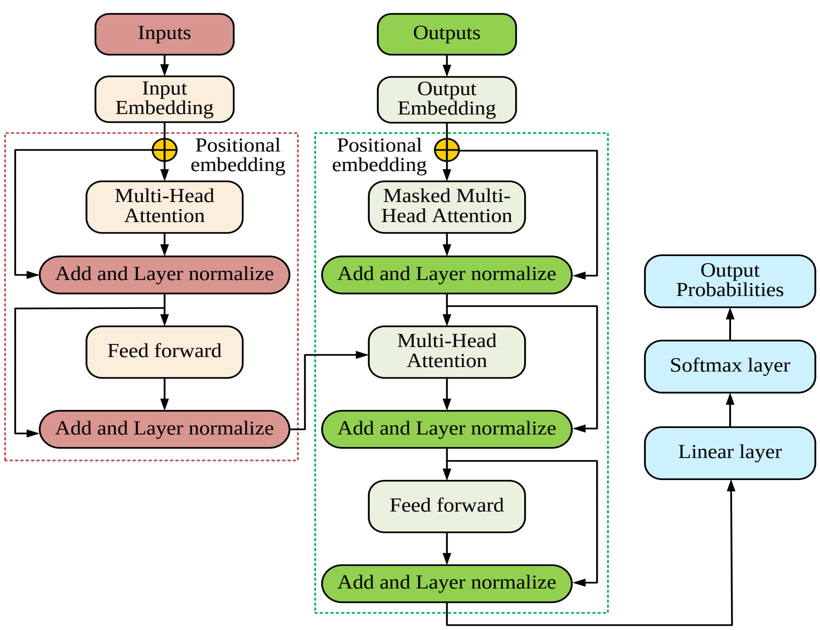


Fig. 2. Structure of the transformer model.

Fig. 2 depicts the Transformer model's structural layout. The operation flow of the encoder and decoder is shown in Eq. (9) and Eq. (10) [18].

$$e_1, e_2, \cdots, e_m = encoder\left(X1, X2, \cdots, Xm\right) \quad (9)$$

In Eq. (9), $e_1, e_2, \cdots, e_m$ denotes a string of input text sequence. $encoder\left(X1, X2, \cdots, Xm\right)$ denotes the encoder for encoding.

$$Y_t = decoder\left(e_1, e_2, \cdots, e_m, Y1, Y2, \cdots Y_{t-1}\right) \quad (10)$$

In Eq. (10), Y_t denotes the probability distribution vector of the decoded data, which is obtained by calculating the SoftMax function. $decoder\left(e_1, e_2, \cdots, e_m, Y1, Y2, \cdots Y_{t-1}\right)$ denotes the decoding operation on the probability distribution vector.

B. Study of UNMT Modelling Incorporating the Transformer Structure and GK

Since the traditional Transformer model tends to ignore the semantic information of the sentence in the translation process, which leads to the translation result deviating from the actual meaning, this study further proposes an optimized UNMT model based on the traditional Transformer structure combined with GK, notated as Transformer combines grammatical knowledge (TCGK). Traditional UNMT is an approach for training in MT tasks without using parallel corpus, i.e., corresponding sentence pairs between source and target languages. Fig. 3 depicts the UNMT model's fundamental structure.

Fig. 3 depicts the UNMT model's overall structure. Two monolingual semantic repositories, a language modelling board, and a reverse translation board are the primary components of the UNMT architecture shown in Fig. 3 [19]. When the words in the two monolingual semantic repositories are input into the model, they first need to be initialised. In the initialisation process, the main purpose is to encode the words, phrases and words so that each word can be recognised by the UNMT model, thus achieving the purpose of training the model. After the initialisation process, language modelling is required. In the modelling process, the encoder-decoder structure is used for denoising and at the same time allowing the encoder to learn the semantic information of the monolingual data. The mathematical expression for language modelling is shown in Eq. (11).

$$L_{\min} = E\left[-\log P_{s\to s}\left(x \middle| C\left(x\right)\right)\right] + E\left[-\log P_{t\to t}\left(x \middle| C\left(x\right)\right)\right] \quad (11)$$

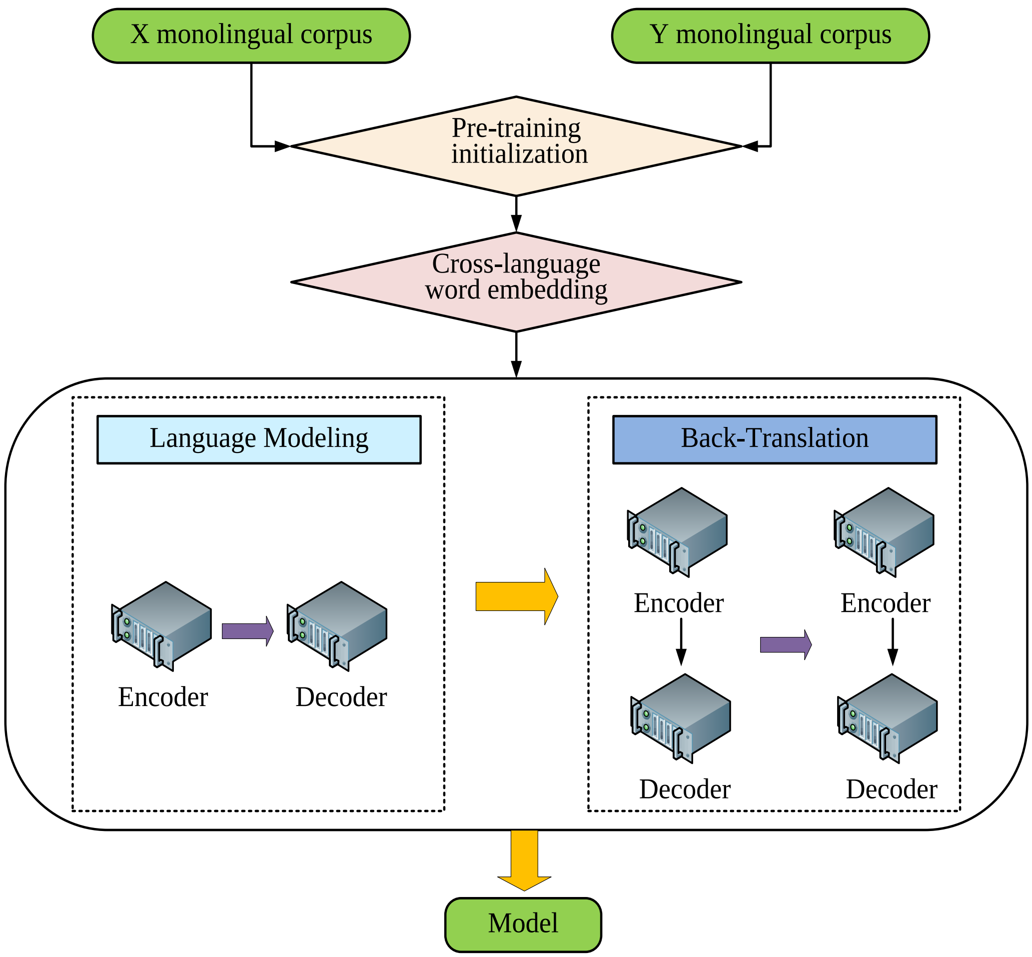


Fig. 3. Unsupervised neural machine translation model structure.

In Eq. (11), $L_{\min}$ denotes the minimum loss in the modelling process. C denotes the noise model. x denotes the sentence in the X monolingual semantic base. $P_{s\to s}$ denotes the source-side encoder-decoder combination. $P_{t\to t}$ denotes the target-side encoder-decoder combination. E denotes the energy consumption in the modelling process. The mathematical expression for reverse translation is shown in Eq. (12).

$$L'_{\min} = E\left[-\log P_{s\to t}\left(y \middle| u*\left(y\right)\right)\right] + E\left[-\log P_{t\to s}\left(x \middle| v*\left(x\right)\right)\right] \quad (12)$$

In Eq. (12), $L'_{\min}$ denotes the minimum loss in the reverse translation process. y denotes a sentence in the Y monolingual semantic base. $u*\left(y\right)$ denotes translating the source language according to the target language. $v*\left(x\right)$ denotes translating the target language according to the source language. $P_{s\to t}$ denotes the direction of translation from the target language to the source language. $P_{t\to s}$ denotes the direction of translation from the source language to the target language.

In order to learn the translation relationships between the source and destination languages, neural MT models typically need a sizable parallel corpus for training that comprises corresponding sentence pairs between the two languages. However, UNMT does not rely on parallel corpus, but is trained by using monolingual corpus. The UNMT model is obtained by optimising the encoder-decoder structure, and the basic idea is to learn the correspondence between source and target languages through self-supervised learning of monolingual corpus, so as to achieve the MT task. This study takes GK and syntactic structure into account for the model's optimisation on the basis of the conventional UNMT model in order to give the UNMT model a better translation effect that can accurately translate according to the syntactic structure and be close to the actual semantic environment. Firstly, the sentence syntactic structure is introduced, as shown in Fig. 4.

Fig. 4 shows the syntactic tree structure diagram of the sentence. To translate a complete sentence according to the actual context and grammatical structure, it is necessary to split its sentence syntactic structure first [20]. In Fig. 4, it can be seen that a complete sentence is composed of sentences or phrases. For phrase structure, its extracted syntactic labels contain constituent categories and phrase structure information. The hierarchical output of the phrase structure syntax contains the information of the various categories of words and the attributes of the words.

In order to allow unsupervised MTs to have a better knowledge of syntax, this research decided to use the results of the syntactic analysis to optimise the translation results of the model. After linearising these results and extracting their syntactic labels, and then combining them with the corresponding sentences, the combined data is used to train the source side of the denoising autoencoder. In this way, sentences and syntactic information can be jointly encoded into a new vector, thus creating a language model that incorporates syntactic information. The process of training the model with fused syntactic knowledge is shown in Fig. 5.

A flowchart of the model training process incorporating syntactic knowledge is demonstrated in Fig. 5. In Fig. 5, the optimised model has adapted the inputs of the denoising autoencoder, and multiple encoders are used to process the monolingual corpus, lexicality, phrase structure, and dependency syntax, respectively. During training, the model absorbs lexical and syntactic information and optimises the shared encoder and decoder parameters to better capture the implicit syntactic information in sentences. When decoding, the model utilises semantic, lexical and syntactic information to assist in the predictive generation of target words. By incorporating lexical and syntactic knowledge, the challenge of not being able to explicitly learn syntax can be addressed and the accuracy of the translation can be improved.

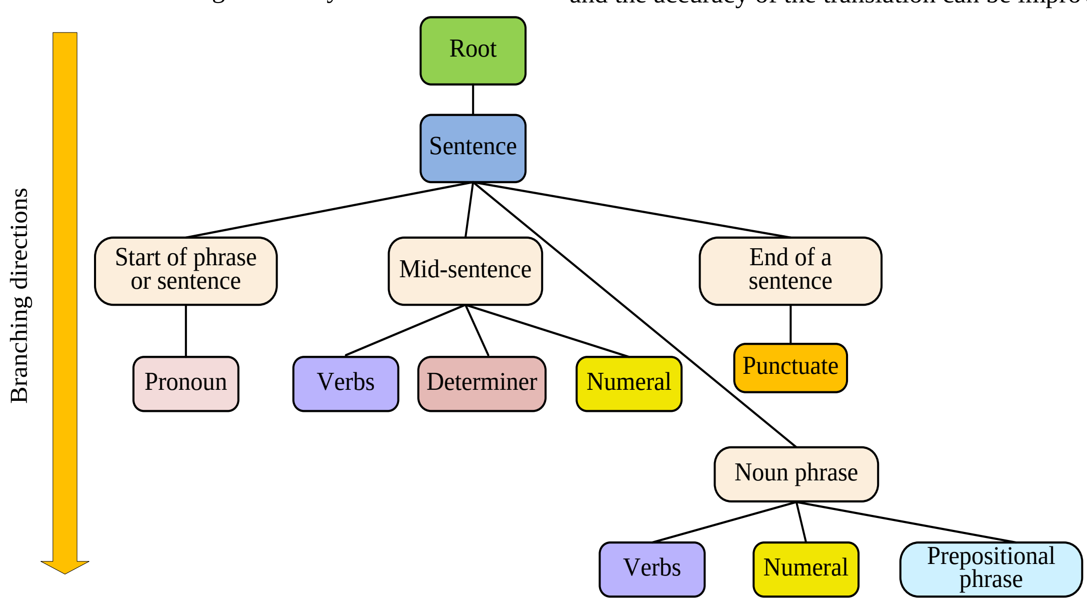


Fig. 4. Sentence syntactic tree structure.

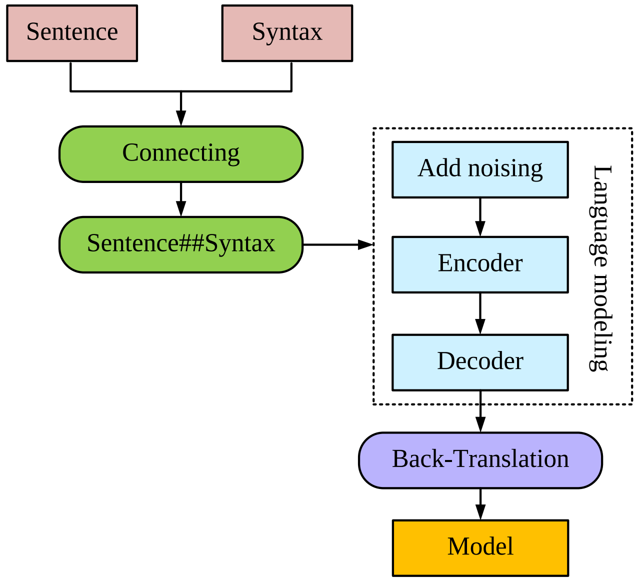


Fig. 5. Flow chart of model training incorporating syntactic knowledge.

In addition to adopting the approach of fusing syntactic knowledge in Fig. 5, the study also uses Bi-Long Short-Term Memory (Bi-LSTM) in NN to extract vectors of features for syntactic analysis sequences and splice them with sentence vectors to obtain a vector containing semantic and syntactic information, and uses this vector to train TCGK, so as to improve the translation quality of the TCGK model. The flow chart of Bi-LSTM extraction of syntactic vectors is shown in Fig. 6. The flowchart of Bi-LSTM for extracting syntactic vectors is shown in Fig. 6.

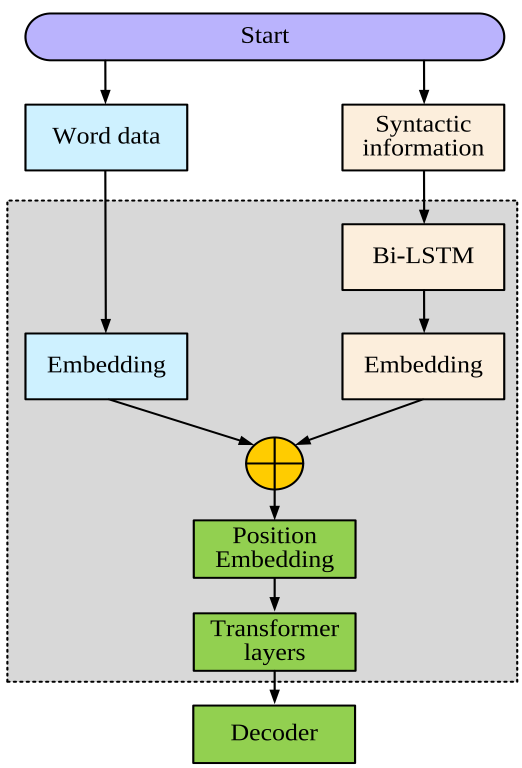


Fig. 6. Flow chart of sentence normal vector extraction under Bi-LSTM.

As shown in Fig. 6, Bi-LSTM is used to extract syntactic vectors. In order to more closely combine monolingual sentences and explicit syntactic information, the syntactic tree sequence is first linearised. At the input of Transformer, this study combines sentence vectors with syntactic vectors processed through Bi-LSTM to form new fusion vectors. The syntactic features are first transformed into high-dimensional vectors through the word embedding layer of the NN and then spliced with the sentence vectors, and this resulting integrated feature vector does not involve modifying the syntactic content. Using this fused vector, the encoder-decoder starts iterative training and stops iterative training until the model has better translation results.

IV. Performance Analysis of UNMT Models using Transformer Structure and GK

The result analysis section tested the performance of various types of translation models before testing the UNMT model created in the aforementioned study. This demonstrated that the translation performance of the model used in this study was superior through the indicators of detection accuracy, change of loss curves, and F1 value. In addition, the study further compares the translation effect of each translation model in practical applications. The results of the study found that the UNMT model combined with Bi-LSTM has higher translation accuracy and teacher-student satisfaction.

A. Performance Analysis of Different Translation Models

The News Crawl dataset was first chosen as the experimental dataset, and Newstest2020 and Newstest2021 were chosen as the experimental training dataset and test dataset, respectively, to test the performance of the model under the Transformer structure. In Newstest2020 and Newstest2021, there were 5000 corpora each. The corpus for Newstest2020 and Newstest2021 is 5000. Table I displays the settings for the experimental model's parameters.

The basic network parameters in the Transformer model are given in Table I, including the number of its encoder-decoder layers, the number of layers of the multi-head attention mechanism, the dimensions of the word embedding and hidden layers, and the learning rate. The detailed composition of the Newstest2020 and Newstest2021 experimental dataset information is shown in Table II.

Table II shows the details of the Newstest2020 and Newstest2021 experimental datasets, describing the source of the datasets, the composition of the language pairs, the number of samples, and the purpose of the dataset usage, respectively. In addition to utilizing the Newstest2020 and Newstest2021 experimental datasets for testing, the study also selected some public language datasets for testing. In order to compare the performance of the two syntactic fusion methods in the model TCGK, this study notated the syntactic fusion approach in Fig. 5 as TCGK+Common coding, and the syntactic fusion approach in Fig. 6 as TCGK+Bi-LSTM, and introduced the traditional Transformer model as well as the LSTM model for the comparison of the model translation performance. The detection accuracy of the four models in different datasets is shown in Table III.

TABLE I. PARAMETER SETTING OF THE EXPERIMENTAL MODEL

Transformer model parameters	Parameter values
Number of encoder and decoder layers	4 layers
Number of layers of multi-head attention mechanism	8 layers
Dimension of word embedding and hidden layers	1024
Learning rate under Adam's optimization method	0.001
Batch Size	32
Model regularization Dropout	0.1

TABLE II. EXPERIMENTAL DATA SET INFORMATION TABLE

Data set information	Introduction
Source	It mainly contains news texts, which are sourced from various news websites and agencies.
Language pair	These two datasets cover multiple language pairs. For example, English to German, English to French, English to Chinese, etc.
Sample size	The number of corpus in Newstest2020 and Newstest2021 is 5000.
Aim	Evaluate the translation performance of machine translation models.

In Table III, a total of five public datasets, Newstest2020, Newstest2021, Para Crawl, Europarl, and Common Crawl, are selected for testing. Europarl dataset is a dataset based on the records of the European Parliament, covering 21 European languages. Common Crawl is a multilingual aligned dataset based on web crawling. Para Crawl is a multilingual parallel corpus for large-scale web crawling. As shown in Table III, the detection accuracies of LSTM in Newstest2020, Newstest2021, ParaCrawl, Europarl, and Common Crawl are 0.78, 0.77, 0.71, 0.66, and 0.69, respectively. Transformer in Newstest2020, Newstest2021, ParaCrawl, Europarl, and Common Crawl with detection accuracies of 0.82, 0.83, 0.75, 0.71, and 0.72, respectively. TCGK+Common coding in Newstest2020, Newstest2021, Para Crawl, Europarl, and Common Crawl were 0.88, 0.89, 0.86, 0.82, and 0.83, respectively. The detection accuracy of TCGK+Bi-LSTM in Newstest2020, Newstest2021, Para Crawl, Europarl, Common Crawl are 0.96, 0.97, 0.91, 0.92, and 0.93, respectively. Among them, the TCGK+Bi-LSTM model is able to achieve the highest detection accuracy in the datasets Newstest2020 and Newstest2021. The translation performance of the four models will be further tested in combination with the datasets Newstest2020 and Newstest2021.

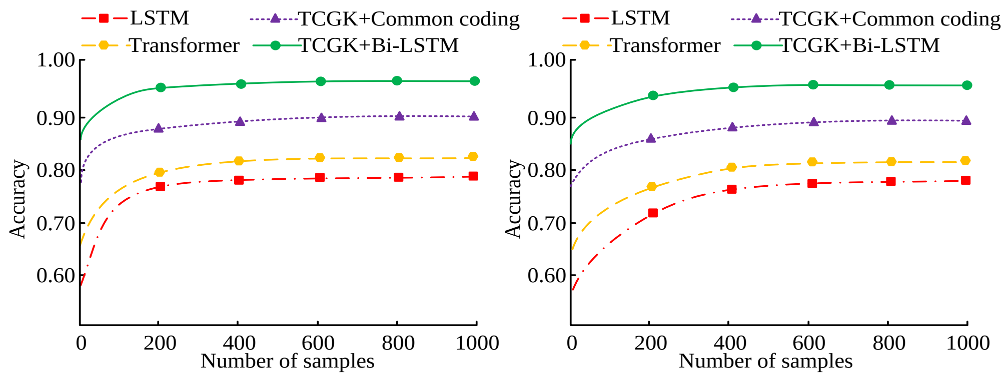
In Fig. 7, the incorrect syntax detection accuracy values for the various translation models in the training dataset and test dataset are displayed. Fig. 7(a) and Fig. 7(b) among them illustrate the detection accuracy of the four translation models for the training dataset and the test dataset, respectively: LSTM, Transformer, TCGK+Common Coding, and TCGK+Bi-LSTM. The four translation models' detection accuracies for faulty grammar samples exhibit an increasing trend as the number of detected samples rises, as shown in Fig. 7(a) and Fig. 7(b). The four translation models, LSTM, Transformer, TCGK+Common coding, and TCGK+Bi-LSTM, each have detection accuracy scores in the training dataset that are 0.78, 0.82, 0.90, and 0.97, respectively. In the testing dataset, the highest detection accuracy values of the four translation models, LSTM, Transformer, TCGK+Common coding, and TCGK+Bi-LSTM, are 0.78, 0.82, 0.90, and 0.97, respectively. In the testing dataset, the highest detection accuracy values of the four translation models, LSTM, Transformer, TCGK+Common coding, and TCGK+Bi-LSTM, are 0.77, 0.81, 0.88, and 0.96, respectively.

The graphs of the variation of loss values for different translation models are shown in Fig. 8. Among them, all the figures in Fig. 8 show the actual loss curves and the specific changes of the training loss curves of the four translation models, namely, LSTM, Transformer, TCGK+Common coding, and TCGK+Bi-LSTM, in the training process, respectively. Comparing the loss change curves of the four models, it can be found that compared to the other three models, the training loss curve and the actual loss curve of TCGK+Bi-LSTM basically overlap during the training process, so the stability of this model is better during the training process, and there will not be large data fluctuations.

The variance of F1 values for various translation models in the training dataset and test dataset is depicted in Fig. 9. The study introduces F1 values for testing to better represent the detection performance of each model. The F1 values obtained by the four models in the training dataset and test dataset are displayed in Fig. 9(a) and Fig. 9(b), respectively. In Fig. 9(a), the highest F1 values of the four translation models, LSTM, Transformer, TCGK+Common coding, and TCGK+Bi-LSTM, are 0.76, 0.81, 0.90, and 0.97, respectively. In Fig. 9(b), the highest F1 values of the four translation models, LSTM, Transformer, TCGK+Common coding, and TCGK+Bi-LSTM, the highest F1 values of the four translation models are 0.75, 0.80, 0.89, and 0.96, respectively.

TABLE III. DETECTION ACCURACY OF FOUR MODELS IN DIFFERENT DATA SETS

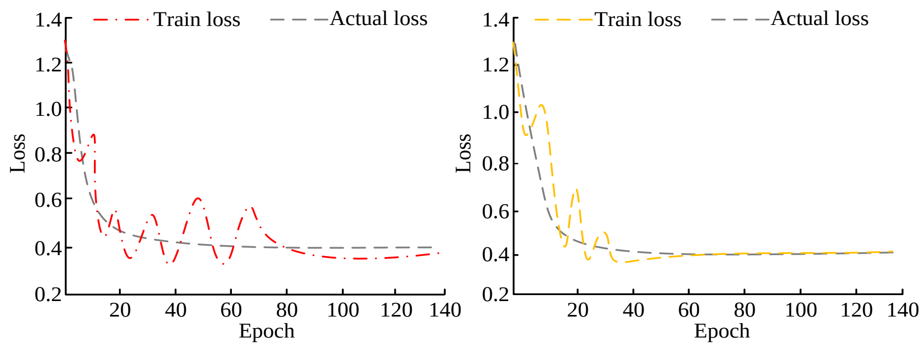
Model	Data set type				
	Newstest2020	Newstest2021	ParaCrawl	Europarl	Common Crawl
LSTM	0.78	0.77	0.71	0.66	0.69
Transformer	0.82	0.83	0.75	0.71	0.72
TCGK+Common coding	0.88	0.89	0.86	0.82	0.83
TCGK+Bi-LSTM	0.96	0.97	0.91	0.92	0.93



(a) Translation accuracy of different models in the training dataset

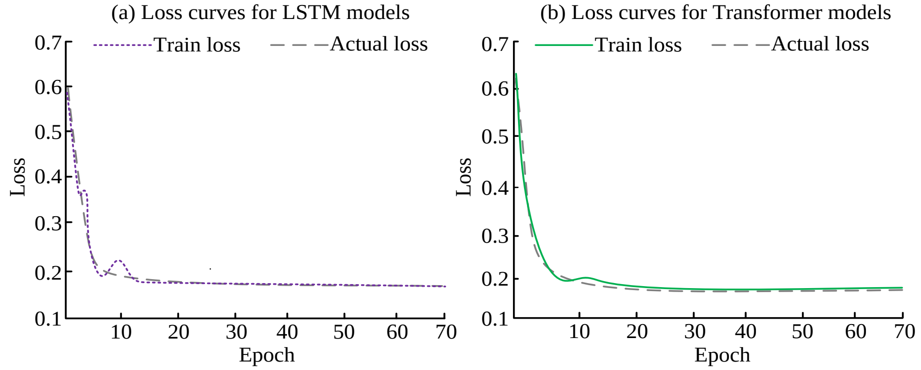
(b) Translation accuracy of different models in the testing dataset

Fig. 7. Translation accuracy values for the different translation models.



(a) Loss curves for LSTM models

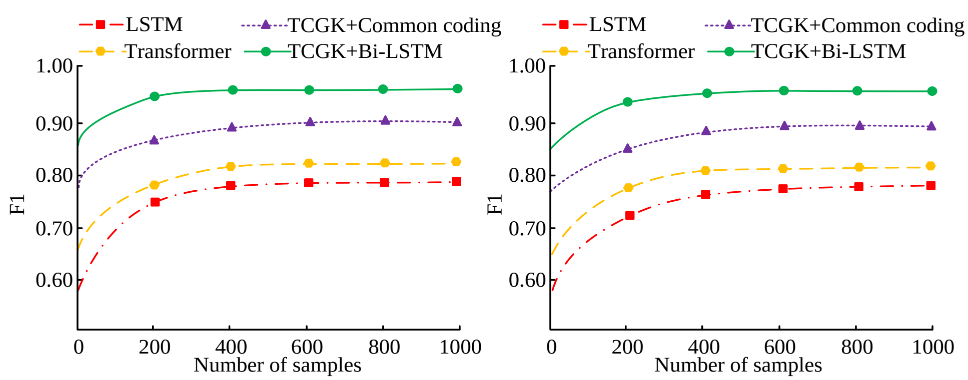
(b) Loss curves for Transformer models



(c) Loss curves for TCGK+Common coding models

(d) Loss curves for TCGK+Bi-LSTM models

Fig. 8. Loss values for the different translation models.



(a) F1 score of different models in the training dataset

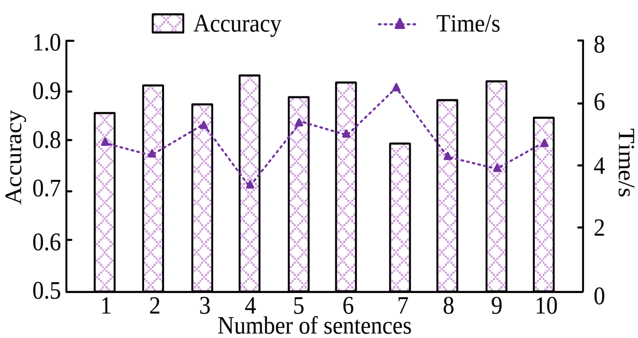
(b) F1 score of different models in the testing dataset

Fig. 9. Translation F1 values for the different translation models.

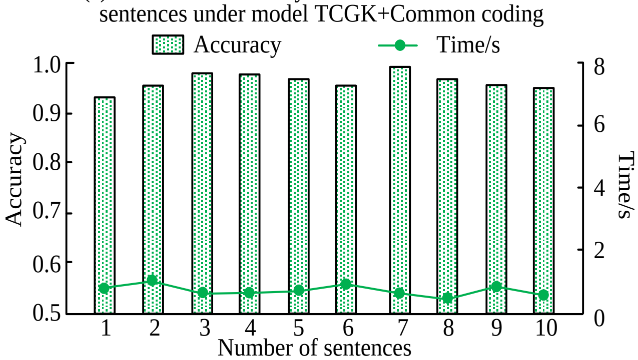
B. Analysis of the Effectiveness of the Application of Different Translation Models

The results of the analysis of the above performance indicators show that TCGK+Bi-LSTM has better performance compared with the other three translation models. TCGK+Bi-LSTM not only has better error grammar recognition accuracy values and F1 values, the change of the loss curve of this network during training is also basically the same as the actual change. To test the effectiveness of the four models in real English sentence translation, the study randomly selected 10 English utterances from a high school English textbook for testing. The translation accuracy and translation time of the two optimised UNMT models in real translation are shown in Fig. 10.

The translation accuracy and translation time of TCGK+Common coding and TCGK+Bi-LSTM in different English utterances are demonstrated in Fig. 10. Fig. 10(a) and (b) shows the translation accuracy and translation time of TCGK+Common coding and TCGK+Bi-LSTM, respectively. Comparing the translation effects of the two models in ten English utterances, it can be seen that the highest translation accuracies of TCGK+Common coding and TCGK+Bi-LSTM are 0.93 and 0.99, respectively. The shortest translation times of TCGK+Common coding and TCGK+Bi-LSTM take 3.2s and 0.5s, respectively. In addition, the translation accuracies of TCGK +Common coding model has a large change in the accuracy value during the translation process, and its translation elapsed time fluctuates more. Therefore, compared with TCGK +Common coding, TCGK +Bi-LSTM has better translation effect in practical applications.



(a) Translation accuracy and translation time for ten sentences under model TCGK+Common coding



(b) Translation accuracy and translation time for ten sentences under model TCGK+Bi-LSTM

Fig. 10. Translation accuracy and translation time of the two translation models in practice.

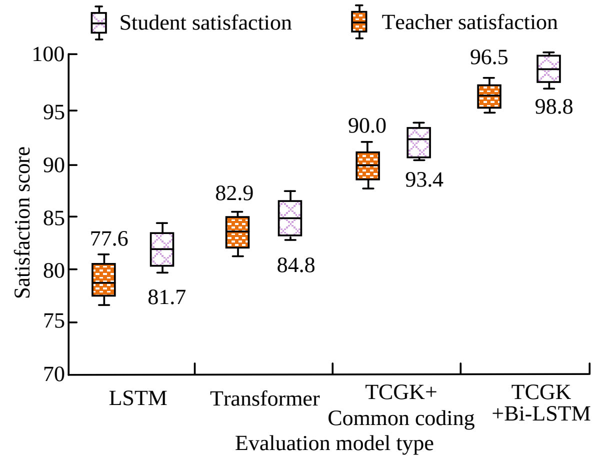


Fig. 11. Satisfaction of university teachers and students with the four translation models.

Fig. 11 shows the satisfaction scores of university students and teachers for the four translation models in practical applications. As shown in Fig. 11, the satisfaction scores of university students for the four translation models LSTM, Transformer, TCGK+Common coding, and TCGK+Bi-LSTM are 77.6, 82.9, 90.0, and 96.5, respectively. The satisfaction scores of university teachers for the four translation models LSTM, Transformer, TCGK+Common coding, and TCGK+Bi-LSTM are 81.7, 84.8, 93.4, and 98.8, respectively. In conclusion, TCGK+Bi-LSTM not only have better translation performance in practical applications, but also have higher satisfaction of university teachers and students for this model.

V. DISCUSSION

In order to improve the accuracy and efficiency of machine translation, this research combines the fusion of Transformer structure and grammar knowledge to optimize the unsupervised neural machine translation model, and finally builds the TCGK+Bi-LSTM translation model. By comparing and analyzing the performance of various types of models as well as their practical application effects, the following discussion is derived from this research.

From the experimental results of error syntax detection accuracy, it is obvious that the TCGK+Bi-LSTM model has better error syntax detection effect compared to LSTM, Transformer, and TCGK+Common coding. The TCGK+Bi-LSTM model outperforms the other three models in terms of test accuracy and F1 value in both the training dataset and the test dataset. The reason behind the high detection accuracy and F1 value of the TCGK+Bi-LSTM model is that the combination of the deep self-attention mechanism of the Transformer structure and the Bi-LSTM network enables the model to better capture long-distance dependencies and complex syntactic structures in sentences. In addition, the TCGK+Bi-LSTM model has a better loss profile compared to LSTM, Transformer, and TCGK+Common coding, which further illustrates that the Transformer structure and the Bi-LSTM network can improve the stability of the model during the training process, which enables the model to obtain more accurate test values.

In addition, although translation accuracy is the primary index of the machine translation task, fast translation speed is also very critical in practical applications. Especially in situations where a large number of translations are required, such as online services or real-time translation applications, efficient translation speed can greatly improve the user experience. The TCGK+Bi-LSTM model also has a significant advantage in translation time compared to the other three models. This is because the introduction of the Transformer structure and the Bi-LSTM network enables the model to process the information features faster, thus achieving fast translation. Finally, the TCGK+Bi-LSTM model was also able to achieve a high level of teacher and student satisfaction in real-world applications, thus proving the value of this technique in real-world applications.

Although this study provides valuable insights into the performance of the proposed model in Chinese-English translation tasks, there are some limitations. First, the experiments were mainly conducted based on specific datasets and specific tasks, and the performance of the proposed model should be further validated on more datasets and multiple language pairs in the future. In addition, although this experiment examined the performance of several models, there are still more existing and emerging modeling approaches that deserve further exploration and comparison. Based on the current findings, future research can further examine the performance of the models on other language pairs or larger datasets. In addition, it can also explore how to further optimize the structure or parameters of the model to improve its performance in specific tasks or scenarios. In summary, this study provides valuable insights into unsupervised neural machine translation models that incorporate Transformer structural and syntactic knowledge, and provides useful directions for future research.

VI. Conclusion

This research utilizes the knowledge of Transformer structure and syntax to construct a new UNMT model that aims to improve the performance and translation accuracy of existing translation models. The results of the study show that the proposed TCGK+Bi-LSTM model model significantly outperforms the other three models in terms of detection accuracy and F1 value on both training and testing datasets. In addition, the TCGK+Bi-LSTM model exhibits higher translation accuracy and translation speed than the TCGK+Common Coding model in real translation tests involving English sentences. Finally, the TCGK+Bi-LSTM model gained high satisfaction among university teachers and students, further validating its effectiveness. Since this study mainly focused on the performance of the model in Chinese-English translation tasks, it does not have comprehensive coverage, and subsequent studies should further extend the scope of the study to examine the performance of the model on more language pairs.

References

[1] Y. Zheng, G. Li, and W. Zhang, “A New Efficient Algorithm Based on Multi-Classifiers Model for Classification,” Int. J. Uncertainty Fuzziness Knowledge Based Syst., vol. 28, no. 1, pp. 25-46, 2020.

[2] M. Heidari, S. Lakshmivarahan, S. Mirniaharikandehei, G. Danala, S. K. R. Maryada, H. Liu, and B. Zheng, “Applying a random projection algorithm to optimize machine learning model for breast lesion classification,” IEEE Trans. Biomed. Eng., vol. 68, no. 9, pp. 2764-2775, 2021.

[3] C. Faustmann, M. Bajzek, H. Hick, J. Edtmayer, and S. Walch, “System models and model classification in tribological system development,” Syst. Eng., vol. 23, no. 6, pp. 783-794, 2020.

[4] S. P. A. Alewijnse, K. Buchin, M. Buchin, S. Sijben, and M. A. Westenberg, “Model-based segmentation and classification of trajectories,” ALGORITHMICA., vol. 80, no. 8, pp. 2422-2452, 2018.

[5] C. Lv, Z. Wu, X. Wang, Z. Dan, M. Zhou, “Ethnicity classification by the 3D discrete landmarks model measure in Kendall shape space,” Pattern Recognit. Lett., vol. 129, no. 1, pp. 26-32, 2020.

[6] J. Lee, H. S. Kim, and T. Y. Kang, “Classification algorithm using halftone features of counterfeit bills and CNN,” J. Forensic Sci., vol. 67, no. 1, pp. 345-352, 2022.

[7] Y. Xiao, Y. Jin, R. Cheng, and K. Hao, “Hybrid attention-based transformer block model for distant supervision relation extraction,” Neurocomputing., vol. 470, no. 22, pp. 29-39, 2022.

[8] Z. Zhang, Y. Wang, and J. Yang, “Text-conditioned transformer for automatic pronunciation error detection,” Speech Communication, vol. 130, no. 10, pp. 55-63, 2021.

[9] Y. Li, Z. Kai, X. Hao, H. Sheng, L. Shen, and S. Yu, “Abnormal Detection Based on Graph Attention Mechanisms and Transformer,” Acta Electronica Sinica, vol. 50, no. 4, pp. 900-908, 2022.

[10] N. Li, Y. Liu, Y. Wu, S. Liu, S. Zhao, and M. Liu, “Robutrans: A robust transformer-based text-to-speech model,” AAAI., vol. 34, no. 5, pp. 8228-8235, 2020.

[11] Y. Xiao, C. Tan, Z. Fan, Q. Xu, and W. Zhu, “Joint entity and relation extraction with a hybrid transformer and reinforcement learning based model,” AAAI., vol. 34, no. 5, pp. 9314-9321, 2020.

[12] S. Ren, Z. Zhang, S. Liu, M. Zhou, and S. Ma, “Unsupervised neural machine translation with smt as posterior regularization,” AAAI., vol. 33, no. 1, pp. 241-248, 2019.

[13] M. Li, P. Y. Huang, X. Chang, J. Hu, and Y. Yang, “Hauptmann A. Video pivoting unsupervised multi-modal machine translation,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 45, no. 3, pp. 3918-3932, 2022.

[14] H. Sun, R. Wang, M. Utiyama, B. Marie, K. Chen, Sumita E, and Zhao T, “Unsupervised neural machine translation for similar and distant language pairs: An empirical study,” ACM Trans. Asian Low-Resour. Lang. Inf. Process., vol. 20, no. 1, pp. 1-17, 2021.

[15] C. Li, and Z. Mao, “Generative adversarial network-based real-time temperature prediction model for heating stage of electric arc furnace,” Trans. Inst. Meas. Control., vol. 44, no. 8, pp. 1669-1684, 2022.

[16] M. Sajjad, F. Ramzan, M. U. G. Khan, R. Amjad, K. Mahyar, M. F. Suliman, and A. B. Saeed, “Deep convolutional generative adversarial network for Alzheimer’s disease classification using positron emission tomography (PET) and synthetic data augmentation,” Microsc. Res. Tech., vol. 84, no. 12, pp. 3023-3034, 2021.

[17] A. Bozikas, and G. Pitselis, “Incorporating crossed classification credibility into the Lee–Carter model for multi-population mortality data,” INSUR MATH ECON., vol. 93, no. 7, pp. 353-368, 2020.

[18] X. Wang, M. Cheng, J. Eaton, C. Hsieh, and S. F. Wu, “Fake node attacks on graph convolutional networks,” JCCE., vol. 1, no. 4, pp. 165-173, 2022.

[19] W. C. Chang, and I. H. R. Jiang, “iClaire: A Fast and General Layout Pattern Classification Algorithm with Clip Shifting and Centroid Recreation,” IEEE Trans. Comput. Aided Des. Integr. Circuits Syst., vol. 39, no. 8, pp. 1662-1673, 2019.

[20] F. Ay, G. İnce, M E. Kamaşak, and K. Y. Eksi, “Classification of pulsars with Dirichlet process Gaussian mixture model,” MNRAS., vol. 493, no. 1, pp. 713-722, 2020.